# Social Psychology

**Editor-in-Chief**
Christian Unkelbach

HOGREFE

# Social Psychology

HOGREFE

# The Effects of Counterfactual Attacks on Social Judgments

Patrizia Catellani and Mauro Bertolotti

Department of Psychology, Catholic University of Milan, Italy

**Abstract.** Two experiments were conducted to compare the effects of different styles of verbal criticism (factual vs. counterfactual) on the perceptions of target, source, and quality of the attack. Counterfactual attacks resulted in more negative overall judgment of the target and ratings of the target's morality than either factual attacks or no attack. Counterfactual attacks were also rated more positively than factual attacks, and the source of the counterfactual attack was rated as being less biased against the target. Regression analyses confirmed that the observed effect on overall judgment was mediated by the perceived bias of the source. The greater effectiveness of counterfactual attacks was moderated by awareness of prior hostility of the source of the attack toward the target.

Keywords: counterfactual communication, criticism, morality, person perception

Criticism of someone's past actions does not always result in a negative impression of the criticized person but it may induce a negative impression of the criticizing person instead (Carraro, Gawronski, & Castelli, 2010; Gawronski & Walther, 2008; Wyer, Swan, & Gruenfeld, 1995). One way of avoiding this backlash effect and to increase the effectiveness of criticism may be to formulate it in a subtle and indirect way rather than an open and direct way. For instance, instead of openly criticizing one person stating that s/he acted badly, one could resort to more indirect criticism, saying that things would have been better if the person had behaved differently. This last is an example of a counterfactual statement. Counterfactuals are mental simulations of how a given event might have had a different outcome if one or more antecedents of the event had been different (e.g., "If the driver had been more careful, the accident wouldn't have happened"). The generation of counterfactuals influences the way people explain past events (Kahneman & Tversky, 1982; Roese, 1997) and attribute responsibility for those events (Wells & Gavanski, 1989). When people focus counterfactuals on how an actor of an event could have behaved in a different way, they are likely to blame the same actor for the actual outcome of the event (Markman & Tetlock, 2000; Nario-Redmond & Branscombe, 1996). One may therefore presume that counterfactuals might be effectively employed in communication to attack other people. However, while research on counterfactual thinking is well established, research on counterfactual communication has been limited so far (but see Bertolotti, Catellani, Douglas, & Sutton, 2013; Catellani & Bertolotti, 2014; Catellani, Bertolotti, & Covelli, 2013; Tal-Or, Boninger, Poran, & Gleicher, 2004; Wong, 2010).

In the present research, we investigated the effects of counterfactual attacks on social judgments. In two studies, we compared the effects of factual and counterfactual attacks on people's attitudes toward the target and the source of the attack. We expected that, compared with factual attacks, counterfactual attacks would lead to a more negative evaluation of the target, and less backlash toward the source of the attack. They would also be perceived as more appropriate than factual attacks.

## Attacks and Their Effects

Attack messages provide individuals who hear or read them (observers) with negative information about the target of the attack, particularly their behavior. Given that negative information is generally more salient than positive information – the *negativity effect* (Cacioppo, Gardner, & Berntson, 1997; Fiske, 1980; Ito, Larsen, Smith, & Cacioppo, 1998; Peeters, 1971; Peeters & Czapinski, 1990) – one would expect attacks to readily induce negative evaluations of the target of the criticism. However, this is often not the case, because attempts to criticize a person may result in negative evaluations of the source, rather than the target of the criticism – the *backlash effect* (Budesheim, Houston, & DePaola, 1996; Carraro et al., 2010; Haddock & Zanna, 1997; Hill, 1989; Matthews & Dietz-Uhler, 1998; Roese & Sande, 1993; Skowronski, Carlston, Mae, & Crawford, 1998). This effect derives from two simultaneous processes. On one hand, observers may develop a negative impression of the source of the criticism, who may be perceived as

aggressive, unfair, and driven by ulterior motives (Douglas & Sutton, 2003; Hornsey, 2005). This is particularly evident in the case of intergroup communication when criticism comes from an outgroup member (Hornsey & Imani, 2004; Hornsey, Oppes, & Svensson, 2002; Rabinovich & Morton, 2010). On the other hand, observers may discount negative information about the target of the criticism if they consider the source to be biased and, therefore, untrustworthy (Douglas & Sutton, 2010; Esposo, Hornsey, & Spoor, 2013).

A way of avoiding the backlash effect may be the recourse to subtle communication strategies. For example, mixing criticism with praise of the target or self-criticism reduces the perceived negativity of the message, preventing recipients from rejecting it (Hornsey, Robson, Smith, Esposo, & Sutton, 2008). Other subtle communicative strategies that are often employed by people include politeness (Brown & Levinson, 1987) and equivocation (Bavelas, Black, Bryson, & Mullett, 1988). However, so far the effectiveness of subtle communication strategies in conveying negative information about the target without inducing a negative reaction against the source has not been systematically investigated. In the present research, we explored the possibility that one specific form of subtle communication, namely presenting criticism as a counterfactual rather than factual statement, would effectively achieve the communication goal of attacking a target without being penalized by the backlash effect.

## Counterfactual Attacks

Counterfactuals are mental simulations of alternatives to actual scenarios or events, in which one or more antecedents of a past event is postulated to have changed in order to hypothetically alter the outcome of the event (Kahneman & Tversky, 1982; Roese, 1997). The antecedents that are more likely to trigger counterfactual thoughts are those deviating from a norm. The norm can be intrapersonal (Kahneman & Miller, 1986), such as a routine course of action (e.g., "If I had taken the usual route home, I would not have had a car accident"), or social (Catellani, Alberici, & Milesi, 2004; Catellani & Milesi, 2005), such as stereotypical expectations about individuals or groups and their behavior (e.g., "If the woman had not taken a lift from a stranger, she would not have been raped").

The fact that certain antecedents of an event are more likely to be focused on in counterfactuals leads people to: (a) overestimate the importance of these antecedents as causes of the event; and (b) ignore or undervalue other possible antecedents that contributed to the event (McClure, Hilton, & Sutton, 2007). Counterfactuals are therefore involved in the explanations of events and related attributions of responsibility (Markman & Tetlock, 2000; Nario-Redmond & Branscombe, 1996; Wells & Gavanski, 1989). In particular the target of an upward counterfactual, which focuses on how things might have been better, is more likely to be blamed for a negative event compared with other actors involved in the event (Markman &

Tetlock, 2000; Nario-Redmond & Branscombe, 1996; Wells & Gavanski, 1989). This effect is even stronger when counterfactuals focus on the controllable, rather than uncontrollable, behaviors of an actor (Alicke, Buckingham, Zell, & Davis, 2008; Girotto, Legrenzi, & Rizzo, 1991).

