WEB APPENDIX

This Web Appendix contains information about the sample and additional analyses supporting the robustness of the results. The following paragraphs provide some background about these analyses to facilitate navigation through the explanations, tables and figures.

Web Appendix A: List of Keywords to Identify Negative Celebrity Endorser Publicity
Web Appendix A presents the list of keywords that this study used to identify cases of negative celebrity endorser publicity.

Web Appendix B: List of Events
Table B1 provides details about the sample, such as endorser name, short description of the event and year of occurrence, as well as how many firms were affected and the name of the specific firm(s). Web Appendix B also explains why 30 data points had to be excluded.

Web Appendix C: Specification Details for Event Study Regression Model
Web Appendix C shows that, assuming static covariates, the use of "traditional" *CARs* is equivalent to *ARs* in Event Study Regression.

Web Appendix D: Propensity Score Matching (PSM) to Adjust for a Firm's Decision to React
In the robustness analysis, this study specifies several models which correct for observable self-selection bias. For these models, Web Appendix D outlines the technical details of the PSM method and shows the results. Web Appendix D also tests the balancing properties to provide support that matching and bias reduction was successful.

Web Appendix E: Sensitivity Analysis of Selection on Unobservables
This sensitivity analysis investigates how robust the results are to the possible existence of an unobserved confounding variable. The analyses find that unobservable information is unlikely to drive the findings.

Web Appendix F: Robustness Coefficient
Although the evaluation of robustness is important, there is no commonly accepted definition thereof. To test the robustness of the findings, this study follows Neumayer and Plümper's (2017) "robustness coefficient" which ranges from 0 (not robust at all) to 1 (very robust). Figure F1 gives an illustrative example.

Web Appendix G: Generalized PSM to Adjust for a Firm's Decision When to React
PSM is limited to binary treatment variables (i.e., firm response: yes, no), but the decision about response timing is non-binary (i.e., firm could react on day 0, …, 20) and might be susceptible to self-selection bias as well. Thus, this study applies the Generalized PSM method to identify the optimal response day, and to divide active firms into fast (day 0-2) and slow (day 3-19) responders.

Web Appendix H: Results of Regression Models Used for Robustness Analysis
Tables H1 to H6 present the estimates and sample-size adjusted standard errors used for the robustness analyses.

Web Appendix I: Analyses of Same Event with Multiple Firms and Different Reactions

One could argue that a firm's reaction depends on the reaction of competing firms employing the same celebrity. Thus, instead of considering the reacting in isolation, this study also considers the reaction relative to other firms using the same celebrity. This subsample analysis confirms the results with two exceptions (subsample analysis provides support for the interaction *Maintain* Endorser Apology* which was not supported by the baseline model; subsample analysis does not find support for the interaction *Maintain*Product Fit* which was supported by the baseline model). Further, the confidence is higher than 90% that 8 out of 11 tested effects in subsample replicate the full sample's estimates (robustness coefficients > .90).

Web Appendix J: Removing Lower Blame Events

For this robustness analysis, all lower blame events (N = 15) that ranked below 4 on the 7-point blame scale are excluded. The findings for these analyses do not change any of the findings for the full sample analysis.

Web Appendix K: Including Control Variable "Contract Duration"

For this robustness analysis, the model is extended by a dummy variable which measures how long the endorser has had a contract with the brand. The data stem from a scan of the online data bases LexisNexis and Factiva as well as an extensive Internet search. The dummy variable indicates if the endorser has been with the brand for less (1, N = 126) or more than one year (0, N = 104). Adding this variable to the analysis (including the probit models used for PSM) does not change any of the findings for the full sample analysis.

Web Appendix A: List of Keywords to Identify Negative Celebrity Endorser Publicity

A search of major world newspapers using the databases LexisNexis and Factiva, and on the Internet identified the cases of negative celebrity endorser publicity. This search used the following keywords:

- celebrity scandal
- celebrity misbehavior
- negative celebrity publicity
- Hollywood scandal
- athlete scandal
- doping scandal
- TV scandal
- music scandal
- top celebrity scandals 1988 [up to 2016]

Web Appendix B: Sample

This study excluded 30 data points, because they would have biased the results. For instance, Phelps' pictures smoking marijuana leaked on 2/1/2009, just three trading days before Rodriguez' steroids allegations. Both athletes were sponsored by Activision Blizzard (Guitar Hero), so these two data points had to be excluded in order to assure that one event/firm reaction did not impact the other. The same rationale was applied to the other excluded cases. Table B1 lists all 128 events included in the sample.

Table B1: Sample

#	Endorser	Event	Year	Firms	Name of Firm(s) [Subsidiary Brand]
1	Alan Alda	Buying competing product in public	1988	1	IBM
2	Mike Ditka	Heart attack due to unhealthy food	1988	1	Campbell Soup
3	Mike Tyson	Domestic violence	1988	3	Kodak, Toyota, PepsiCo
4	Madonna	Scandalous music video	1989	2	PepsiCo, Warner Communications
5	Magic Johnson	Retirement due to HIV	1991	2	Dayton Hudson Corporation [Target Corporation], PepsiCo [KFC and Pepsi]
6	Dan O'Brien	Did not qualify for Olympic games	1992	4	Bausch & Lomb, Fujifilm, Reebok, Texaco
7	Dave Johnson	Won only Bronze at Olympic games	1992	1	Reebok
8	John Daly	Domestic violence	1992	1	Reebok
9	Michael Jackson	Accused of child abuse	1993	3	Disney [Captain EO], PepsiCo, Sony [Epic Label]
10	Michael Jordan	Unexpected retirement from NBA	1993	5	General Mills [Wheaties, Ball Park Franks], McDonald's, Nike, Quaker Oats Company [Gatorade], Sara Lee
11	Anna Nicole Smith	Married 89-old man	1994	1	Dress Barn [Lane Bryant]
12	Kathie Lee Gifford	Clothing line produced in sweatshop	1996	1	Wal-mart
13	Michael Irvin	Drug charges	1996	1	Nike
14	Roberto Alomar	Spit on umpire	1996	1	Reebok
15	Michael Irvin	Accused of sexual assault	1997	1	Nike
16	Chipper Jones	Extramarital affair produced child	1998	3	Coca Cola, Lowe's, Wendy's
17	Evander Holyfield	Extramarital affair produced child	1998	1	Coca Cola
18	Bill Maher	Outrages comments after 9/11	2001	3	Federal Express, Sears, Roebuck and Co.
19	Britney Spears	Buying competing product in public	2001	1	PepsiCo
20	Ludacris	Glamorizing "violence, drugs and disrespect of women"	2002	1	PepsiCo
21	Michael Jordan	Divorce after 12 years of marriage	2002	1	Nike
22	Barry Bonds	Performance enhancing drugs	2003	2	Charles Schwab, Yum! Brands [KFC]
23	Benjamin Curtis	Arrested on misdemeanor drug charge	2003	1	Dell
24	Dixie Chicks	Negative comments about President	2003	2	PepsiCo, Unilever [Lipton]
25	Kobe Bryant	Accused of sexual assault	2003	3	Coca Cola, McDonald's, Russell [Spalding]
26	Jackson/Timberlake	"Nipplegate"	2004	2	McDonald's, Sony [RCA or Jive]
27	Jason Giambi	Performance enhancing drugs	2004	1	Nike
28	Mary Kate Olsen	Entered treatment for an eating disorder	2004	1	Omnicom Group
29	Whoopi Goldberg	Negative comments about President	2004	1	Unilever [Slimfast]
30	Kanye West	"President doesn't care about black people"	2005	1	PepsiCo
31	Lance Armstrong	Performance enhancing drugs	2005	4	Bristol-Myers Squibb, Coca Cola, Nike, Oakley
32	Michael Vick	Alias "Ron Mexico" for STD testing and treatment	2005	5	AirTran Airways, Coca Cola [Powerade], Jarden [Rawling], Kraft, Nike
33	Bode Miller	Admitted being drunk while competing at Olympics	2006	1	Nike
34	Justin Gatlin	Performance enhancing drugs	2006	1	Nike
35	Nicole Richie	Driving under the influence	2006	1	Comcast [E!]
36	Britney Spears	Shaved head and got tatoos after rehab	2007	1	Sony [Jive]
37	Dog Chapman	Racial slur about his son's girlfriend	2007	1	Disney [A&E]
38	Don Imus	Insulted members of Rutgers' Women's Basketball team	2007	2	CBS, General Electric [NBC]
39	Hulk Hogan	Divorce after 24 years of marriage	2007	2	Viacom [VH1], World Wrestling Entertainment
40	Jamie Lynn Spears	Teenage pregnancy	2007	1	Viacom [Nickelodeon]
41	Michael Vick	Illegal dog fights and gambling	2007	4	AirTran Airways, EA Sports, Jarden [Rawling], Nike
42	Mischa Barton	Driving under the influence	2007	1	Wolverine World Wide [Keds]
43	Rosie O'Donnell	Unexpectedly left ABS's "The View"	2007	1	Disney [ABC]
44	Vanessa Hudgens	Nude pictures widely spread on the internet	2007	2	Disney, Johnson & Johnson [Neutrogena]
45	Brady Quinn	Homophobic remarks	2008	4	Abbott Labs [EAS Myoplex], Microsot [xBox], Nike,