Considering the above, one can hypothesize that upward counterfactuals that focus on the controllable behaviors of a target can be used effectively as an attack. A counterfactual attack could be more effective than a factual attack, for two main reasons. First, whereas a factual attack places the focus on the question whether a person behaved in a wrong way, a counterfactual attack takes the wrong behavior for granted, and places the focus on what might have happened if the person had acted differently. Thus, compared with a factual attack, a counterfactual attack shifts the focus on the opportunity the person had to act in a better way, which may increase the degree of blame attributed to the person. Second, owing to its indirect formulation, a counterfactual attack raises less suspicion on the communicative intent of the source than a factual attack, which reduces the likelihood of backlash. Previous research has indeed shown that resistance to persuasion, and corresponding negative reactions against the source, are lower when linguistic and rhetorical strategies are used to conceal persuasive intentions (Fiedler, 2008; Hornsey et al., 2008).

## Factual and Counterfactual Attacks on Morality

Our investigation of the effects of factual and counterfactual attacks focuses on attacks related to moral issues, given the crucial role that morality plays in social perception (Abele & Bruckmüller, 2011; Vonk 1999; Wojciszke & Abele, 2008; Ybarra et al., 2008). People are more sensitive to negative information about morality than they are to negative information about other dimensions of personality (Brambilla, Rusconi, Sacchi, & Cherubini, 2011; Kervyn, Dolderer, Mahieu, & Yzerbyt, 2010; Skowronski, 2002; Wojciszke, Brycz, & Borkenau, 1993). A cognitive explanation for this phenomenon is that negative information about morality is more diagnostic of underlying personal dispositions (Reeder & Brewer, 1979; Skowronski & Carlston, 1987, 1989). Even a single case of immoral behavior is considered to be indicative of someone's morality, because such behavior is considered to be under the control of the actor (Reeder & Brewer, 1979; Skowronski, 2002; Wojciszke, 2005)

Given the importance attributed to negative information about morality, one's moral character should be easily damaged by attacks focusing on it. On the contrary, some studies have found that this is not always the case, partly because there is a backlash against the source of the attack, as discussed above (Carraro et al., 2010; Funk, 1996). The attacks used in those studies were, however, formulated in a direct way. Attacks formulated in an indirect way may be less susceptible to backlash, as their source may not be perceived as being biased. Thus, in the present study we

hypothesized that counterfactual attacks would be more effective than factual attacks in lowering the perceived morality of the attacked target.

# The Present Research

The present research tested the effectiveness of counterfactual attacks about morality in a political context. This context was used for two reasons. First, politics is naturally contentious (Bull & Wells, 2012; Catellani & Covelli, 2013). Second, the personal image of politicians, including their morality, is often scrutinized (Barisione & Catellani, 2008; Birch & Allen, 2010; Bull & Fetzer, 2010).

Two studies were conducted. Each involved asking participants to read a fictional interview between a journalist and a politician. In both studies, the final statement of the journalist was manipulated, so that it was either a factual or a counterfactual attack. In Study 1, a no-attack (control) condition was also included. It was predicted that: (1) the counterfactual attack would induce a comparatively more negative evaluation of the target than the factual attack (or the control condition); and (2) this effect would be mediated by the perception that the source of the counterfactual attack was less biased against the target than the source of the factual attack.

In Study 2, we also manipulated how the source of the attack was initially presented by describing the journalist who interviewed the politician as having been either neutral or hostile toward the politician in the past. We predicted that when the journalist was described as neutral, the results would replicate those of Study 1. Namely, participants would perceive the source as less biased and would evaluate the target more negatively after a counterfactual than after a factual attack. When the journalist was described, instead, as being hostile to the politician, we predicted that the effects of factual and counterfactual attacks would not differ significantly. This should be the case because information about the hostile attitude of the source toward the target that is provided before exposure to the message would induce participants to perceive that the source is biased regardless of the factual versus counterfactual style of the message. Results from Study 2 should, therefore, help to clarify under what conditions a counterfactual attack against morality can be more effective than a factual attack.

# Study 1

In Study 1, we presented different groups of participants with three versions of a fictitious interview made by a journalist with a politician, as mentioned above. Two of them ended with either a factual or a counterfactual attack on the politician's morality, whereas the third one ended without an attack. We expected that participants in the counterfactual-attack condition would have a more negative overall judgment of the politician and a more negative evaluation of the politician's morality than participants in the

factual-attack and no-attack conditions. We also expected participants in the counterfactual-attack condition to perceive the journalist as less biased and to perceive the remark as more appropriate than those in a factual-attack condition. Finally, we tested whether participants' overall judgments of the politician, after a factual or a counterfactual attack, were mediated by the perception that the journalist was biased against the politician.

# Method

## Participants and Design

A sample of 108 students from the Catholic University of Milan (48.6% males; age $M = 25.5$, $SD = 8.93$) were assigned randomly, in equal numbers, to one of the three conditions that presented different versions of the attack (factual vs. counterfactual vs. no attack).

## Procedure

Participants read a 250-word fictional narrative that was presented as an excerpt from an interview of a journalist with an incumbent Prime Minister. In the interview, the journalist questioned the Prime Minister about the government's interventions in the economy. The politician answered by citing the positive results the government had obtained, reducing the high costs of bureaucracy and the political process and thus improving the national budget. At the end of the interview, the final remark of the journalist was manipulated. In the *factual-attack* condition the journalist said: "You acted incorrectly on the fiscal problem. You disregarded your previous commitment about the issue of taxes. You misrepresented the problem of taxation burdens!" In the *counterfactual-attack* condition, the closing remark was: "If you had acted correctly on the fiscal problem, our country would be in a better condition today. If you had honored your previous commitment about the issue of taxes, things would be better now. Today citizens would be more satisfied, if you had told the truth on the problem of taxation burdens!" Finally, in the *no-attack condition* the journalist simply thanked the politician: "All right. Thank you for participating. We hope to have you again as a guest soon."

## Measures

After reading the interview, participants were asked to answer a series of questions about the target of the attack, the source of the attack, and the attack itself. The order of these three blocks of questions was counterbalanced. As to the attacked target, participants were asked to give their overall judgment of the politician ("What is your overall judgment of the interviewed politician?") using a 7-point scale ranging from *very negative* (1) to *very positive* (7). Participants were also asked to rate the politician on a

series of four items that measured the extent to which they found the politician to be honest, sincere, loyal, and trustworthy on a 7-point scale ranging from *not at all* (1) to *very much* (7). These trait adjectives were chosen to measure the politician's morality, according to previous research on the operationalization and measurement of personality dimensions (Abele, Uchronski, Suitner, & Wojciszke, 2008). A single index was computed from the mean of the four items (Cronbach's $\alpha$ = .89).