Table B1 (continued): Sample

#	Endorser	Event	Year	Firms	Name of Firm(s) [Subsidiary Brand]
46	Brett Favre	Unexpected retirement from NFL	2008	2	Briggs & Stratton [Snapper Lawn Mower], VF Corporation [Wrangler]
47	Christian Bale	Domestic violence	2008	1	Time Warner [Warner Communication]
48	David Duchovny	Checked into a rehab facility for sex addiction	2008	1	Genesco Inc [Johnston & Murphy]
49	Miley Cyrus	Half naked Vanity Fair shooting at age 15	2008	1	Disney
50	Plaxico Burress	Shot himself in the leg	2008	1	Nike
51	Robert Jarvik	Medical advice without a license to practice	2008	1	Pfizer [Lipitor]
52	A-Rod	Performance enhancing drugs	2009	1	Jarden [Rawling]
53	Ben Roethlisberger	Accused of sexual assault	2009	1	Nike
54	Britney Spears	Topless photos leaked on the internet	2009	3	Elizabeth Arden, Kohl's [Candie's], Sony [Jive]
55	Charlie Sheen	Domestic violence	2009	1	HanesBrands
56	Chris Brown	Domestic violence	2009	1	Sony [Jive]
57	Cody Linley	Drunk, shirtless, out of control	2009	2	Disney, Iconix Brand Group [Ocean Pacific]
58	David Letterman	Admitted to multiple affairs on national TV	2009	1	CBS
59	Dwyane Wade	Sued his wife over STD allegations	2009	1	PepsiCo [Gatorade]
60	Jon/Kate Gosselin	Divorce after starring in TV show	2009	1	Discovery Communications [TLC]
61	Kanye West	Crashed the stage during award show	2009	1	Nike
62	Michael Phelps	Pictures surfaced showing him smoking marijuana	2009	2	Kellogg Corporation, Visa Inc
63	Miley Cyrus	Inappropriate performance during Teen Choice Awards	2009	1	Disney
64	Miley Cyrus	Wore inappropriate outfit meeting the Queen	2009	2	Disney, Wal-mart
65	Oprah Winfrey	Announced end of her show	2009	1	CBS
66	Thierry Henry	Illegal play during World Cup qualifier	2009	1	PVH Corp. [Tommy Hilfiger]
67	Tiger Woods	Sexual relations outside marriage	2009	5	Accenture, AT&T, Berkshire Hathaway [NetJets], EA
68	Ben Roethlisberger	Accused of sexual assault	2010	1	Nike
69	Brett Favre	Inappropriate texts and pictures to woman	2010	2	Briggs & Stratton [Snapper Lawn Mower], VF Corporation [Wrangler]
70	Chelsea Handler	Angelina Jolie rant on stage	2010	1	Comcast [E!]
71	Demi Lovato	Checked into a rehab facility for unknow reason	2010	1	Disney [Hollywood Records]
72	Lawrence Taylor	Rape allegations	2010	1	Nutrisystem
73	Miley Cyrus	Smoking salvia with a bong	2010	1	Disney
74	Oprah Winfrey	New Tell-All book claimed she has ''hidden' life''	2010	1	Sirius XM Holdings [XM Satellite Radio]
75	T.I.	Violated probation (marijuana)	2010	1	Unilever [Axe Body Spray]
76	Wayne Rooney	Prostitute scandal while his wife was pregnant	2010	1	Coca Cola
77	Alec Baldwin	Kicked off plane for not shutting phone off	2011	1	Capital One
78	Demi Moore	Divorce after 6 years of marriage	2011	1	Ann Taylor
79	Gilbert Gottfried	Comments about Tsunami in Japan	2011	1	Aflac
80	Jennifer Lopez	Divorce after 7 years of marriage	2011	3	Kohl's, Macy's, PepsiCo
81	Kim Kardashian	Divorce after 72 days of marriage	2011	2	Comcast [E!], Skechers
82	Ndamukong Suh	Suspended for 2 games for dirty play	2011	2	Dick's Sporting Goods, Nike
83	Rashard Mendenhall	Twitter comments about 9/11 and Bin Laden death	2011	1	HanesBrands [Champion]
84	Vanessa Hudgens	Rumors of her doing drugs in public surface	2011	1	Iconix Brand Group [Candie's]
85	Chad Ochocinco	Domestic violence	2012	2	Coca Cola [Zico Coconut Water], Viacom [VH1]
86	Charles Barkley	Talking negatively about endorser on national TV	2012	1	Weight Watchers
87	Courteney Cox	Divorce after 12 years of marriage	2012	1	Procter & Gamble [Pantene]
88	Demi Moore	Hospitalized for anorexia and substance abuse	2012	1	Ann Taylor
89	Lance Armstrong	Performance enhancing drugs	2012	3	Anheuser Busch, Luxottica [Oakley], Nike
90	Manny Pacquiao	Faced Criminal Tax Charges	2012	3	Hewlett Packard, Monster Beverage, Nike
91	Paula Deen	Diagnosed with type 2 diabetes	2012	7	Home Depot, J.C. Penney, Liberty Media [QVC], Scripps Networks Interactive [Food Network], Smithfield Foods, Target, Walgreen, Wal-mart
92	Alec Baldwin	Threatened journalist via Twitter	2013	1	Capital One
93	Beyonce	Lip Synced National Anthem at Inauguration Day	2013	2	PepsiCo, Sony
94	Kanye West	Assaulting a photographer	2013	1	Nike
95	Oscar Pistorius	Accused of killing girlfriend	2013	3	British Telecom, Luxottica [Oakley], Nike
96	Paula Deen	Admitted to using racial slurs	2013	11	Caesars Entertainment [Horseshoe Casinos], Home Depot, J.C. Penney, Liberty Media [QVC], Scripps Networks Interactive [Food Network], Novo Nordisk, Sears, Smithfield Foods, Target, Walgreen, Wal-mart
97	Adrian Peterson	Child abuse	2014	2	BP [Castrol Oil], General Mills
98	DeMarco Murray	Affair with a teammate's wife	2014	3	Foot Locker [Eastbay], PepsiCo, Verizon
99	Donald Sterling	Racist telephone recordings	2014	4	Anheuser Busch [Corona], CarMax, Lumber Liquidators,
100	Dustin Johnson	Use of cocaine	2014	1	Berkshire Hathaway [Net Jets]

Table B1 (continued): Sample

#	Endorser	Event	Year	Firms	Name of Firm(s) [Subsidiary Brand]
101	Greg Hardy	Domestic violence	2014	5	American Airlines, AT&T, Ford, Molson Coors [Miller Brewing Co], PepsiCo
102	Hope Solo	Domestic violence	2014	3	EA Sports, Nike, Pepsi Co
103	Justin Bieber	DUI	2014	1	Elizabeth Arden
104	Luis Suarez	Biting another player	2014	1	Apple [Beats by Dre]
105	Michael Phelps	DUI and suspension	2014	1	Under Armour
106	Nicki Minaj	Smashed her boyfriend's car	2014	1	Liberty Interactive [Home Shopping Network]
107	Ray Rice	Domestic violence	2014	2	Dick's Sporting Goods, EA Sports
108	P. Diddy	Brawl with Drake	2014	1	Diageo [Ciroc]
109	Tony Stewart	Accidentally killed another driver	2014	1	ExxonMobil
110	Brian Williams	Exaggerated Iraq War story	2015	1	Comcast [NBC]
111	Charlie Sheen	HIV	2015	1	Fiat Chrysler
112	Floyd Mayweather	Performance enhancing drugs	2015	1	Restaurant Brands International [Burger King]
113	Robin Thicke	Ripped off Marvin Gaye's song	2015	1	Apple [Beats by Dre]
114	Shia LaBeouf	Public intoxication	2015	1	Nike
115	Tom Brady	Deflategate	2015	7	Coca Cola [Glacéau], Deckers Brand [Ugg], Fiat Chrysler [Dodge], Fiat Chrysler [Dodge], General Mills [Wheaties], Movado, Under Armour
116	Warren Sapp	Prostitution and assault charges	2015	1	Anheuser Busch [Budweiser]
117	Brandon Marshall	Kneeling during national anthem	2016	1	CenturyLink
118	Colin Kaepernick	Sitting during national anthem	2016	1	Apple [Beats by Dre]
119	Conor McGregor	Fake retirement	2016	3	Anheuser Busch [Budweiser], EA Sports, Monster
120	Gabby Douglas	Improper posture during national anthem	2016	2	Kellogg, Procter & Gamble
121	Gigi Hadid	Mocking Melania Trump	2016	1	PVH [Calvin Klein, Tommy Hilfiger]
122	Hope Solo	Kicked off USWNT	2016	1	PepsiCo [Gatorade]
123	Kevin Durant	Move to Warriors	2016	2	BBVA, Sonic
124	Maria Sharapova	Performance enhancing drugs	2016	1	Avon
125	Michael Strahan	Left ABC to join GMA	2016	2	Dr Pepper, Yum! Brands [Pizza Hut]
126	Nick Kyrgios	Tanking a tennis match	2016	1	Apple [Beats by Dre]
127	Ryan Lochte	Faked robbery during 2016 Olympics	2016	1	Polo Ralph Lauren
128	Taylor Swift	Lyrics drama with Kanye West	2016	2	Coca Cola, Sony

Web Appendix C: Event Study Regression Model

Assuming static covariates, the use of "traditional" *CARs* is equivalent to the use of *ARs* in Event Study Regression.

Common practice in event study methodology is to aggregate daily *ARs* to some window (t_1, t_2):

$$CAR_i(t_1,t_2)=\sum_{t=t_1}^{t_2} AR_{it}$$

The average *CAR* for N events measures the average impact of the event on firm value and can be defined as:

$$ACAR\ (t_1,t_2)=\frac{1}{N}\sum_{i=1}^{N} CAR_i\ (t_1,t_2)$$

An alternative way of estimating the average impact of the event on *CARs* is to specify an *Event Study Regression* model (Beber and Pagano 2013, p. 363 based on Boehmer, Jones and Zhang 2013, p. 1372) which uses AR_{it} as dependent variable and a dummy variable D_{it} as covariate which has the value of 1 for the event window (t_1, t_2) of interest:

$AR_{it}=\alpha+\beta\cdot D_{it}+\varepsilon_{it}$ with t $\in$ [$t_1 - a$, $t_2 +$ b] and a, b $\geq$ 0

The regression coefficient α can be estimated by:

$$\hat{\alpha}=\overline{AR}_{it}-\hat{\beta}\cdot\overline{D}_{it}$$

while $\hat{\beta}$ results from:

$$\hat{\beta}=\frac{\frac{1}{N\cdot(t_2+b-t_1+a+1)}\sum_{i=1}^{N}\sum_{t=t_1-a}^{t_2+b} D_{it}\cdot AR_{it}-\overline{D}_{it}\cdot\overline{AR}_{it}}{\frac{1}{N\cdot(t_2+b-t_1+a+1)}\sum_{i=1}^{N}\sum_{t=t_1-a}^{t_2+b} D_{it}^2-\overline{D}_{it}^2}$$

Since D_{it} is 1 for $t \in [t_1 , t_2]$ and else 0, $\hat{\beta}$ can be expressed as function of *ACAR*(t_1, t_2):

$$\hat{\beta}=\frac{\frac{1}{t_2+b-t_1+a+1}\cdot\frac{1}{N}\sum_{i=1}^{N}\sum_{t=t_1}^{t_2} AR_{it}-\overline{D}_{it}\cdot\overline{AR}_{it}}{\frac{t_2-t_1+1}{t_2+b-t_1+a+1}-\overline{D}_{it}^2}=\frac{\frac{1}{t_2+b-t_1+a+1}\cdot ACAR(t_1,t_2)-\overline{D}_{it}\cdot\overline{AR}_{it}}{\frac{t_2-t_1+1}{t_2+b-t_1+a+1}-\overline{D}_{it}^2}$$

Finally, *ACAR*(t_1, t_2) can be expressed as a function of the regression coefficients:

$$ACAR(t_1,t_2)=(t_2-t_1+1)\cdot(\hat{\alpha}+\hat{\beta})$$

While the sum of α and β measures overall performance during the event window, a significant β indicates actual outperformance during the event window (t_1, t_2) relative to the whole window ($t_1 - a$, $t_2 + b$).

This relationship shows that both approaches (*CAR* and event study regression) produce the same effects. Depending on the specification of the window (t_1, t_2), the regression-based specification hence also controls for delayed effects although the dependent variable is modeled on a daily basis.

Web Appendix D: Propensity Score Matching (PSM)

In short, the PSM approach matches each individual with similar propensities to receive the treatment T = 1. The propensity score $\lambda(x) = pr(T = 1 \mid X = x)$ is a function of observed covariates X. Conditional on X respectively $\lambda(x)$, the treatment assignment is randomized. Rosenbaum and Rubin (1983) show that under unconfoundedness, independence of potential outcomes and treatment indicators also holds for conditioning solely on the propensity score $\lambda(x)$. In sum, PSM is advantageous by reducing the dimensionality of the matching problem and provides consistent estimates for the treatment effect (Wooldridge 2010). In this study, the propensity score is the probability that a firm will react to the negative endorser publicity given the value of the covariates.

Variables for the selection models. This study uses various firm (e.g., size), event (e.g., blame), and endorser (e.g., product fit) variables to create matched controls with similar propensity scores. One should only incorporate variables in the PSM model that could potentially have a direct impact on both firm response and *ARs*. Further, one should only use covariates which have been observed before the event and which could not have been affected by the firm response. For instance, endorser reaction cannot be included as covariate because it occurs after the event. This study uses variables that have been used previously by celebrity endorsement and other crisis management literature. Table D1 lists variables, operationalization, and data source.