As to the attacking source, participants were asked three questions to measure the degree to which they perceived that the journalist was biased against the politician. The questions were: "In your opinion, to what extent the journalist is attacking the politician?"; "To what extent do you think the journalist has a hostile attitude toward the politician?"; and "To what extent do you think the journalist's remarks are based on personal judgments?". Answers were reported on a 7-point scale, ranging from *not at all* (1) to *very much* (7). A measure of the perceived bias of the journalist was computed from the mean of the three items (Cronbach's $\alpha$ = .68). Finally, using the same 7-point scale, participants rated to what extent they found the final remark by the journalist to be relevant, intelligent, polite, and balanced. A single index of the perceived appropriateness of the remark was computed from the mean of the four items (Cronbach's $\alpha$ = .75).

## Results and Discussion

### Overall Judgment of the Attacked Target

A one-way ANOVA was performed on participants' overall judgment of the attacked target, with style of the attack (factual, counterfactual, or no attack) as a between-subjects factor. A significant effect of attack style emerged $F(1, 105) = 3.45$, $p < .05$, $\eta^2 = .06$. Consistent with our expectations, post hoc comparisons showed that participants' overall judgment of the target was more negative after a counterfactual attack ($M = 2.87$, $SD = 1.21$) than after a factual attack ($M = 3.52$, $SD = 1.43$) and after no attack ($M = 3.65$, $SD = 1.45$), both $p < .05$ (Figure 1, left side). The overall judgments of the target in the factual and no-attack conditions were comparable, $p = .695$.

### Evaluation of the Target's Morality

A one-way ANOVA was also performed on the evaluation of the target's morality. Again, a significant effect of attack style emerged, $F(1, 103) = 6.98$, $p < .01$, $\eta^2 = .12$. Post hoc comparisons showed that the ratings of the politician's morality were more negative after a counterfactual attack ($M = 2.81$, $SD = 1.01$) than after a factual attack ($M = 3.45$, $SD = 1.19$), $p < .05$, or no attack ($M = 3.85$, $SD = 1.34$), $p < .001$. Ratings of the politician's morality did not differ between the factual and the no-attack conditions, $p = .488$.

### Perceived Bias of the Attacking Source

Another one-way ANOVA was performed on the journalist's perceived bias against the politician. A strong effect of attack style was found, $F(1, 105) = 12.93$, $p < .001$, $\eta^2 = .20$. Participants perceived the journalist as more biased in the factual-attack condition ($M = 5.50$, $SD = 0.81$) than in the counterfactual-attack ($M = 4.74$, $SD = 0.93$) or the no-attack conditions ($M = 4.37$, $SD = 1.09$), $p < .01$ and $p < .001$ respectively. Perceived bias in the counterfactual and no-attack conditions did not differ significantly, $p = .109$.

### Perceived Appropriateness of the Remark

A one-way ANOVA was performed on participants' ratings of the perceived appropriateness of the final remark by the journalist. A main effect of the style of attack was again found, $F(1, 105) = 6.27$, $p < .01$, $\eta^2 = .10$. Post hoc comparisons showed that the factual attack was perceived as less appropriate ($M = 2.79$, $SD = 0.82$) than the counterfactual attack ($M = 3.46$, $SD = 1.49$), $p < .05$, or the no-attack statement ($M = 3.75$, $SD = 1.21$), $p < .001$ (Figure 1, right side). The ratings of the remark in the counterfactual and no-attack conditions did not differ significantly, $p = .586$.

### Mediation Analysis

Finally, we performed mediation analysis to test our prediction that the effectiveness of the counterfactual attack, compared with the factual attack, would depend on the perceived bias of the source of the attack. We predicted that the difference in the overall judgment of the target, between the two attack conditions, would be fully accounted for by the difference in the perceived bias of the source. Following the procedure described by Baron and Kenny (1986), we first regressed the participants' perception of the journalist's bias on the attack style (either factual or counterfactual), and found that the counterfactual attack was related to the perception the journalist was less biased, $\beta = -.41$, $t(73) = 3.79$, $p < .001$. We then regressed the participants' overall judgment of the politician on attack style and found it had a significant negative association with attack style, $\beta = -.24$, $t(73) = 2.12$, $p < .05$. Finally, we added perceived bias to the model and found that its association with the overall judgment of the politician was significant, $\beta = .32$, $t(72) = 2.68$, $p < .01$, whereas the association of attack style was no longer significant, $\beta = -.11$, $t(72) = 0.93$, $p = .354$ (Figure 2). Sobel's (1982) test of mediation, $z = 2.14$, $p < .05$, indicated that the effect of attack style on the overall judgment of the politician was fully mediated by the perceived bias of the journalist.

We also tested the feedback model by reversing the order of the mediator and the dependent variable. When we regressed perceived bias on attack style and the overall judgment of the politician, we found that although the overall judgment of the politician was a significant positive
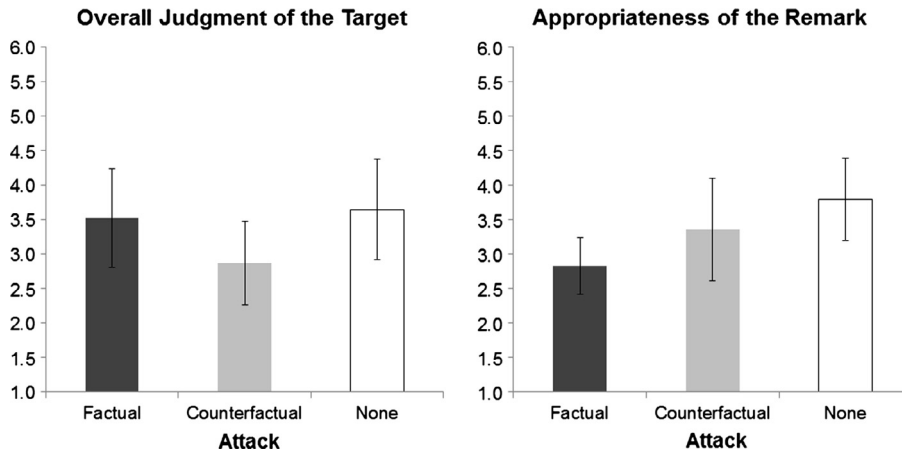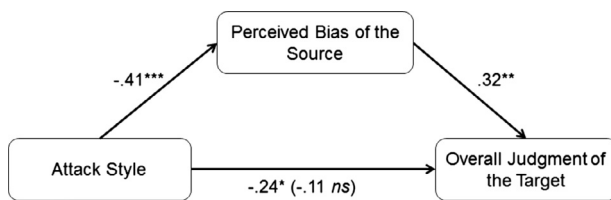
*Figure 2.* Perceived bias of the source as a mediator of the effect of attack style (factual vs. counterfactual) on the overall judgment of the target (Study 1). *p < .05; **p < .01.