Probit regressions. This study uses probit regression to estimate the conditional probability of being in the treatment group. Table D2 shows the results for the PSM models with firm reaction (active vs. passive) as dependent variable for the full sample excluding advertising spending (N = 230, model (1)) and the subsample including it (N = 206, model (2)). The subsample is smaller because Kantar Media's data start in 1995 (as opposed to 1988). This study also includes a PSM model with changes in advertising spending (i.e., whether the firm reduced advertising spending or not after the negative event) as dependent variable (model (3)), because a firm's spending decision could also be affected by a self-selection bias. McFadden Pseudo-R^2 are .223 (full sample), .233 (subsample), and .112 (advertising reduction). As expected, the fit of the advertising reduction model is worse because unlike the decisions for firm response, decisions concerning advertising spending largely depend on factors that are unrelated to the negative event. Concerning the full sample, firms are less likely to respond if a non-response is deemed appropriate (-.014, $p < .01$), if the event turns out to be a rumor (-.788, $p < .01$), and if a female endorser is involved (-.883, $p <$ 01). The latter finding is surprising but can be explained by the fact that female endorsers are less likely to be involved in high blame events ($F_{Blame} = 27.8$, $p < .001$). Firms are more likely to respond if the firm and celebrity have received a lot of media attention during the year prior to the event (.258, $p < .01$).

Table D1: Variables for the PSM Models

Variable	Operationalization	Source
Appropriateness of Non-Response$_e$	Perceived appropriateness of a firm's decision not to respond to event *e* (-100 = not at all, +100 = very); metric; mean centered	SSI
Blame$_e$	Perceived blame of the endorser for event *e* (1 = not at all, 7 = very much); metric; mean centered	SSI
Brand$_{i*e}$	Whether the endorsing firm *i* is a subsidiary (1) or the parent firm (0) during event *e*; binary	Coded
Cumulative Abnormal Returns (-10, -1)$_{i*e}$,	Sum of actual minus expected stock return for firm *i* during the period of 10 days before event; metric; mean centered	Compustat
Crime$_e$	Whether the negative publicity of event *e* is a criminal act (1) or not (0); binary	Coded
Disagreement about Appropriateness of Non-response$_e$	Standard deviation of perceived appropriateness of a firm's decision not to respond to event *e*; metric; mean centered	SSI
Event Year$_e$	Year of the event *e*; metric; mean-centered	Coded
Gender$_e$	Whether the endorser involved in event *e* is female (1) or male (0); binary	Coded
Industry$_{i*e}$	Whether the endorsing firm *i* is part of one of five different industries during event *e*; categorical	Coded
Leverage$_{i*e}$	Total Liabilities / Total Assets for firm *i* prior event *e* (1 year); metric; mean centered	Compustat
Market-to-Book Ratio$_{i*e}$	Book value of a firm *i* divided by its market value prior event *e* (1 year); metric; mean centered	Compustat
Number of Firms Endorsed$_e$	Whether the endorser involved in event *e* is endorsing more than one firm (1) or not (0); binary	Coded
Occupation$_e$	Whether the negative publicity of event *e* is related to the occupation of the endorser (1) or not (0); binary	Coded
Prior Year Advertising Spending$_{i*e}$	Ln(ad spending of firm *i* during the year before the event *e*); metric; mean centered	Kantar
Priot Year Media Coverage: Celebrity$_e$	Ln(Sum of Associated Press (AP) articles mentioning the celebrity during the year before the event *e*); metric; mean centered	Factiva
Priot Year Media Coverage: Celebrity + Firm$_{i*e}$	Ln(Sum of Associated Press (AP) articles mentioning the celebrity and firm *i* during the year before the event *e*); metric; mean centered	Factiva
Priot Year Media Coverage: Firm$_{i*e}$	Ln(Sum of Associated Press (AP) articles mentioning firm *i* during the year before the event *e*); metric; mean centered	Factiva
Product Fit$_{i*e}$	Perceived fit between endorser involved in event *e* and product of firm *i*; metric; mean centered	SSI
Return on Asset$_{i*e}$	Net income divided by total assets for firm *i* prior event *e* (1 year); metric; mean centered	Compustat
Rumor$_e$	Whether the negative publicity of event *e* was a rumor (1) or fact (0); binary	Coded
Sales$_{i*e}$	Logarithm of unit sales for firm *i* prior event *e* (1 year); metric; mean centered	Compustat
Scandal History$_e$	Whether the endorser involved in event *e* has caused negative publicity before (1) or not (0); binary	Coded
Type$_e$	Whether the endorser involved in event *e* is an athlete, musician, or TV/radio personality; categorical	Coded

Table D2: Log Estimates for Propensity Score Matching

		Full sample (N = 230)		Subsample (N = 206)			
	Dependent Variable	***Firm Reaction***		***Firm Reaction***		***Ad Spending Reduction***	
		(1)		**(2)**		**(3)**	
Independent Variables		Coeff.	SE	Coeff.	SE	Coeff.	SE
	Appropriateness of Non-response	-.014**	.006	-.015**	.006	.001	.006
	Disagreement about Appropriateness of Non-response	.027$	.014	.029$	.015	-.012	.015
Event Details	*Blame*	-.116	.159	-.039	.177	-.037	.174
	Crime	-.279	.334	-.307	.339	-.109	.310
	Occupation	-.399	.259	-.524$	.272	.005	.253
	Rumor	-.788**	.354	-.661$	.385	.337	.374
Endorser Details	*Scandal History*	.301	.207	.159	.220	-.166	.204
	# of Firms Endorsed	.170	.291	-.160	.330	-.584$	.312
	Gender (female=1; male=0)	-.883**	.307	-.738**	.312	-.630**	.300
	Type (Baseline: Sports): Music	.194	.397	.047	.441	.113	.398
	Type (Baseline: Sports): TV/Radio	1.138**	.337	1.056**	.350	-.400	.341
	Product Fit	-.003	.004	-.004	.004	.004	.004
	Prior Year Media Coverage: Celebrity	-.026	.086	.011	.099	.053	.088
Firm Details	*Pr. Year Media Cov.: Celebrity + Firm*	.258**	.124	.208$	.126	-.018	.125
	Prior Year Media Coverage: Firm	-.046	.070	.032	.086	.090	.081
	Return on Assets	1.208	1.256	1.135	1.286	-.328	1.371
	Market-to-Book Ratio	-.015	.017	-.013	.015	.007	.011
	ln(Sales)	.000	.088	-.064	.095	-.039	.099
	Leverage	.108	.576	-.186	.589	-.533	.596
	CAR(-10,-1)	.027	.017	.019	.018	.002	.017
	Brand	-.045	.202	-.042	.218	-.104	.207
	Prior Year Advertisng Spending			-.025	.060	.052	.064
	Industry dummies	YES		YES		YES	
Period	*Event Year*	-.062***	.015	-.028	.030	.033	.029
Intercept		.677	.512	1.129	.893	.877	.919
Model Fit	Log Likelihood	-120.582		-104.412		-126.749	
	Wald χ^2	73.838		68.465		31.823	
	Pseudo R^2	.223		.233		.112	
	N	230		206		206	

Note: *** p < .001, ** p < .01, * p < .05, $ p < .10 (based on heteroscedasticity robust standard errors)

Matching observations and testing the balancing properties. Using above probit regression results, this study applies both (i) the Biweight[1] Kernel estimator and (ii) one-to-one estimator to match each treatment observation with the control observations that were weighted (Kernel estimator) or selected (one-to-one estimator) by propensity score similarity (here: similar probability that a firm will react to the celebrity misbehavior given the value of the covariates). Kernel matching is a non-parametric technique that uses weighted averages of all individuals in the control group to construct the counterfactual outcome (Heckman, Ichimura and Todd 1998). This method is more efficient than one-to-one matching, because it uses more information, but has the drawback that potentially more bad matches are used. Hence, this study also uses one-to-one matching. This study furthermore disregards any control observation that falls outside the smallest connected area of common support. The analysis applies the Gaussian Kernel density estimator to obtain smooth densities of the propensity scores for treatment and control units and to identify the area of common support.

To provide evidence that matching was successful and balancing properties are fulfilled, this study (1) reports graphs of Kernel density estimates of propensity scores for treatment and control units before and after matching (Figure D1) and (2) discusses t-tests and matching metrics such as Rubin's B and Rubin's R comparing covariates' means and variances between treatment and control groups (Table D3). In the following, the reporting focuses on the findings for the full sample.[2] Figure D1 (Panel A) shows the propensity score distributions for treatment (TG – active firms) and control group (CG – passive firms) before the matching procedure. This study disregards any treatment or control observation that fell outside the smallest connected area of common support (Heckman et al. 1998). Figure D1 (Panel B) illustrates the kernel density estimates of propensity scores $\lambda(x)$ for TG and CG after matching. The trimmed propensity score distributions of TGs and CGs are virtually identical. After matching, probit regression produces a fairly low Pseudo-R^2 (.029) and insignificant χ^2-value (6.381; p =1.000), and t-tests comparing covariates' means between treatment and control groups are insignificant (p > .10; see Table D3). Further, Rubin's R (ratio of treated to (matched) non-treated variances of the propensity score index) is 1.00 (should be close to 1). Rubin's B (absolute standardized difference of the means of the linear index of the propensity score in the treated and (matched) non-treated group) improved from 119.0 to 40.7, which indicates that matching reduced bias substantially (more than 50%). Matching TG with a similar CG enables us to avoid bias in the estimated treatment effect that may occur when linking a potentially dissimilar CG to a TG (Smith 1997).

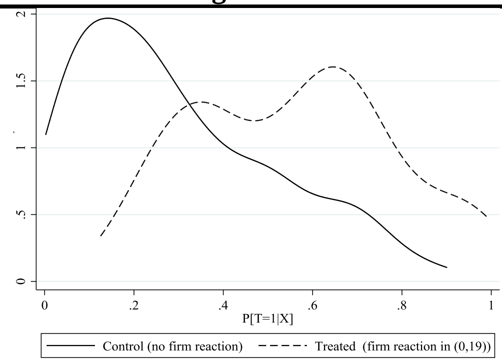
Finally, this study uses the individual observation weights w_{i*e} from this analysis step in the regression models (see Table 3 in the main manuscript). This re-weighting of treatment and control units allows drawing inferences (Morgan and Harding 2006) and is particularly useful in combining PSM with other methods such as fixed effects regression (Nichols 2007).

[1] Results are largely robust using other Kernel estimators (e.g., Gaussian Kernel) to match units. Results are available on request.

[2] Graphs of Kernel densities before and after matching and the other balancing property tests for the other PSM models (models (2) and (3) in Table D3) are available upon request. In sum, results suggest that their balancing properties are largely fulfilled.