predictor of perceived bias, $\beta = .28$, $t(72) = 2.66$, $p < .01$, the effect of attack style remained significant, $\beta = -.34$, $t(72) = 3.19$, $p < .01$. Sobel's test confirmed that the overall judgment of the politician did not mediate the effect of attack style on perceived bias, $z = 1.66$, $p = .10$. This result corroborated our hypothesis regarding the direction of the mediation, in that perceived bias mediated the effectiveness of counterfactual attack (compared to factual attack) on the overall judgment of the politician, but not vice-versa.

To sum up, the results fully supported our predictions. The counterfactual attack on morality was more effective than the factual attack, the effectiveness of which did not differ from that observed in the absence of an attack. Compared with participants in the factual-attack and no-attack conditions, participants in the counterfactual-attack condition gave lower overall ratings of the target and the target's morality. Moreover, participants in the counterfactual condition rated the final remarks of the source as more appropriate, and rated the source as less biased, than participants in the factual condition. Finally, as expected, the perception of the source's bias fully mediated the effect of the attack on the evaluation of the attacked target.

# Study 2

Like Study 1, Study 2 compared the effects of a factual and a counterfactual attack on the perceived morality of a

politician. However, in Study 2 we also manipulated the neutrality of the journalist, who was described as having been either neutral or hostile to the interviewed politician in the past. In the neutral-source condition, we expected to replicate the findings of Study 1, in which a counterfactual attack produced less perceived bias by the source and, therefore, more negative judgments about the target than a factual attack. In the hostile-source condition, on the other hand, we expected that the factual and the counterfactual attack would not differ as to their effects on the perceived bias of the source and the judgments about the target. Being aware of the hostility of the source toward the target from the beginning, participants should perceive the source as biased regardless of the way the attack was formulated. Therefore, in the hostile-source condition the effect of the attack style on the evaluation of the attacked target would not anymore be mediated by the perceived bias of the source. The mediation effect should instead be still present in the neutral-source condition, again replicating the results of Study 1.

## Method

### Participants and Design

A sample of 81 students from the Catholic University of Milan (58.8% females; age $M = 22.8$, $SD = 3.92$) were assigned randomly and equally to one of the four experimental conditions produced by the manipulation of the attack style (factual vs. counterfactual) and the journalist's attitude (hostile vs. neutral) toward the attacked politician.

### Procedure

Students were invited to volunteer to participate in an online study in which they were asked to read the same 250-word fictional interview that was used in Study 1. The factual and counterfactual versions of the final attack on the politician's morality were the same as those employed in Study 1. This time, however, the journalist was described at the beginning of the interview as a person

who had always been either neutral or hostile toward the politician, depending on the experimental condition. After reading the interview, participants were asked to complete a short questionnaire.

### Measures

To verify the effectiveness of the manipulation about the journalist's neutrality, participants were asked to indicate whether the journalist who did the interview had been neutral or hostile to the politician in the past. Participants were then asked to rate their overall judgment of the politician, the morality of the politician, the journalist's bias against the politician, and the appropriateness of the journalist's final remark. These measures were the same as those used in Study 1. The order of presentation of the measures was randomized. The scales measuring the perceived morality of the politician, the perceived bias of the journalist, and the appropriateness of the remark were highly reliable (Cronbach's $\alpha = .93$, $\alpha = .92$, and $\alpha = .89$, respectively).

## Results and Discussion

### Preliminary Analyses

We performed a preliminary check to determine whether participants recognized the prior attitude of the journalist toward the politician that was described in our experimental manipulation. In the hostile-source condition, the great majority of participants (94.4%) acknowledged that the journalist had been initially hostile to the politician. In the neutral-source condition, the majority of participants acknowledged the initial neutrality of the journalist (82.2%), although some participants (17.8%) responded that the journalist previously had been hostile to the politician. Participants who did not correctly recognize the manipulations ($n = 10$) were removed from further analyses, leaving 71 participants in the main analyses.

### Overall Judgment of the Attacked Target

A 2 (Source Neutrality: neutral vs. hostile) × 2 (Attack Style: factual vs. counterfactual) between-subjects ANOVA was performed on participants' overall judgment of the politician. A significant main effect of source neutrality was found, $F(1, 67) = 11.84$, $p < .005$, $\eta^2 = .18$, with the overall judgment of the politician being more negative in the neutral-source condition ($M = 3.24$, $SD = 1.58$) than in the hostile-source condition ($M = 4.38$, $SD = 1.38$). A significant effect of attack style was also found, $F(1, 67) = 5.16$, $p < .05$, $\eta^2 = .08$, with the overall judgment being more negative in the counterfactual-attack condition ($M = 3.46$, $SD = 1.69$) than in the factual-attack condition ($M = 4.19$, $SD = 1.39$). These effects were qualified by a significant interaction effect between source neutrality and attack style, $F(1, 67) = 5.44$, $p < .05$, $\eta^2 = .08$. As shown in Figure 3 (left side), the counterfactual attack resulted in a more negative evaluation of the politician ($M = 2.47$, $SD = 1.38$) than the factual attack ($M = 4.00$, $SD = 1.41$), $t(32) = 3.20$, $p < .01$, in the neutral-source condition, thus replicating the findings of Study 1. No significant difference was found, $t(35) = 0.04$, $p = .96$, in participants' evaluations after a factual ($M = 4.37$, $SD = 1.38$) and a counterfactual attack ($M = 4.39$, $SD = 1.42$), in the hostile-source condition. Thus, as we expected, when the journalist was described as neutral, the counterfactual attack was more effective than the factual attack in lowering the overall judgment of the politician. Instead, when the journalist was described as hostile the difference between the two attacks was not significant.

### Evaluation of the Target's Morality

The same 2 (source neutrality) × 2 (attack style) ANOVA was performed on the ratings of the politician's morality. A significant main effect of source neutrality was found, with lower perceived morality after an attack by a neutral source ($M = 3.03$, $SD = 1.65$) than after an attack
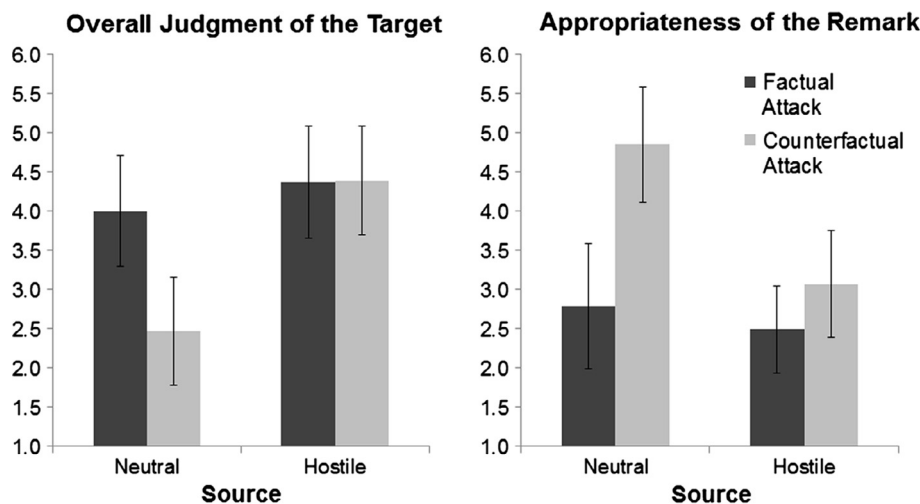


*Figure 3.* Overall judgment of the target and perceived appropriateness of the remark as a function of attack style (factual vs. counterfactual) and source neutrality (neutral vs. hostile) (Study 2).

by a hostile source ($M = 3.87$, $SD = 1.63$), $F(1, 67) = 4.94$, $p < .05$, $\eta^2 = .7$. No main effect of attack style was found, $F(1, 67) = 2.42$, $p = .12$, $\eta^2 = .03$, but the interaction effect between source neutrality and attack style approached significance, $F(1, 67) = 3.72$, $p = .06$, $\eta^2 = .06$. Subsequent $t$-tests for each subgroup confirmed that the evaluation of the politician's morality was lower after a counterfactual attack ($M = 2.37$, $SD = 1.26$) than after a factual attack ($M = 3.69$, $SD = 1.77$), $t(32) = 2.51$, $p < .05$, in the neutral-source condition. Instead, the evaluations of the politician's morality after a factual or a counterfactual attack were similar in the hostile-source condition ($M = 3.80$, $SD = 1.80$ vs. $M = 3.94$, $SD = 1.44$, respectively), $t(35) = 0.26$, $p = .80$. These results further confirmed that counterfactual attacks were more effective than factual attacks in reducing the perceived morality of the target, but only when the source of the attack was described as neutral.

### Perceived Bias of the Attacking Source

Another 2 (source neutrality) × 2 (attack style) ANOVA was performed on the journalist's perceived bias against the politician. Not surprisingly, a strong effect of source neutrality was found, $F(1, 67) = 21.01$, $p < .001$, $\eta^2 = .21$, in which the journalist described as hostile was perceived as being more biased against the target ($M = 5.82$, $SD = 1.13$) than the journalist described as neutral ($M = 4.11$, $SD = 1.88$). Attack style also had a main effect, with a counterfactual attack leading to less perceived bias of the journalist ($M = 4.60$, $SD = 1.82$) than a factual attack ($M = 5.40$, $SD = 1.62$), $F(1, 67) = 5.80$, $p < .05$, $\eta^2 = .09$. The two main effects were qualified by a significant interaction, $F(1, 67) = 6.71$, $p < .05$, $\eta^2 = .10$. The difference between counterfactual and factual attack was significant in the neutral journalist condition ($M = 3.25$, $SD = 1.51$ vs. $M = 4.98$, $SD = 1.87$), $t(32) = 2.93$, $p < .01$, whereas it was not significant in the hostile journalist condition ($M = 5.85$, $SD = 0.99$ vs. $M = 5.79$, $SD = 1.28$), $t(35) = 0.09$, $p = .87$. As expected, then, using a counterfactual attack instead of a factual attack reduced the perceived bias of the journalist, but this reduction was significant only if the journalist was initially described as neutral.

### Perceived Appropriateness of the Remark

A 2 (source neutrality) × 2 (attack style) ANOVA also was performed on the perceived appropriateness of the remark made by the journalist. Once again, we found main effects of source neutrality and attack style. First, an attack made by a journalist initially described as neutral was rated more appropriate ($M = 3.82$, $SD = 1.84$) than an attack made by a journalist described as hostile ($M = 2.77$, $SD = 1.27$), $F(1, 67) = 9.95$, $p < .005$, $\eta^2 = .15$. Second, a counterfactual attack was rated as more appropriate ($M = 3.94$, $SD = 1.62$) than a factual attack ($M = 2.63$, $SD = 1.41$), $F(1, 67) = 15.88$ $p < .001$, $\eta^2 = .24$. The interaction

between the two variables was also significant, $F(1, 67) = 4.96$, $p < .05$, $\eta^2 = .07$. As shown in Figure 3 (right side), a counterfactual attack was perceived as much more appropriate than a factual attack in the neutral journalist condition ($M = 4.85$, $SD = 1.59$ vs. $M = 2.79$, $SD = 1.47$, respectively), $t(32) = 3.92$, $p < .001$, whereas the difference between a counterfactual and a factual attack was not significant in the hostile journalist condition ($M = 3.07$, $SD = 1.12$ vs. $M = 2.49$, $SD = 1.37$), $t(35) = 1.41$, $p = .17$. Thus, counterfactual attacks were perceived as more appropriate than factual attacks, especially when they were made by a journalist initially described as neutral.

### Mediation Analysis

Finally, we performed two mediation analyses to test whether the effect of attack style on the evaluation of the target was mediated by the perceived bias of the source, separately for the neutral and the hostile-source conditions. We predicted that the mediation effect would be similar to the one found in Study 1 in the neutral-source condition, whereas no mediation effect would be found in the hostile-source condition. Following the procedure described by Baron and Kenny (1986), we first regressed the participants' perception of the journalist's bias on the attack style in the neutral-source condition ($n = 33$). We found that the counterfactual attack was related to the perception that the journalist was less biased, $\beta = -.46$, $t(32) = 2.93$, $p < .01$. Then, we regressed the participants' overall judgment of the politician on attack style and found it had a significant negative association with counterfactual attack, $\beta = -.50$, $t(32) = 3.27$, $p < .01$. Finally, when we added perceived bias to the model we found that its association with the overall judgment of the politician was significant, $\beta = .46$, $t(31) = 2.98$, $p < .01$, whereas the association of attack style was no longer significant, $\beta = -.29$, $t(31) = 1.87$, $p = .07$. Sobel's (1982) test of mediation, $z = 2.09$, $p < .05$, revealed that the perceived bias of the journalist fully mediated the effect of attack style on the overall judgment of the politician (Figure 4, upper panel), thus replicating the results of Study 1.