Figure D1: Propensity Score Distribution Before and After Matching

Panel A: Before Matching



Panel B: After Matching

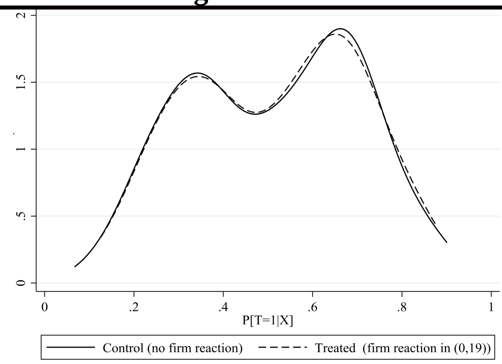


Table D3: Test of Balancing Properties after Matching (Full Sample with N = 230)

Covariates		Treat-ment	Control	t-test	p-value	% Bias	% |Bias| Red.
	Appropriateness of Non-response	-1.551	-3.468	.370	.709	5.7	65.3
	Disagreement about Appropriateness of Non-response	.369	.148	.190	.851	2.9	74.4
Event Details	*Blame*	.020	.119	-.620	.535	-9.6	2.0
	Crime	.274	.321	-.660	.508	-10.4	-102.8
	Occupation	.357	.319	.520	.602	7.8	51.9
	Rumor	.071	.065	.150	.878	2.2	80.2
Endorser Details	*Scandal History*	.512	.477	.440	.658	6.9	46.2
	# of Firms Endorsed	.786	.799	-.210	.836	-3.4	88.4
	Gender	1.286	1.263	.330	.744	5.0	64.1
	Type (Baseline: Sports): Music	.143	.152	-.170	.863	-2.5	70.3
	Type (Baseline: Sports): TV/Radio	.298	.237	.890	.374	14.0	62.2
	Product Fit	-4.788	-2.312	-.530	.600	-8.4	63.9
	Prior Year Media Coverage: Celebrity	.022	.112	-.350	.724	-5.5	57.0
Firm Details	*Prior Year Media Coverage: Celebrity + Firm*	.097	.138	-.210	.831	-3.6	87.6
	Prior Year Media Coverage: Firm	-.038	-.073	.100	.918	1.6	83.6
	Return on Assets	.005	.002	.250	.803	3.4	1.3
	Market-to-Book Ratio	-.376	-.569	.250	.804	2.5	77.7
	ln(Sales)	.006	-.188	.880	.382	13.3	-473.0
	Leverage	.010	-.021	1.080	.282	16.2	-101.1
	CAR(-10,-1)	.547	.929	-.520	.606	-6.8	63.2
	Brand	.321	.271	.710	.479	10.6	-145.4
	Industry 3	.167	.207	-.660	.510	-9.5	77.6
	Industry 4	.214	.203	.190	.852	3.1	91.5
	Industry 5	.190	.180	.180	.858	2.9	77.2
	Industry 6	.107	.084	.510	.614	7.6	-82.6
Period	*Event Year*	-1.148	-1.995	.710	.476	11.7	71.0
Overall Criteria	Likelihood ratio χ^2-value			6.830			
	p-value			1.000			
	Pseudo R^2			.029			
	Mean Bias			6.800			
	% Mean Bias Reduction			59.0			
	Median Bias			6.200			
	% Median Bias Reduction			51.2			
	Rubin's B			40.70			
	Rubin's R			1.00			

Web Appendix E: Sensitivity Analysis of Selection on Unobservables

Unobserved heterogeneity can bias the results. Therefore, this study discusses two different methods (Rosenbaum Bound Analysis and Instrumental Variables) to test how robust the findings are with regard to the influence of unobservable variables. These analyses reveal that it seems unlikely the findings are driven by unobserved factors.

Rosenbaum Bounds Analysis

Although PSM can eliminate the bias from observable variables, results may not be robust to the hidden bias from unobservable variables. The Rosenbaum Bounds analysis (DiPrete and Gangl 2004; Rosenbaum 2002) enables us to estimate how sensitive the treatment effect is to the influence of unobservable variables assuming they exert the same influence on the treatment and outcome variables (e.g., Manchanda, Packard and Pattabhiramaiah 2015; Sun and Zhu 2013). Rosenbaum Bounds analysis varies the parameter Γ, which represents the level of hidden bias as measured by the odds ratio of differential treatment assignment due to unobservable variables, to estimate the sensitivity of treatment effects. The underlying propensity score models for the observable self-selection bias are specified according to Web Appendix D. This analysis focuses the hypothetical *p*-values for treatment effects assuming positive self-selection into treatment status for different levels of Γ. Since the treatment effect for $\Gamma = 1$ represents a conservative estimate for negative self-selection, the sensitivity in case of positive self-selection is of primary interest (DiPrete and Gangl 2004). This study also presents Hodges-Lehmann point estimates and confidence intervals for effect sizes at different values of Γ.

For this analysis, the difference $\Delta AR_{post\text{-}pre}$ between the daily averages of *ARs* after and before the firm responds serves as input. For firms without response, this difference equals the average of daily *ARs* for event window (0, 20). It is important to note that this study is able to control for constant firm and event-specific unobserved variables because it measures the difference between pre and post firm response stock returns (diff-in-diff design). Hence, the Rosenbaum Bounds analysis gives an estimate concerning the sensitivity of the treatment effect with regard to the influence of unobserved time-varying variables on the estimated relationship between treatment and outcome.

Table E1 reports the results for the full sample and Kernel matching with common support (a) including and (b) excluding outliers (the results for any other model specification are available upon request).[3] For the full sample including outliers, the treatment effect becomes insignificant at $\Gamma > 1.2$ ($\alpha =$ 5%). As the mean propensity score for firm response is .514, this study would not find significant evidence for a treatment effect if hidden time-varying bias increased the propensity score of firm response above .514*1.2 = .617. Excluding outliers, the mean propensity score of firm response is .462 and the critical Γ is about 1.7. The treatment effect would therefore become insignificant if the hidden time varying bias increased the propensity score of firm response above .462*1.7 = .785. Note, the time varying hidden bias not only had to affect treatment assignment but also the outcome in a similar manner. Therefore, reported bounds are the worst-case scenarios. This analysis suggests that the results are robust with regards to unobserved heterogeneity.

[3] This study identifes outliers as observations for which Cook's D > 4/230 in the regression $\Delta AR_{post\text{-}pre}$ = f(*Firm Response*).

Table E1: Result for the Rosenbaum Bounds Analysis

Gamma (Γ)	**p-Value[a]** (U-Bound)	**p-Value[a]** (L-Bound)	**H-L Point Estimate** (U-Bound)	**H-L Point Estimate** (L-Bound)	**Confidence Interval[a]** (U-Bound)	**Confidence Interval[a]** (L-Bound)
Including Outliers (N = 230)						
1.0	.011	.011	.155	.155	.028	.296
1.1	.027	.004	.128	.182	-.003	.335
1.2	.056	.001	.107	.204	-.033	.367
1.3	.101	.000	.084	.223	-.057	.390
1.4	.160	.000	.067	.243	-.079	.409
1.5	.233	.000	.054	.260	-.097	.432
1.6	.315	.000	.036	.282	-.116	.455
1.7	.401	.000	.022	.310	-.133	.482
1.8	.487	.000	.003	.329	-.152	.512
1.9	.569	.000	-.017	.348	-.179	.530
2.0	.644	.000	-.028	.365	-.196	.553
Excluding Outliers (N = 210)						
1.0	.001	.001	.178	.178	.072	.296
1.1	.002	.000	.166	.200	.055	.318
1.2	.005	.000	.145	.215	.040	.335
1.3	.010	.000	.133	.231	.024	.359
1.4	.018	.000	.116	.249	.010	.370
1.5	.031	.000	.106	.265	-.008	.388
1.6	.048	.000	.095	.277	-.018	.400
1.7	.071	.000	.084	.289	-.027	.418
1.8	.099	.000	.071	.297	-.039	.434
1.9	.131	.000	.062	.306	-.051	.448
2.0	.169	.000	.052	.319	-.060	.453

Note: H-L, Hodges-Lehmann.

[a] *p*-values and confidence intervals are one-sided and at the 95% level.

Instrumental Variables

To account for unobserved heterogeneity, management scholars frequently use the instrumental variables estimation technique (e.g., Chintagunta, Dubé and Goh 2005; Kos, Li and Pang 2017; Semadeni, Withers and Certo 2014). The advantage of this technique is that it excludes the influence of unobservable variables and provides an unbiased estimate for the endogenous variable. However, the challenge is to identify appropriate instruments for the endogenous variable which satisfy two basic conditions. First, the instrument must be a good predictor of the endogenous variable. Second, the instrument must be exogenous, i.e., theoretically uncorrelated with the outcome. Ideally, this study had some private information about the events which has a bearing on firm response respectively response timing, but is unknown to investors (e.g., private information about contractual regulations between firms and endorsers could be such information). Unfortunately, such information is not available for most cases in our context.

One feasible approach in the context of our study is to exploit the time series nature of the data and to use lagged endogenous variables as instruments (e.g., Adebambo and Yan 2017; Hutton, Jiang and Kumar 2015). Our date set contains 116 firms that were involved in more than one event (e.g., Nike is involved in 23 events). Hence, this study uses information about firms' response to prior events to make predictions about firms' response to the current event. For instance, if prior response timing is a good predictor of current response timing and if prior response timing is exogenous to the current event's stock return, then it is a valid instrumental variable.

Nevertheless, the validity of this instrument could be affected by the existence of an omitted variable, which is not only correlated with firms' response to prior events but also with stock returns to the current event. However, the existence of such an omitted variable is bound to two restrictions:

First, prior and current events are usually not comparable, because they involve different endorsers and/or types of misbehavior. Hence, if such an omitted variable exists, it is most likely not related to the endorser or type of misbehavior but rather to firm characteristics, which, in addition, must be stable over time.

Second, if the omitted variable affects prior event's response timing and current event's stock returns simultaneously, it is unreasonable to assume that investors ignored this omitted variable during the prior event. However, if this omitted variable affects both prior and current event's stock returns, then the stock market would behave very inefficiently, because it would not have been able to include information about this omitted variable into the stock price since it became first available during the prior event. Instead, (part of) of the information about the omitted variable would have been carried over to the current event, which would violate the Efficient Market Hypothesis (Fama 1970). This violation is not impossible but unlikely.

Even in the worst-case scenario, which assumes the existence of such an omitted variable, using prior event's response timing as instrumental variable still creates important insights about the value of response timing as early indicator: since the instrument is observed before the current event, we can rule out reverse causality (stock return → response timing).

We apply two-stage least squares estimate to predict *CAR*(0,20) with firm response timing. This analysis uses *Prior Response Timing* as instrument, which is the firm's day of response to the last event before the current event (note, for firms without response this analysis imputes day 21 as response day for both *Prior Response Timing* and *Current Response Timing*). The first-stage regression is specified as *Current Response Timing* = $\beta_0 + \beta_1 \cdot$ *Prior Response Timing* + ε. The second-stage regression is specified as *CAR*(0,20) = $\beta_0 + \beta_1 \cdot$ *Predicted Current Response Timing* + ε, where *Predicted Current Response Timing* is the obtained from the first-stage regression.

This study estimates this two-stage model twice: once with all 116 observations and once with 112 observations excluding outliers which fall outside the 95% confidence interval (for two observations *CAR*(0,20) is above 15% and for two observations *CAR*(0,20) is below -15%). Table E2 below shows the results.

First, *Prior Response Timing* is a reasonably good predictor of *Current Response Timing*: R^2 is about 9.5% and the F-value is above the common threshold of 10 (Staiger and Stock 1997; Stock and Yogo 2005). This finding seems plausible because if firms are able and motivated to respond to negative endorser publicity, then firms should decide to do so not only once but repeatedly over time. Second, the instrumented *Current Response Timing* variable is negatively related to *CAR*(0,20) (Panel B: -.449 with $p < .10$ including outliers; -.447 with $p < .05$ excluding outliers), which suggests that the longer the firm waits with its response the more negative the stock market reacts.

The results of this instrumental variable analysis are in line with the results of the numerous other analyses presented in this study. As indicated previously, the results also rule out reverse causality, because the instrument is observed before the outcome. This instrumental variables analysis further suggests that it seems unlikely for the results to be fully driven by unobserved heterogeneity, because beyond the restrictions about the nature of the omitted variable presented above, the omitted variable would have to be unrelated to the many observed control variables incorporated in the other analyses of this study (see main manuscript).