We performed the same mediation analysis for the hostile-source condition ($n = 37$). Although the perception that the journalist was biased and the overall judgment of the politician were correlated, $r(35) = .45$, $p < .01$, attack style was not found to have a significant effect on either the perception of the journalist being biased, $\beta = .03$, $t(35) = 0.17$, $p = .87$, or the overall judgment of the politician, $\beta = .01$, $t(35) = 0.04$, $p = .96$ (see Figure 4, lower panel). Consequently, Baron and Kenny's (1986) criteria for mediation were not met. These results confirmed our expectation that, in the hostile-source condition, the perceived bias of the source would not mediate the effect of attack style on the overall judgment of the target.

In summary, the results of Study 2 replicated those of Study 1 with regard to the greater effectiveness of a counterfactual compared to a factual attack. As we predicted, however, we found that describing the journalist as a person
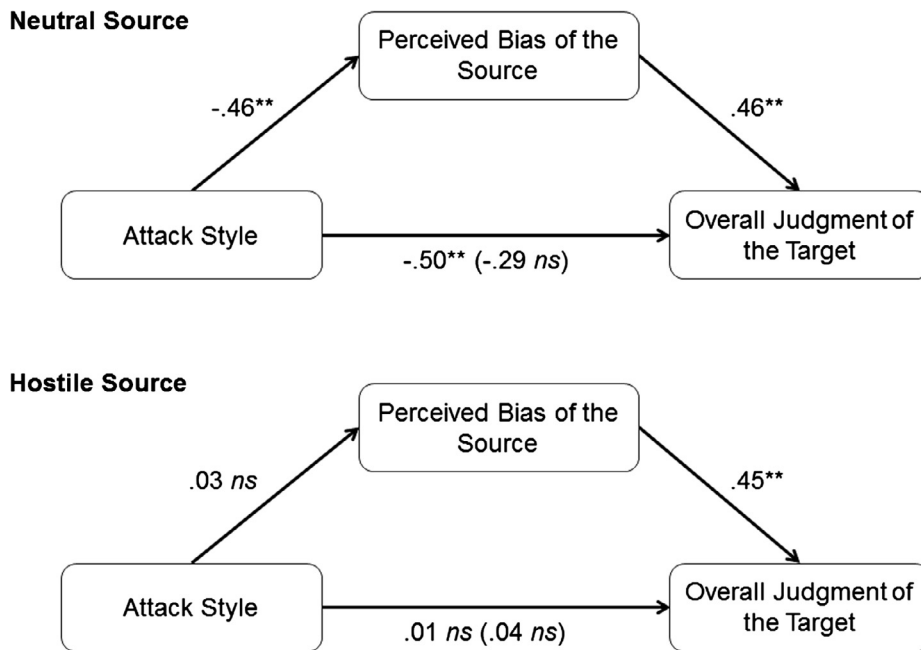
**Neutral Source**



**Hostile Source**



*Figure 4.* Perceived bias of the source presented as either neutral (upper panel) or hostile (lower panel) as a mediator of the effect of attack style (factual vs. counterfactual) on the overall judgment of the target (Study 2). **$p < .01$.

who had been hostile to the politician in the past reduced the effectiveness of counterfactual attacks, compared to factual attacks. When the journalist was described as hostile to the politician, the overall judgment of the politician, the evaluation of the politician's morality, and the perceived bias of the source were not influenced by the style of the attack. However, there was still a tendency to perceive the attack as more appropriate after a counterfactual rather than a factual attack.

Results of the mediation analyses replicated those of Study 1 when the source of the attack was described as neutral. In this condition, a counterfactual attack reduced the perceived bias of the source, which, in turn, induced a more negative judgment on the attacked target than a factual attack. No such mediation effect was found when the source was described from the beginning as being hostile to the target. Evidently, when participants were initially told that the source of the attack had been hostile to the target in the past, the counterfactual style of the attack was not enough to reduce their perceived bias of the source and, therefore, to influence their judgments on the attacked target.

## General Discussion

The results of our research show that counterfactuals can be effectively employed to attack other people's morality. Compared with a factual or no attack, a counterfactual attack leads to more negative evaluations of the target of the attack, and more positive evaluations of the source of the attack and the appropriateness of the attack message. The greater effectiveness of counterfactual attack over factual attack can be attributed to its ability to reduce the perception that the source is biased against the target.

Our results contribute to research on the effects of criticism on the evaluation of other people. Previous research has shown that attacks – even on a dimension perceived as very relevant, such as morality – are often ineffective, in part because they induce backlash toward the source of the attack (Budesheim et al., 1996; Carraro et al., 2010; Haddock & Zanna, 1997; Roese & Sande, 1993). The present results suggest that part of the observed ineffectiveness of attacks may be attributed to style.

The advantage of a counterfactual attack over a factual attack is twofold. First, whereas factual criticism focuses on what an actor did, counterfactual criticism takes past actions for granted and shifts attention to the opportunities the actor had to behave in a different way and, hence, obtain a better outcome. As shown by previous research, counterfactual thinking makes violations of norms (including social norms) salient (Kahneman & Miller, 1986), increasing the blame for actors who violate the norms (e.g., Catellani & Milesi, 2001; Catellani et al., 2004; Mandel, 2003; Tal-Or et al., 2004). Our results show that this mechanism can be used in communication to criticize someone indirectly, making the target more likely to be blamed for the violation of a norm. Secondly, given that it is formulated in the conditional rather than the indicative mode, which mitigates illocutionary force (Thaler, 2012), a counterfactual attack raises less suspicion about the source's motives than a factual attack. This makes the counterfactual attack more persuasive, and reduces the likelihood of a backlash effect. In Study 1, counterfactual attack evidently prevented observers from discounting the negative information cited by the attacking source. This interpretation is further corroborated by the fact that, in Study 2, counterfactual attacks maintained their effectiveness when the attack source was described as neutral, and were still evaluated positively even when the source was described as hostile to the target. Together, these results

on the effectiveness of counterfactual attacks further contribute to previous research that suggests that the effectiveness of an attack does not depend on its overt force, but rather on its subtlety (Fiedler, 2008; Fiedler & Mata, 2013).