Table E2: Results for Instrumental Variables Regression

Panel A: Stage one regression	***DV: Current Response Timing (day of response)***	
	Including outliers	Excluding outliers
Prior Response Timing (day of response)	.302** (.088)	.302** (.089)
Intercept	9.761*** (1.477)	9.762*** (1.479)
N	116	112
F	11.94***	11.64***
R^2	.095	.096

Panel B: Stage two regression	***DV: CAR(0,20)***	
	Including outliers	Excluding outliers
Current Response Timing (day of response)	-.449$^{\$}$ (.247)	-.447* (.216)
Intercept	6.727$^{\$}$ (3.544)	6.698* (3.072)
N	116	112
Wald Chi^2	3.30$^{\$}$	4.30*

Note: *** $p < .001$, ** $p < .01$, * $p < .05$, $^{\$}$ $p < .10$ (standard errors in parenthesis)

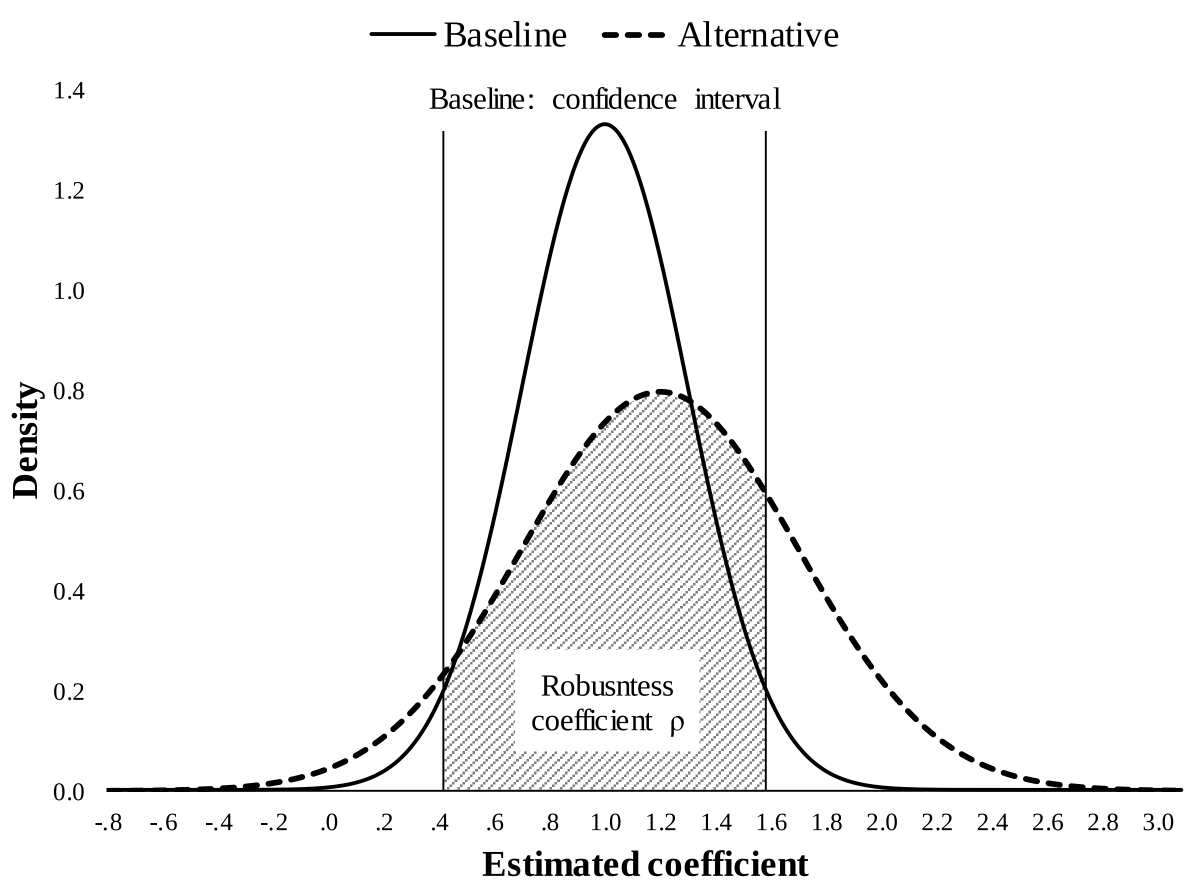
Web Appendix F: Robustness Coefficient

Neumayer and Plümper (2017) operationalize the robustness coefficient ρ as the degree to which the probability density function of the alternative model's estimate $\hat{\beta}_{alt}$ overlaps with the baseline model's confidence interval $[\hat{\beta}_{base} - C\hat{\sigma}_{base};\ \hat{\beta}_{base} + C\hat{\sigma}_{base}]$:

$$\rho(\hat{\beta}_{alt}) = \frac{1}{\hat{\sigma}_{alt}\sqrt{2\pi}} \int_{\hat{\beta}_{base} - C\hat{\sigma}_{base}}^{\hat{\beta}_{base} + C\hat{\sigma}_{base}} e^{-(x-\hat{\beta}_{alt})^2 / C\hat{\sigma}_{alt}^2} dx$$

C corresponds with the critical value of interest. This study lets *C* = 1.96 to estimate the 95% confidence interval. The parameters $\hat{\sigma}_{alt}$ and $\hat{\sigma}_{base}$ are the square root of the standard errors of the two coefficient estimates. Figure F1 below provides a graphical representation of how the robustness coefficient ρ is defined. In this example, the baseline model's coeffcient β_{base} = 1.0 and its SE_{base} = .3. The alternative model's coefficient β_{alt} = 1.2 and its SE_{alt} = .5. The resulting robustness coefficient is ρ_{alt} = .724 (shaded area).

Figure F1: Example for Robustness Coefficient *ρ*



Web Appendix G: Generalized PSM to Adjust for a Firm's Decision When to React

PSM is limited to binary treatment variables (i.e., firm response: yes, no), but the decision about response timing is non-binary (i.e., firm response on day 0, …, 20) and might be susceptible to self-selection bias as well. Hirano and Imbens (2004) propose a generalization of the binary treatment propensity score, which they label generalized propensity score (GPS). Adjusting for GPS removes biases associated with differences in the observed covariates. This study uses GPSM to correct for observable self-selection bias concerning the exact response day and to divide active firms into slow and fast responders. As this propensity score approach for continuous treatments found less consideration in business literature to date, this Web Appendix provides more technical details about this procedure (see also Bia and Mattei 2008):

Technical details. To ease interpretation of timing treatment effects, this study assigns the values $T = 0,\ldots, 20$ for those firms who react on day 0, …, 20. Firms can also choose not to respond within this time frame. For these cases, $T = 21$. The first step of the GPS approach involves the estimation of the GPS r which is the conditional density of treatment t given covariates x:

$$r(t, x) = f_{T|X}(t|x)$$

Conditional on the covariates, this study assumes that the timing treatment (or its transformation) is normally distributed:

$$g(T)|X \sim N\{h(\gamma, X), \sigma^2\}$$

The analysis models $g(\cdot)$ as the log function to account for non-linearity. It estimates parameters γ and σ^2 by maximum likelihood and test the normal distribution assumption with the Kolmogorov-Smirnov test. To test the balancing property of the propensity score, this study follows Hirano and Imbens (2004) and divides all observations into several timing treatment intervals with comparable sample sizes for responding firms, namely (0) with N = 33, (1,3) with N = 27, (4,20) with N = 35 and (20+) with N = 135. For each of the covariates X and for each timing treatment interval group, this analysis tests whether the covariate's group mean differs from the means of other groups combined if it adjusts for the estimated GPS. The results include the Bayes factors to evaluate the magnitude of evidence in favor or against the balancing property (Jeffreys 1961).[4]

The second step estimates the conditional expectation of the outcome Y (i.e., $CAR(0,20)$) as a function of the timing treatment level T and the GPS R:

$$\beta(t, r) = E[Y|T = t, R = r]$$

This study uses the following linear function to estimate the conditional expectation of Y:

$$E[Y|T = t, R = r] = \psi(T, R, \alpha) = (1, T, R, T * R) \cdot \alpha$$

[4] Bayes-factor method is a Bayesian alternative to classical hypothesis testing. Instead of all-or-nothing interpretation of results, Bayes-factor methods provide a scale for interpretation and allow comparing hypothesis pair H_0 and H_1 (strength of evidence). The Bayes factor is computed by calculating the effect of the sample data on the prior odds $P(H_0)/P(H_1)$ and thus measures the ratio of posterior odds and prior odds. For technical details refer to Greene (2002).

There is no direct meaning of the coefficients α in this model, except that the coefficients involving R test whether the covariates introduce any bias (Hirano and Imbens 2004). Unlike the binary case, this regression does not allow for causal inference.

The third and final step involves the estimation of the so-called dose-response function (DRF), which is the expectation of R conditional X at a certain level of t:

$$\mu(t) = E[t, r(t, X)]$$

Using results of steps one and two, the average potential outcome for each level of the timing treatment is computed as:

$$E\{\hat{Y}(t)\} = \frac{1}{N}\sum_{i=1}^{N} \hat{\beta}\{t, \hat{r}(t, X)\}, \qquad with\ \hat{\beta}\{t, \hat{r}(t, X)\} = \hat{\psi}\{t, \hat{r}(t, X), \hat{\alpha}\}$$

This function also provides estimates for unobserved treatment levels t. The DRF is less instructive in terms of economic interpretation (Hirano and Imbens 2004). To examine the effect of interest, results report the timing treatment effect function, which is the derivative of the DRF and defined as the difference in expected outcomes for two dose levels. By letting $c = 1$, it shows how *CAR*s would change if the firm delayed its reaction by one day. This analysis uses bootstrapping to estimate confidence intervals for the estimated coefficients.

The GPS approach is advantageous because it provides a selection-bias adjusted estimate for the optimal timing of firm response, which can easily be interpreted visually. One limitation is that by aggregation of *AR*s to *CAR*s, it neglects potential differences in pre- and post-treatment. With this respect, the two approaches (Event Study Regression and GPS) complement each other.

Findings. Table G1 shows the results for the maximum likelihood estimated self-selection model using day of firm reaction as dependent variable. A positive (negative) coefficient indicates that it increases the likelihood of a later (earlier) firm reaction. R^2 is .31 and indicates a moderate explanatory power. A Kolmogorov-Smirnov test assesses the validity of the assumed normal distribution model for the treatment conditional on the covariates. The analysis does not find evidence against this assumption, because the results cannot reject the null hypothesis at a 95% confidence level ($p = .084$).

Next, this analysis inspects the tests of the balancing property (see Table G2). As indicated earlier, this analysis divides all observations into four timing treatment intervals and assesses - after adjustment for the GPS - the extent to which each covariate is balanced, i.e., the probability that $T = t$ does not depend on the value of X. This test examines for each of the 27 covariates whether the mean in one of the four TGs was different from the mean in the other three TGs combined (108 comparisons). In only two cases there is strong evidence against the balancing property (Industry #3 for treatment group 1, appropriateness of non-response for treatment group 2). The conclusion is that the adjustment for the GPS creates by large a balanced sample. Table G3 reports the regression results (there is no direct meaning of estimated coefficients) for the DRF, which is used to estimate the timing treatment effect.

Figure G1 reports the resulting treatment effect function. The graph reveals that firms would receive significantly more negative *ARs* if they delayed their response from day 0 to 1 (-1.396%, $p < .05$), from day 1 to 2 (-.981%, $p < .05$), or day 2 to day 3 (-.630%, $p < .05$), but not after that (*p's* > .10). Therefore, this study divides firms into quick (day 0, 1, 2) and slow (after day 2) responders.

Table G1: Self-Selection Model

		Full Sample (N = 230)	
	Dependent Variable	**ln(Response Timing)**	
Independent Variables		Coeff.	SE
	Appropriateness of Non-response	.009$	.005
	Disagreement about Appropriateness of Non-response	-.015	.010
Event Details	*Blame*	.178	.139
	Crime	.170	.261
	Occupation	.180	.207
	Rumor	.720**	.219
Endorser Details	*Scandal History*	-.135	.163
	# of Firms Endorsed	-.025	.238
	Gender (1 = female; 0 = male)	.569**	.204
	Type (Baseline: Music): Sports	-.055	.315
	Type (Baseline: Music): TV/Radio	-.764**	.260
	Product Fit	.002	.003
	Prior Year Media Coverage: Celebrity	.060	.071
Firm Details	*Prior Year Media Coverage: Celebrity + Firm*	-.351**	.106
	Prior Year Media Coverage: Firm	.052	.052
	Return on Assets	-.776	1.012
	Market-to-Book Ratio	.009	.008
	ln(Sales)	.017	.065
	Leverage	-.437	.460
	CAR(-10,-1)	-.013	.013
	Brand	-.043	.155
	Industry dummies	YES	
Period	*Event Year*	.045**	.013
Intercept		.034	.464
Model Fit	F-value	5.270***	
	R^2	.307	
	Kolmogorov-Smirnov test p-value	.098 (corrected: .084)	
	N	230	

Note: *** p < .001, ** p < .01, * p < .05, $ p < .10 (based on heteroscedasticity robust standard errors)

Table G2: Testing the Balancing Property

		Timing Treatment Group 1 [Response on Day 0]				Timing Treatment Group 2 [Response on Days 1 to 3]				Timing Treatment Group 3 [Response on Days 4 to 20]				Timing Treatment Group 4 [Response on Day 20+]			
Covariates		**Mean**	**SD**	**Bayes-Factor**	**p-value**	**Mean**	**SD**	**Bayes-Factor**	**p-value**	**Mean**	**SD**	**Bayes-Factor**	**p-value**	**Mean**	**SD**	**Bayes-Factor**	**p-value**
	Appropriateness of Non-response	-18.527	8.500	.400	.029	20.209	6.195	.025	.001	14.409	7.749	.692	.063	-14.821	5.468	.168	.007
	Disagreement about Appropr. of Non-resp.	1.613	2.009	3.083	.422	.013	1.479	3.957	.993	-.954	1.756	3.372	.587	-.644	1.255	5.285	.608
Event Details	*Blame*	.469	.258	.822	.070	-.428	.187	.297	.022	-.594	.236	.172	.012	.353	.158	.509	.026
	Crime	.167	.120	1.611	.164	-.177	.087	.515	.043	-.141	.104	1.558	.177	.146	.076	.971	.056
	Occupation	.016	.128	4.239	.899	.057	.093	3.272	.543	.242	.113	.398	.032	-.126	.085	1.980	.138
	Rumor	.060	.078	3.159	.440	.094	.058	1.042	.104	-.133	.067	.538	.046	.064	.053	2.929	.233
Endorser Details	*Scandal History*	.304	.129	.266	.018	-.162	.095	.906	.087	-.191	.114	.968	.096	.001	.085	6.053	.990
	# of Firms Endorsed	.068	.083	3.055	.416	-.095	.063	1.258	.133	.130	.084	1.171	.122	-.047	.056	4.245	.406
	Gender	-.180	.120	1.387	.135	.215	.088	.208	.014	-.013	.105	3.889	.899	-.128	.080	1.670	.111
	Type (Baseline: Music): Sports	.135	.124	2.347	.276	-.192	.089	.399	.032	.102	.113	2.597	.369	.024	.081	5.787	.768
	Type (Base: Music): TV/Radio	-.122	.104	2.131	.240	.132	.074	.803	.075	-.102	.097	2.242	.295	.030	.061	5.338	.621
	Product Fit	-10.559	7.659	1.640	.168	13.036	5.557	.260	.019	-4.059	6.755	3.261	.548	1.106	4.918	5.898	.822
	Pr. Year Media Cov.: Celebrity	-.563	.415	1.686	.174	.311	.297	2.259	.295	-.010	.359	3.920	.977	.144	.276	5.258	.601
Firm Details	*Pr. Y. Med. Cov.: Celeb. + Firm*	-.438	.215	.540	.042	.194	.164	1.935	.237	.390	.258	1.237	.131	-.026	.152	5.964	.865
	Pr. Y. Med. Cov.: Firm	-.013	.536	4.273	.981	.545	.391	1.472	.163	-.847	.471	.774	.072	.427	.358	2.927	.233
	Return on Assets	.001	.021	4.269	.960	-.001	.016	3.951	.959	-.008	.020	3.632	.699	.008	.015	5.185	.583
	Market-to-Book Ratio	.621	2.302	4.119	.787	.784	1.673	3.534	.639	.389	1.972	3.844	.844	-.381	1.354	5.812	.779
	ln(Sales)	.329	.399	3.024	.409	.081	.299	3.811	.787	-.299	.349	2.694	.391	.060	.273	5.906	.827
	Leverage	-.006	.049	4.240	.899	.009	.035	3.825	.798	.017	.044	3.643	.704	-.020	.035	5.096	.563
	CAR(-10,-1)	-1.768	1.482	2.082	.233	.198	1.088	3.889	.855	-1.426	1.291	2.109	.269	.882	1.051	4.220	.402
	Brand	-.072	.124	3.603	.562	.045	.091	3.483	.619	.061	.110	3.346	.577	-.062	.082	4.523	.451
	Industry 3	5.176	1.626	.030	.001	-2.524	1.135	.340	.026	1.520	1.624	2.511	.349	-1.500	1.023	2.033	.142
	Industry 4	-.118	.118	2.572	.316	.087	.086	2.346	.312	.078	.104	2.939	.452	-.043	.079	5.204	.588
	Industry 5	-.063	.082	3.193	.448	-.016	.058	3.799	.779	.008	.078	3.898	.915	.021	.050	5.538	.677
	Industry 6	.148	.095	1.263	.120	-.020	.071	3.804	.782	-.142	.084	.906	.088	.008	.060	6.001	.896
Period	*Event Year*	-.074	.073	2.518	.306	-.054	.056	2.422	.328	.085	.071	1.902	.233	-.025	.051	5.329	.619

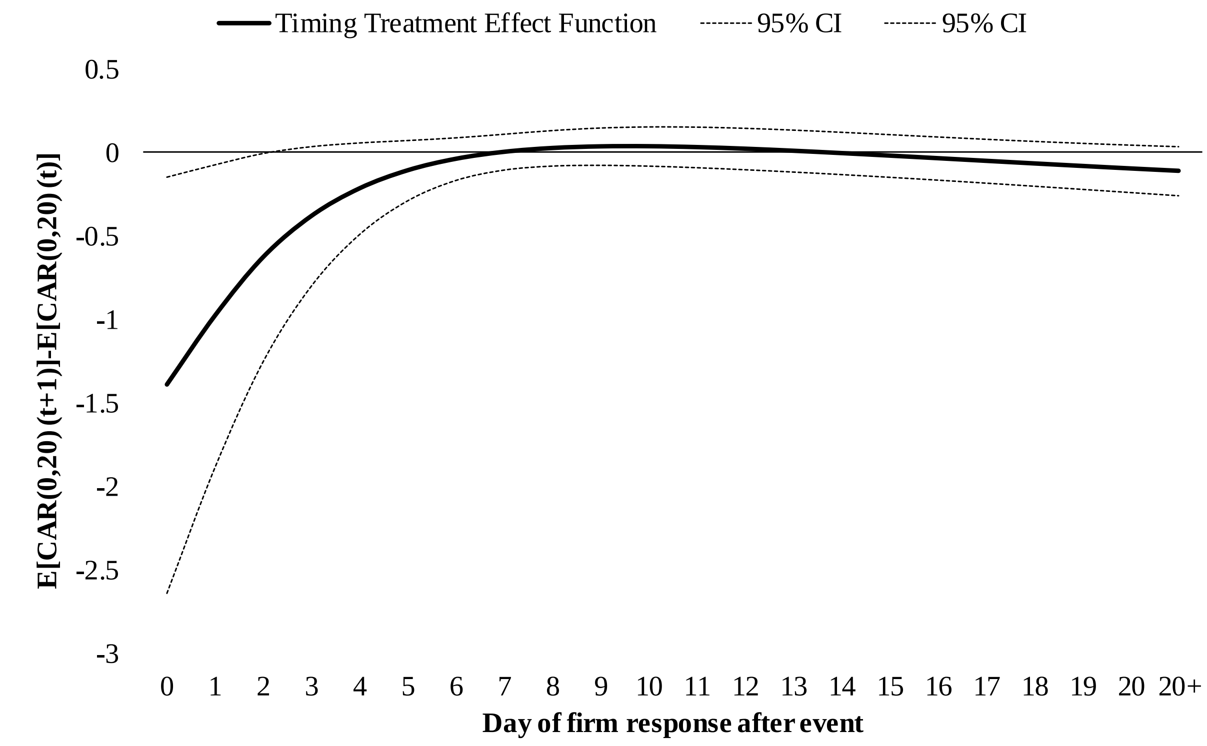
Note: This analysis applies the order of magnitude interpretations of the Bayes factor (BF) as proposed by Jeffreys (1961). Evidence for the balancing property (BP): BF > 1.00 (Evidence supports the BP); .32 < BF < 1.00 (Very slight evidence against the BP); .10 < BF < .32 (Moderate evidence against the BP); .01 < BF < .10 (Strong to very strong evidence against the BP); BF < .01 (Decisive evidence against the BP)

Table G3: Dose Response Function

	Full sample (N = 230)	
Dependent Variable	**CAR(0,20)**	
Independent Variables	Coeff.	SE
Response Timing	-.330**	.139
GPS	-15.526**	7.389
*Response Timing*GPS*	1.065**	.465
Intercept	4.604**	2.135
Model Fit F-value	2.010	
R^2	.032	
N	230	

Note: *** p < .001, ** p < .01, * p < .05, $ p < .10 (based on heteroscedasticity robust standard errors)

Figure G1: Timing Treatment Effect Function



Web Appendix H: Results of Regression Models Used for Robustness Analysis

The following Tables H1 to H6 present the estimates and sample-size adjusted standard errors used for the robustness analyses. See Table 3 in the manuscript for model specifications. For parsimonious reasons, control variables and model fit criteria are not displayed but available upon request.