Results from Study 2 further clarified under what conditions the factual or counterfactual style of an attack influences its effectiveness. In general, using a counterfactual attack decreases the perceived bias of the source and improves the effectiveness of the attack compared to a factual attack. This effect, however, is reduced when other information, provided by the context, induces observers to think that the source is biased. In other words, linguistic cues derived from the style of an attack message are used by observers to infer the communicative intention of the source when other information about the source's attitude toward the target is initially unavailable. However, when such information is available, it is likely to influence how the message is processed, and to lead to an interpretation that is consistent with expectations, independent on the style of the message (Fischer, Jonas, Frey, & Schulz-Hardt, 2005; Taber & Lodge, 2006).

Our results also contribute to research that investigates how the morality of a person is perceived, by focusing on the hitherto underdeveloped theme of identifying which ways of attacking morality prove to be effective. As discussed in the Introduction, morality is the most important dimension in person perception (Abele & Bruckmüller, 2011; Cislak & Wojciszke, 2008; Wojciszke & Abele, 2008), and the most heavily affected by negative information (Kervyn et al., 2010; Skowronski, 2002; Skowronski & Carlston, 1989). However, past research has not found attacks on morality to be very effective (Carraro et al., 2010; Funk, 1996). Our results suggest the potential importance of how an attack on morality is formulated. A counterfactual attack affects the evaluation of the target's morality, and the target generally, more than a factual attack. It is possible that, owing to the importance attributed to morality, people are especially attentive to the way criticism is made, which provides a cue to the source's possible bias toward the target. If a criticism regarding morality overcomes this close scrutiny, as in the case of the counterfactual attack in our research, its effectiveness is likely to be high.

In conclusion, the present research showed that a counterfactual attack reduces the perception of morality, which is a personality dimension that is particularly important in the evaluation of other people. Although counterfactual attacks focus on the target's alleged violation of a social norm, they do so in a way that is perceived as softer and more fair than other kinds of attacks. Counterfactuals should, therefore, be included among the subtle communication strategies that allow criticism to be effective while avoiding the backlash effect against the source of criticism.

## Acknowledgments

## References

Abele, A. E., & Bruckmüller, S. (2011). The bigger one of the "Big Two"? Preferential processing of communal information. *Journal of Experimental Social Psychology, 47*, 935–948.

Abele, A. E., Uchronski, M., Suitner, C., & Wojciszke, B. (2008). Towards an operationalization of the fundamental dimensions of agency and communion: Trait content ratings in five countries considering valence and frequency of word occurrence. *European Journal of Social Psychology, 38*, 1202–1217.

Alicke, M. D., Buckingham, J., Zell, E., & Davis, T. (2008). Culpable control and counterfactual reasoning in the psychology of blame. *Personality and Social Psychology Bulletin, 34*, 1371–1381.

Barisione, M., & Catellani, P. (2008). L'offerta personalizzata degli sfidanti. In ITANES. (Ed.), *Il ritorno di Berlusconi. Vincitori e vinti nelle elezioni del 2008* [The return of Berlusconi: Winners and losers in 2008 elections]. (pp. 137–148). Bologna, Italy: Il Mulino.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.

Bavelas, J. B., Black, A., Bryson, L., & Mullett, J. (1988). Political equivocation: A situational explanation. *Journal of Language and Social Psychology, 7*, 137–145.

Bertolotti, M., Catellani, P., Douglas, K. M., & Sutton, R. M. (2013). The "Big Two" in political communication: The effects of attacking and defending politicians' leadership or morality. *Social Psychology, 44*, 117–128.

Birch, S., & Allen, N. (2010). How honest do politicians need to be? *The Political Quarterly, 81*, 49–56.

Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). Looking for honesty: The primary role of morality (vs. sociability and competence) in information gathering. *European Journal of Social Psychology, 41*, 135–143.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usages*. Cambridge, UK: Cambridge University Press.

Budesheim, T. L., Houston, D. A., & DePaola, S. J. (1996). Persuasiveness of in-group and out-group political messages: The case of negative political campaigning. *Journal of Personality and Social Psychology, 70*, 523–534.

Bull, P. E., & Fetzer, A. (2010). Face, facework and political discourse. *International Review of Social Psychology, 23*, 155–185.

Bull, P. E., & Wells, P. (2012). Adversarial discourse in Prime Minister's Questions. *Journal of Language and Social Psychology, 31*, 30–48.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1*, 3–25.

Carraro, L., Gawronski, B., & Castelli, L. (2010). Losing on all fronts: The effects of negative versus positive person-based campaigns on implicit and explicit evaluations of political candidates. *British Journal of Social Psychology, 49*, 453–470.

Catellani, P., Alberici, A. I., & Milesi, P. (2004). Counterfactual thinking and stereotypes: The nonconformity effect. *European Journal of Social Psychology, 34*, 421–436.

Catellani, P., & Bertolotti, M. (2014). The effects of counterfactual defences on social judgements. *European Journal of Social Psychology.* doi: 10.1002/ejsp.1995.

Catellani, P., & Covelli, V. (2013). The strategic use of counterfactual communication in politics. *Journal of Language and Social Psychology, 32*, 495–504.

Catellani, P., Bertolotti, M., & Covelli, V. (2013). Counterfactual communication in politics: Features and effects on voters. In I. Poggi, F. D'Errico, L. Vincze, & E. A. Vinciarelli (Eds.), *Multimodal communication in political speech* (pp. 75–85). Springer Verlag: Berlin.

Catellani, P., & Milesi, P. (2001). Counterfactuals and roles: Mock victims' and perpetrators' accounts of judicial cases. *European Journal of Social Psychology, 31*, 247–264.

Catellani, P., & Milesi, P. (2005). When the social context frames the case: Counterfactuals in the courtroom. In D. Mandel, D. Hilton, & P. Catellani (Eds.), *The psychology of counterfactual thinking* (pp. 183–198). London, UK: Routledge.

Cislak, A., & Wojciszke, B. (2008). Agency and communion are inferred from actions serving interests of self or others. *European Journal of Social Psychology, 38*, 1103–1110.

Douglas, K. M., & Sutton, R. M. (2003). Effects of communication goals and expectancies on language abstraction. *Journal of Personality and Social Psychology, 84*, 682–696.

Esposo, S. R., Hornsey, M. J., & Spoor, J. R. (2013). Shooting the messenger: Outsiders critical of your group are rejected regardless of argument quality. *British Journal of Social Psychology, 52*, 386–395.

Fiedler, K. (2008). Language – A toolbox for sharing and influencing social reality. *Perspectives on Psychological Science, 3*, 38–47.