Table H1: Estimates and Sample-size Adjusted Standard Errors for Alternative Models 1-16 testing Firm Response and Timing

Model # (see Table 3 main manuscript)			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Firm Response	(+)	β	.681	.676	.663	.662	.573	.597	.544	.497	.605	.703	.619	.705	.841	.633	.547	.652
		σ	.263	.234	.283	.257	.160	.141	.184	.162	.249	.240	.265	.255	.286	.270	.176	.153
*Firm Response*Timing (slow vs. fast)*	(-)	β	-.184	-.176	-.169	-.150	-.178	-.180	-.158	-.132	-.235	-.220	-.221	-.189	-.186	-.159	-.203	-.200
		σ	.087	.083	.086	.083	.061	.060	.068	.063	.095	.091	.090	.089	.082	.084	.076	.070

Table H2: Estimates and Sample-size Adjusted Standard Errors for Alternative Models 17-32 testing Firm Response and Timing

Model # (see Table 3 main manuscript)			17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Firm Response	(+)	β	.557	.563	.472	.477	.595	.657	.662	.673	.666	.668	.544	.630	.574	.532	.502	.571
		σ	.178	.169	.222	.194	.225	.219	.238	.230	.282	.256	.159	.140	.163	.152	.181	.164
*Firm Response*Timing (slow vs. fast)*	(-)	β	-.199	-.160	-.130	-.125	-.178	-.163	-.191	-.160	-.178	-.154	-.176	-.173	-.174	-.133	-.162	-.154
		σ	.073	.069	.073	.074	.085	.082	.082	.081	.085	.083	.068	.062	.064	.060	.068	.064

Table H3: Estimates and Sample-size Adjusted Standard Errors for Alternative Models 1-16 testing Suspend and Maintain

Model # (see Table 3 main manuscript)			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Firm Response: Suspend	(+)	β	.756	.598	.783	.582	.598	.547	.608	.440	.673	.601	.783	.614	1.071	.568	.572	.501
		σ	.309	.272	.318	.281	.184	.167	.208	.181	.298	.270	.293	.274	.319	.291	.207	.185
Firm Response: Maintain	(+)	β	.564	.771	.419	.786	.466	.664	.397	.613	.527	.827	.365	.839	.393	.729	.487	.705
		σ	.297	.246	.321	.283	.227	.150	.235	.168	.272	.254	.295	.281	.314	.298	.221	.159

Table H4: Estimates and Sample-size Adjusted Standard Errors for Alternative Models 17-32 testing Suspend and Maintain

Model # (see Table 3 main manuscript)			17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Firm Response: Suspend	(+)	β	.641	.498	.632	.364	.658	.555	.821	.584	.771	.606	.567	.478	.657	.468	.562	.445
		σ	.213	.181	.248	.216	.266	.246	.265	.246	.311	.276	.185	.169	.195	.163	.202	.177
Firm Response: Maintain	(+)	β	.403	.638	.154	.430	.516	.780	.415	.806	.420	.771	.487	.683	.433	.605	.349	.575
		σ	.217	.179	.230	.166	.251	.233	.267	.254	.331	.293	.205	.145	.198	.161	.239	.164

Table H5: Estimates and Sample-size Adjusted Standard Errors for Alternative Models 1-16 testing Moderators

Model # (see Table 3 main manuscript)			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
*Blame*Suspend*	(+)	β	.875	.736	.682	.500	.709	.611	.653	.543	.598	.497	.438	.273	.464	.052	.662	.579
		σ	.302	.325	.250	.262	.155	.177	.121	.167	.260	.284	.185	.199	.417	.525	.152	.175
*Blame*Maintain*	(-)	β	-.116	.059	.007	.260	-.152	-.102	-.153	.030	-.025	.122	.055	.245	.007	.388	-.135	-.031
		σ	.144	.172	.144	.191	.125	.125	.123	.132	.135	.168	.119	.160	.197	.200	.110	.116
*Occupation*Suspend*	(+)	β	.863	.971	.612	.899	.987	1.010	.810	.939	.394	.484	.377	.574	.296	1.028	.951	.948
		σ	.417	.446	.356	.412	.247	.260	.215	.256	.337	.378	.305	.354	.327	.381	.254	.278
*Occupation*Maintain*	(-)	β	-.662	-.335	-.698	-.475	-.594	-.480	-.666	-.646	-.604	-.343	-.581	-.457	-.641	-.691	-.521	-.408
		σ	.287	.348	.299	.320	.206	.213	.258	.212	.299	.349	.291	.316	.384	.314	.203	.217
*Product Fit*Suspend*	(-)	β	-.319	-.290	-.385	-.323	-.269	-.205	-.289	-.194	-.246	-.200	-.318	-.217	-.240	-.295	-.261	-.220
		σ	.141	.133	.131	.123	.072	.071	.070	.065	.118	.105	.108	.094	.122	.124	.067	.073
*Product Fit*Maintain*	(+)	β	.351	.372	.240	.346	.406	.411	.284	.335	.385	.433	.246	.356	.129	.327	.470	.414
		σ	.176	.207	.196	.219	.107	.106	.135	.117	.175	.203	.189	.208	.183	.206	.104	.102
*E. Apology*Suspend*	(-)	β	-.459	-.490	-.483	-.455	-.590	-.631	-.624	-.572	-.459	-.505	-.474	-.449	-.446	-.400	-.650	-.631
		σ	.229	.190	.223	.179	.119	.116	.114	.123	.253	.182	.248	.172	.221	.191	.158	.124
*E. Apology*Maintain*	(+)	β	.016	.594	-.080	.614	-.114	.207	-.093	.271	-.128	.575	-.195	.608	.081	.582	-.113	.370
		σ	.419	.411	.408	.381	.306	.243	.268	.196	.406	.387	.378	.355	.420	.346	.258	.206

Table H6: Estimates and Sample-size Adjusted Standard Errors for Alternative Models 17-32 testing Moderators

Model # (see Table 3 main manuscript)			17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
*Blame*Suspend*	(+)	β	.604	.474	.576	.283	.609	.508	.445	.285	.663	.494	.681	.598	.619	.495	.647	.535
		σ	.106	.143	.214	.313	.246	.269	.174	.187	.247	.256	.143	.165	.099	.134	.116	.164
*Blame*Maintain*	(-)	β	-.150	.052	-.083	.168	-.040	.109	.056	.248	.078	.319	-.148	-.041	-.154	.049	-.091	.129
		σ	.101	.110	.111	.112	.129	.160	.110	.150	.171	.195	.105	.111	.093	.103	.101	.097
*Occupation*Suspend*	(+)	β	.959	.944	.631	1.175	.389	.494	.372	.576	.652	.890	.946	.950	.953	.944	.782	.941
		σ	.216	.264	.250	.291	.306	.347	.274	.319	.361	.406	.229	.254	.195	.238	.202	.254
*Occupation*Maintain*	(-)	β	-.658	-.589	-.736	-.765	-.591	-.331	-.585	-.456	-.849	-.636	-.513	-.401	-.659	-.589	-.854	-.852
		σ	.225	.205	.222	.194	.274	.320	.263	.283	.344	.303	.185	.198	.205	.183	.235	.189
*Product Fit*Suspend*	(-)	β	-.298	-.169	-.193	-.191	-.240	-.199	-.316	-.225	-.409	-.328	-.257	-.219	-.293	-.176	-.312	-.189
		σ	.065	.063	.073	.080	.108	.097	.100	.087	.133	.123	.061	.067	.060	.058	.068	.065
*Product Fit*Maintain*	(+)	β	.257	.369	.140	.280	.372	.426	.247	.359	.247	.356	.460	.409	.256	.371	.268	.299
		σ	.126	.113	.099	.115	.158	.183	.175	.190	.194	.218	.094	.091	.119	.103	.120	.107
*E. Apology*Suspend*	(-)	β	-.716	-.621	-.591	-.525	-.498	-.501	-.476	-.446	-.471	-.445	-.662	-.629	-.715	-.618	-.614	-.567
		σ	.146	.113	.125	.126	.231	.167	.219	.155	.223	.177	.126	.114	.112	.102	.112	.122
*E. Apology*Maintain*	(+)	β	-.187	.309	.022	.323	-.153	.586	-.197	.612	.034	.673	-.115	.377	-.183	.314	-.092	.368
		σ	.252	.200	.324	.202	.373	.355	.341	.320	.446	.421	.230	.188	.223	.180	.266	.148

Web Appendix I: Analyses of Same Event with Multiple Firms and Different Reactions

One could argue that a firm's reaction depends on the reaction of competing firms employing the same celebrity. Thus, instead of considering the reacting in isolation, this study also considers the reaction relative to other firms using the same celebrity. The dataset contains 13 events (N = 51 observations) where a celebrity was endorsing more than one firm and at least one of the firms showed a diverging response to the same negative event (response vs. no response, slow vs. quick response, maintaining vs. suspending response). This subsample is consequently homogeneous concerning any unmeasured endorser and event characteristics.

However, this subsample analysis has several drawbacks. First, the subsample size is much smaller compared to baseline model (Table 6 in main manuscript). There are only Nsub = 1,071 (13 events, 51 firms, 21 event days) vs. Nbase = 4,380 (128 events, 230 firms, 21 event days) observations. Hence, even if the coefficient estimates were identical, the standard errors in the subsample are more likely much wider. This analysis therefore adjusts standard errors of the subsample by employing the baseline model's sample size. It furthermore calculates the robustness coefficient to evaluate the confidence that the subsample's and baseline model's estimates are identical (for technical details, see Web Appendix F). Second, variation in many of the variables is much smaller. For example, in the subsample, the data do not include a single firm suspending an endorser while the misbehavior has been related to the occupation of the endorser. Hence, this analysis cannot test this effect. Third, the influence of outliers concerning the control variables can be much more severe while correlations among focal variables and control variables create many multicollinearity concerns. Therefore, this analysis excludes any controls except for firm fixed effects, but adds event day fixed effects to control at least for general event day dynamics. Further, to exploit between firm variance and to get a more efficient estimation, this analysis estimates the panel model with random effects instead of fixed effects (Hausman-test confirms with $p > .10$).

Table I1 below shows the results. At a 95% confidence level, results confirm the positive effect of firm response, the negative effect of slow response, and the positive effect of both suspending and maintaining the endorser. Concerning the moderation effect of *Blame*, the analysis finds in the subsample at a 90% confidence level that the interaction *Suspend*Blame* has a positive coefficient. Similar to the baseline model, it does not find support for the interaction *Maintain*Blame*. As already indicated, this analysis cannot test the interaction *Suspend*Occupation* due to missing information in the subsample, but it can confirm the negative interaction effect *Maintain*Occupation* ($p < .10$). It finds support for the negative interaction of *Suspend*Product Fit* ($p < .05$), but the positive interaction *Maintain*Product Fit* cannot be confirmed. Further, it finds support for the negative interaction of *Suspend*Endorser Apology* ($p < .05$) and the positive interaction of *Maintain*Endorser Apology* ($p < .05$). The latter finding is interesting because in the baseline model, the results cannot confirm the interaction of *Maintain*Suspend* ($p > .10$). Another case is also noteworthy: the subsample provides support for the interaction of *Blame*Suspend* ($p < .10$), but the estimate (.215) is much smaller compared to the baseline model (.875). Accordingly, the robustness coefficient is relatively small (.281). This finding implies that the confidence is relatively strong that this effect holds, but the confidence about the true population parameter is much weaker. In sum, the subsample analysis suggests that most of the findings reported so far are confirmed with two exceptions (support for the interaction *Maintain*Endorser Apology*, which was not supported by the baseline model; no support for the interaction *Maintain*Product Fit*, which was supported by the baseline model). Concerning robustness, this analysis provides more than 90% confidence that 8 out of 11 tested effects in subsample replicate the full sample's estimates (robustness coefficients > .90).