Fiedler, K., & Mata, A. (2013). The art of exerting verbal influence through powerful lexical stimuli. In J. P. Forgas, J. Laszlo, & O. Vincze (Eds.), *Social cognition and communication* (pp. 43–62). New York, NY: Psychology Press.

Fischer, P., Jonas, E., Frey, D., & Schulz-Hardt, S. (2005). Selective exposure to information: The impact of information limits. *European Journal of Social Psychology, 35*, 469–492.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38*, 889–906.

Funk, C. L. (1996). The impact of scandal on candidate evaluations: An experimental test of the role of candidate traits. *Political Behavior, 18*, 1–24.

Gawronski, B., & Walther, E. (2008). The TAR effect: When the ones who dislike become the ones who are disliked. *Personality and Social Psychology Bulletin, 34*, 1276–1289.

Girotto, V., Legrenzi, P., & Rizzo, A. (1991). Event controllability in counterfactual thinking. *Acta Psychologica, 78*, 111–133.

Haddock, G., & Zanna, M. (1997). Impact of negative advertising on evaluations of political candidates: The 1993 Canadian federal election. *Basic and Applied Social Psychology, 19*, 205–223.

Hill, R. P. (1989). An exploration of voter responses to political advertisements. *Journal of Advertising, 18*, 14–22.

Hornsey, M. J. (2005). Why being right is not enough: Predicting defensiveness in the face of group criticism. *European Review of Social Psychology, 16*, 301–334.

Hornsey, M. J., & Imani, A. (2004). Criticizing groups from the inside and the outside: An identity perspective on the intergroup sensitivity effect. *Personality and Social Psychology Bulletin, 30*, 365–383.

Hornsey, M. J., Oppes, T., & Svensson, A. (2002). "It's ok if we say it, but you can't": Responses to intergroup and intragroup criticism. *European Journal of Social Psychology, 32*, 293–307.

Hornsey, M. J., Robson, E., Smith, J., Esposo, S., & Sutton, R. M. (2008). Sugaring the pill: Assessing rhetorical strategies designed to minimize defensive reactions to group criticism. *Human Communication Research, 34*, 70–98.

Ito, T. A., Larsen, J. T., Smith, N. K., & Cacioppo, J. T. (1998). Negative information weighs more heavily on the brain: The negativity bias in evaluative categorizations. *Journal of Personality and Social Psychology, 75*, 887–900.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, E. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York, NY: Cambridge University Press.

Kervyn, N., Dolderer, M., Mahieu, T., & Yzerbyt, V. Y. (2010). Atypicality and the two fundamental dimensions: Applying the negativity effect on warmth to group perception. *European Journal of Social Psychology, 40*, 484–489.

Mandel, D. R. (2003). Effect of counterfactual and factual thinking on causal judgments. *Thinking and Reasoning, 9*, 245–265.

Markman, K. D., & Tetlock, P. E. (2000). "I couldn't have known": Accountability, foreseeability and counterfactual denials of responsibility. *British Journal of Social Psychology, 39*, 313–325.

Matthews, D., & Dietz-Uhler, B. (1998). The black-sheep effect: How positive and negative advertisements affect voters' perceptions of the sponsor of the advertisement. *Journal of Applied Social Psychology, 28*, 1903–1915.

McClure, J., Hilton, D. J., & Sutton, R. M. (2007). Judgments of voluntary and physical causes in causal chains: Probabilistic and social functionalist criteria for attributions. *European Journal of Social Psychology, 37*, 879–901.

Nario-Redmond, M., & Branscombe, N. (1996). It could have been better or it might have been worse: Implications for blame assignment in rape cases. *Basic and Applied Social Psychology, 18*, 347–366.

Peeters, G. (1971). The positive-negative asymmetry: On cognitive consistency and positivity bias. *European Journal of Social Psychology, 1*, 455–474.

Peeters, G., & Czapinski, J. (1990). Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects. *European Review of Social Psychology, 1*, 33–60.

Rabinovich, A., & Morton, T. A. (2010). Who says we are bad people? The impact of criticism source and attributional content on responses to group-based criticism. *Personality & Social Psychology Bulletin, 36*, 524–36.

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86*, 61–79.

Roese, N. J. (1997). Counterfactual thinking. *Psychological Bulletin, 121*(1), 133–148.

Roese, N. J., & Sande, G. N. (1993). Backlash effect in attack politics. *Journal of Applied Social Psychology, 23*, 632–653.

Skowronski, J. J. (2002). Honesty and intelligence judgments of individuals and groups: The effects of entity-related behavior diagnosticity and implicit theories. *Social Cognition, 20*, 136–169.

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52*, 689–699.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin, 105*, 131–142.

Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take

on the qualities they describe in others. *Journal of Personality and Social Psychology, 74*, 837–848.

Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhart (Ed.), *Sociological methodology* (pp. 290–312). San Francisco, CA: Jossey-Bass.

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*, 755–769.

Tal-Or, N., Boninger, D.S., Poran, A., & Gleicher, F. (2004). Counterfactual thinking as a mechanism in narrative persuasion. *Human Communication Research, 30*, 301–328.

Thaler, V. (2012). Mitigation as modification of illocutionary force. *Journal of Pragmatics, 44*, 907–919.

Vonk, R. (1999). Effects of other-profitability and self-profitability on evaluative judgements of behaviours. *European Journal of Social Psychology, 29*, 833–842.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology, 56*, 161–169.

Wojciszke, B. (2005). Morality and competence in person and self-perception. *European Review of Social Psychology, 16*, 155–188.

Wojciszke, B., & Abele, A. E. (2008). The primacy of communion over agency and its reversals in evaluations. *European Journal of Social Psychology, 38*, 1139–1147.

Wojciszke, B., Brycz, H., & Borkenau, P. (1993). Effects of information content and evaluative extremity on positivity and negativity biases. *Journal of Personality and Social Psychology, 64*, 327–336.

Wong, E. M. (2010). It could have been better: The effects of counterfactual communication on impression formation. *European Journal of Social Psychology, 40*, 1251–1260.

Wyer, R. S., Swan, S., & Gruenfeld, D. H. (1995). Impression formation in informal conversations. *Social Cognition, 13*, 243–272.

Ybarra, O., Chan, E., Park, D., Burnstein, E., Monin, B., & Stanik, C. (2008). Life's recurring challenges and the fundamental dimensions: An integration and its implications for cultural differences and similarities. *European Journal of Social Psychology, 38*, 1083–1092.

Patrizia Catellani

Department of Psychology
Catholic University of Milan
Largo Gemelli, 1
20123 Milan
Italy
Tel. +39 02 72342906
Fax +39 02 72342280
E-mail patrizia.catellani@unicatt.it