Table I1: Results of Analyses of Same Event with Multiple Firms and Different Reactions

Effect		Subsample Est.	Subsample SE	Subsample p	Baseline model (full sample) Est.	Baseline model (full sample) SE	Baseline model (full sample) p	Robustness ρ	Robust support for significant effect
Firm Response (yes vs. no)	+	.572	.194	.003	.681	.263	.011	.981	✓[b]
Response Timing (slow vs. fast)	-	-.239	.077	.002	-.184	.087	.036	.932	✓[c]
Response Type: Suspend	+	.661	.191	.001	.756	.309	.016	.996	✓[b]
Response Type: Maintain	+	.548	.199	.006	.564	.297	.060	.997	✓[b]
*Suspend*Blame*	+	.215	.118	.069	.875	.302	.004	.281	✓[d]
*Maintain*Blame*	-	.037	.034	.276	-.116	.144	.421	1.000	✘[b]
*Suspend*Occupation*	+	--[a]	--[a]	--[a]	.863	.417	.040	--[a]	--[a]
*Maintain*Occupation*	-	-.194	.108	.073	-.662	.287	.022	.808	✓
*Suspend*Product Fit*	-	-.121	.035	.001	-.319	.141	.026	.988	✓[b]
*Maintain*Product Fit*	+	.000	.073	.999	.351	.176	.048	.467	✘[d]
*Suspend*Endorser Apology*	-	-.314	.090	.001	-.459	.229	.047	1.000	✓[b]
*Maintain*Endorser Apology*	+	.437	.104	.000	.016	.419	.969	1.000	✘[b]

Note: Standard error of sub-sample adjusted for sample size difference. Robustness ρ measures the average (minimum) share of the probability density distributions of the robustness test estimated effects that lie within the 95% confidence interval of the baseline model effect (Neumayer and Plümper 2017). ρ ranges from 0 (lack of robustness) to 1 (strong robustness).

[a] Cannot be estimated for the subsample, because there is not a single firm that suspends their endorser when the misbehavior is related to her/his occupation.
[b] Coefficient estimate of subsample provides support for coefficient estimate of full sample (>95% confidence).
[c] Coefficient estimate of subsample provides support for coefficient estimate of full sample (>90% significance).
[d] Coefficient estimate of subsample does not provide support for coefficient estimate of full sample (<50% confidence).

Web Appendix J: Removing Lower Blame Events

For this robustness analysis, all lower blame events (N = 15) that ranked below 4 on the 7-point blame scale are excluded. The findings for these analyses do not change any of the findings for the full sample analysis (see Table J1 below).

Table J1: Results of Analyses after Removing Lower Blame Events

Effect		**Percentage of *β* with expected sign** total	**Percentage of *β* with expected sign** p < .05	$\beta_\varnothing$	$SE_\varnothing$	$p_\varnothing$	**Robustness** $\rho_\varnothing$	**Robustness** ρ_{min}	**Robust support for significant effect**
Firm Response (yes vs. no)	+	100.0%	93.8%	.568	.211	.007	.971	.906	✓
Timing (slow vs. fast)	-	100.0%	65.6%	-.158	.075	.035	.960	.899	✓
Response Type: Suspend	+	100.0%	93.8%	.578	.232	.013	.959	.804	✓
Response Type: Maintain	+	100.0%	5.0%	.539	.244	.027	.947	.837	✓
*Suspend*Blame*	+	100.0%	81.3%	.472	.165	.004	.873	.515	✓
*Maintain*Blame*	-	37.5%	.0%	.077	.171	.652	.768	.142	✘
*Suspend*Occupation*	+	100.0%	65.6%	.790	.290	.006	.953	.701	✓
*Maintain*Occupation*	-	100.0%	62.5%	-.590	.284	.038	.918	.658	✓
*Suspend*Product Fit*	-	100.0%	96.9%	-.260	.086	.003	.974	.907	✓
*Maintain*Product Fit*	+	100.0%	5.0%	.334	.161	.038	.941	.611	✓
*Suspend*Endorser Apology*	-	100.0%	87.5%	-.533	.157	.001	.970	.814	✓
*Maintain*Endorser Apology*	+	59.4%	9.4%	.196	.311	.528	.876	.519	✘

Note: Average coefficients, average standard errors, overall p-value, average and minimum robustness ρ based on estimates of baseline model M1 and alternative models M2 to M32. $p_\varnothing$ is based on the t-value (~N(0;1)) calculated by the ratio of the unweighted average of regression coefficients and standard errors of M1 to M32 (Sala-i-Martin 1997). Standard errors of M1 to M32 are adjusted for sample size differences. Average (minimum) robustness ρ measures the average (minimum) share of the probability density distributions of the robustness test estimated effects that lie within the 95% CI of the baseline model effect (Neumayer and Plümper 2017). ρ ranges from 0 (lack of robustness) to 1 (strong robustness).

Web Appendix K: Including Control Variable "Contract Duration"

For this robustness analysis, the model is extended by a dummy variable which measures how long the endorser has had a contract with the brand. The data stem from a scan of the online data bases LexisNexis and Factiva as well as an extensive Internet search. The dummy variable indicates if the endorser has been with the brand for less (1, N = 126) or more than one year (0, N = 104). Adding this variable to the analysis (including the probit models used for PSM) does not change any of the findings for the full sample analysis (see Table K1 below).

Table K1: Results of Analyses Including Control Variable "Contract Duration"

Effect		Percentage of β with expected sign: total	Percentage of β with expected sign: p < .05	$\beta_\varnothing$	$SE_\varnothing$	$p_\varnothing$	Robustness: $\rho_\varnothing$	Robustness: ρ_{min}	Robust support for significant effect
Firm Response (yes vs. no)	+	100%	100%	.632	.216	.003	.979	.933	✓
Timing (slow vs. fast)	-	100%	84.4%	-.181	.076	.017	.961	.858	✓
Response Type: Suspend	+	100%	93.8%	.628	.242	.009	.971	.860	✓
Response Type: Maintain	+	100%	71.9%	.585	.241	.015	.958	.861	✓
*Suspend*Blame*	+	100%	78.1%	.576	.202	.004	.883	.476	✓
*Maintain*Blame*	-	53.1%	0.0%	.023	.142	.869	.754	.129	✗
*Suspend*Occupation*	+	100%	68.8%	.790	.296	.008	.965	.861	✓
*Maintain*Occupation*	-	100%	68.8%	-.585	.261	.025	.929	.743	✓
*Suspend*Product Fit*	-	100%	93.8%	-.252	.090	.005	.978	.934	✓
*Maintain*Product Fit*	+	100%	65.6%	.355	.152	.019	.944	.743	✓
*Suspend*Endorser Apology*	-	100%	90.6%	-.532	.163	.001	.981	.925	✓
*Maintain*Endorser Apology*	+	62.5%	3.1%	.211	.308	.492	.910	.478	✗

Note: Average coefficients, average standard errors, overall p-value, average and minimum robustness ρ based on estimates of baseline model M1 and alternative models M2 to M32. $p_\varnothing$ is based on the t-value (~N(0;1)) calculated by the ratio of the unweighted average of regression coefficients and standard errors of M1 to M32 (Sala-i-Martin 1997). Standard errors of M1 to M32 are adjusted for sample size differences. Average (minimum) robustness ρ measures the average (minimum) share of the probability density distributions of the robustness test estimated effects that lie within the 95% CI of the baseline model effect (Neumayer and Plümper 2017). ρ ranges from 0 (lack of robustness) to 1 (strong robustness).

In the baseline regression model, the variable *Contract Length Dummy*Firm Response* is not significantly different form zero. The estimate ranges between .037 (SE = .269) and .091 (SE = .277) and its p-value is above 10%.

There are several measurement issues associated with the contract duration variable. For instance, one celebrity endorsed Pepsi in 2012. The same endorser already appeared in a Pepsi commercial back in 2005. However, there is no information if the endorser was an official Pepsi spokesperson during the entire 2005-12 span or if the celebrity just appeared sporadically throughout the years. Hence, it is not possible to identify the exact start and/or end date of the contract. The data contain multiple such cases where it is impossible to determine the exact contract duration. Based on the dates of old pictures or videos showing the endorser together with the brand and its products, assumption about existing endorsement contracts are possible, but do not allow for an exact contract duration measure.

WEB APPENDIX REFERENCES

Adebambo, B. N. and X. (S.) Yan (2017), "Investor Overconfidence, Firm Valuation, and Corporate Decisions," *Management Science*, forthcoming.

Beber, A. and M. Pagano (2013), "Short-Selling Bans Around the World: Evidence from the 2007–09 Crisis," *The Journal of Finance*, 68 (1), 343-81.

Bia, M. and A. Mattei (2008), "A Stata Package for the Estimation of the Dose–Response Function through Adjustment for the Generalized Propensity Score," *The Stata Journal*, 8 (3), 354-73.

Boehmer, E., Jones, C. M., and X. Zhang (2013), "Shackling Short Sellers: The 2008 Shorting Ban," *Review of Financial Studies*, 26 (6), 1363-1400.

Chintagunta, P., Dubé, J.-P., and K. Y. Goh (2005), "Beyond the Endogeneity Bias: The Effect of Unmeasured Brand Characteristics on Household-Level Brand Choice Models," *Management Science*, 51 (5): 832-849.

DiPrete, T. A. and M. Gangl (2004), "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments," *Sociological Methodology*, 34 (1), 271-310.

Fama, E. F. (1970), "Efficient Capital Markets, A Review of Theory and Empirical Work," *Journal of Finance*, 25 (2), 383-417.

Greene, W. H. (2002), "*Econometric Analysis*", Englewood Cliffs, NJ: Prentice Hall.

Heckman, J. J., Ichimura, H., and P. Todd (1998), "Matching as an Econometric Evaluation Estimator," *The Review of Economic Studies*, 65 (2), 261-94.

Hirano, K. and G. Imbens (2004), "*The Propensity Score with Continuous Treatment*," In: Gelman, A., Meng, X.L. (Eds.), Applied Bayesian Modelling and Causal Inference from Missing Data Perspectives, Wiley, New York, 73-84.

Hutton, I., Jiang, D., and A. Kumar (2015); "Political Values, Culture, and Corporate Litigation," *Management Science*, 61 (12), 2905-2925.

Jeffreys, H. (1961), "*Theory of Probability*," 3rd edition, Oxford: Oxford University Press.

Kos, J., Li, Y., and J. Pang (2017), "Does Corporate Governance Matter More for High Financial Slack Firms?" *Management Science*, 63 (6):1872-1891.

Manchanda, P., Packard, G., and A. Pattabhiramaiah (2015), "Social Dollars: The Economic Impact of Customer Participation in a Firm-Sponsored Online Customer Community," *Marketing Science*, 34 (3), 367-87.

Morgan, S. L. and D. J. Harding (2006), "Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice," *Sociological Methods and Research*, 35 (1), 3-60.

Neumayer, E. and T. Plümper (2017), "*Robustness Tests for Quantitative Research,*" Cambridge University Press.

Nichols, A. (2007), "Causal Inference with Observational Data," *Stata Journal*, 7 (4), 507.

Rosenbaum P. R. (2002), "*Observational Studies,*" In: Observational Studies. Springer Series in Statistics. Springer, New York, 1-17.

Rosenbaum, P. R. and D. B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70 (1), 41-55.

Semadeni M., Withers, M. C., and S. T. Certo (2014), " The Perils of Endogeneity and Instrumental Variables in Strategy Research: Understanding through Simulations," *Strategic Management Journal*, 35 (2014), 1070-1079.

Smith, H. L. (1997), "Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies," *Sociological Methodology*, 27 (1), 325-53.

Staiger, D. and J. H. Stock (1997), "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557-586.

Stock, J. H. and M. Yogo (2005), "Testing for Weak Instruments in Linear IV Regression," in J.H. Stock and D.W.K. Andrews (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge, MA: Cambridge University Press, 80-108.

Sun, M. and F. Zhu (2013), "Ad Revenue and Content Commercialization: Evidence from Blogs," *Management Science*, 59 (10), 2314-31.

Wooldridge, J. M. (2010), "*Econometric Analysis of Cross Section and Panel Data*," Cambridge, MA: MIT Press.