Fooling the feeling of doing: a goal perspective on illusions of agency

Dannenberg, L.A.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3

If Only I Did It.

Counterfactual Thoughts Can Reduce the Susceptibility to Illusions of Agency

The experience of agency is considered an important part of healthy human functioning, because it helps people identify events in the world that they have caused by their actions (Frith, 1992; Michotte, 1963; Nahab et al., 2011). Arguably, this inference of agency is based on a number of agency or authorship indicators that operate within a comparator or matching model (Blakemore, Frith, & Wolpert, 1999; Wegner & Wheatley, 1999). In this model, congruence or incongruence between a predicted and an actual sensory or conceptual outcome is identified. Congruence between prediction and observation heightens the experience that “I did something”, whereas incongruence reduces this feeling of agency (Daprati et al., 1997; Georgieff & Jeannerod, 1998; Sato, 2009; Sato & Yasuda, 2005; Wegner & Wheatley, 1999). Specifically, people are likely to attribute causation for an action to the self if a consistent mental preview or representation of an action’s future outcome occurs prior to the actual outcome and is devoid of other potential causes for the action (Wegner & Wheatley, 1999). For example, if I think about turning on the light I mentally activate an outcome representation of the illuminated room. If then, after pushing the switch, the room is illuminated, this is a match with the activated thought. Lastly, if no one else could possibly have pushed the light switch, the attribution that I was the author of this event is rather obvious and usually correct.

While this heuristic of anticipated and observed outcomes yields a correct conclusion of agency most of the time, mistakes can occur when causation for an event is more ambiguous. In this case, letting people expect a particular outcome (even if it is uncontrollable) and then in reality presenting an expectancy-congruent one, can falsely heighten feelings of agency (Aarts, Custers, & Wegner, 2005; Aarts, 2007; Linser & Goschke, 2007). In a study demonstrating this illusion of agency over an event (Aarts et al., 2005), the participant was asked to rotate and stop a square that was traversing a rectangular path on the computer screen. At the same time, the computer rotated a second square across the path (see Figure 3.1). After some rotations, the participant was asked to stop the rotation by pressing a key, which made both squares vanish for a short moment. Then the stop position of only one of the two squares was presented. As the presented stop location could represent the participant’s square or the computer’s square, causation for this outcome was ambiguous. When the stop location was flashed subliminally on the screen before the participant pressed the stop key, he or she was more likely to claim causation for this outcome than when the stop location was not previously presented. Presumably, the short visual image of the stop location imitated the representation of an expected outcome and increased feelings of personal causation.
Importantly, some initial studies have hinted at boundary conditions for the illusory experience of agency elicited by such outcome priming. Specifically, the level at which people represent their behavior seems to play an important role in their susceptibility to the illusion. Representing behavior on a high-level keeps the focus on why an action is done and thereby activates the representation of action-outcomes. In contrast, a lower level of behavioral representation lets people focus on the action itself and how it is done (Fujita, Eyal, Chaiken, Trope, & Liberman, 2008; Liberman & Trope, 2008; see also Vallacher & Wegner, 1987). A high-level behavioral representation then corresponds to the level of outcome primes in the wheel-of-fortune paradigm, while a low-level focus on concrete action does not correspond to the abstraction level of outcome primes (Van der Weiden, Aarts, & Ruys, 2010). Accordingly, research has shown that when participants were instructed to focus on the high level outcome during the wheel-of-fortune task (i.e. stopping the square) outcome priming could falsely heighten feelings of agency. When however participants were instructed to focus on how they do the wheel-of-fortune task, thus emphasizing the low-level action (i.e. pressing a key), these illusions of agency were reduced (Van der Weiden et al., 2010).

Thus, one may conclude that high-level goal or outcome representations enhance the likelihood for agency illusions to occur, but not low-level representations of actions. In the present research we tested whether conditions that typically reduce such high-level outcome representations can subsequently influence the susceptibility to illusions of agency.

Previous research has shown that especially failure to reach a goal keeps representations of desired outcomes active (Förster, Liberman, & Friedman, 2007; Förster, Liberman, & Higgins, 2005; Goschke & Kuhl, 1993; Lewin, 1935; Marsh, Hicks, & Bink, 1998; Marsh, Hicks, & Bryan, 1999; Zeigarnik, 1927). In response to such goal failure, people are likely to generate counterfactual thoughts about the event (Sanna & Turley, 1996). Such counterfactuals are the mental simulations of alternative realities where a certain event would have turned out differently had the person acted in a different way (Roese, 1994). These specific action plans can aid in the successful future completion of an unattained goal and reduce the representation of high-level outcomes (Epstude & Roese, 2008, 2011). Considering that illusions of agency are reduced when people focus on their actions instead of on the higher-order outcomes, we propose that generating behavior-regulating counterfactual thoughts can reduce the susceptibility to these illusions.
Counterfactuals can take the form of upward comparisons, where people imagine a better alternative reality (“If only I did X, things would have been better”), or downward comparisons, where they simulate a worse alternative reality (“At least I didn’t do Y, or things would have been even worse”) (Epstude & Roese, 2008). Research shows that it is the generation of upward counterfactuals that produces a concrete behavioral intention (Smallman & Roese, 2009; Markman, Gavanski, Sherman, & McMullen, 1993). In contrast to relatively abstract goal intentions (“I need to loose weight”), behavioral intentions are more specific, because they concern the means to reach a goal (“I should eat less cake”) (Epstude & Roese, 2008; see also Gollwitzer & Brandstätter, 1997). Such a concrete behavioral intention translates the goal state from a higher level to a lower level of abstraction. Thereby, people can focus more on the means to reach a goal rather than on the desired outcome itself (Epstude & Roese, 2008, 2011; Kray, Galinsky, & Wong, 2006). In turn, such goal-conducive action is more likely to be performed in the future and therefore serves behavior regulation (Epstude & Roese, 2011). Upward counterfactuals that provide people with a focus on concrete means and let them focus relatively less on the desired outcome would then become a psychological basis for reducing illusions of agency.

Importantly, as we have hinted at already, the concrete focus on action at the expense of higher-order goal cognitions that results from upward counterfactuals serves a preparatory function for actions in the future (Epstude & Roese, 2008; Gollwitzer & Sheeran, 2006; Roese & Olson, 1995; Smallman & Roese, 2009). Behaviors that may reoccur in the future are more probable and consequently elicit more low-level or concrete construals (Wakslak, Trope, Liberman, & Alony, 2006). Therefore, especially if the person expects to encounter the situation to which the counterfactual thoughts apply again in the future, he or she should start developing specific action plans. For instance, failing to stay on a diet might result in thoughts such as “If only I hadn’t eaten that cake. Next time I see one, I will resist/leave the room/etc.” (Markman et al., 1993; Smallman & Roese, 2009). Thus, we further predict that an important qualification of the hypothesis that upward counterfactuals decrease illusory agency is that these alternative actions are of relevance to the future.

In the present research, we therefore propose that behavior-regulating upward counterfactuals that elicit concrete future-relevant action plans can reduce illusions of agency in a subsequent wheel-of-fortune task (Aarts et al., 2005). Specifically, we examined in a first study whether upward counterfactuals, in contrast to downward counterfactuals, that are generated in response to a re-
called goal-failure can reduce illusions of agency in the wheel-of-fortune paradigm. We chose the generation of downward counterfactuals as a comparison group because people also ponder their goal pursuit and are self-focused (Lyubomirsky & Nolen-Hoeksema, 1993). Yet, because downward counterfactuals do not serve behavioral regulation, they should emphasize specific action plans less. Consequently, downward counterfactuals should not influence susceptibility to agency illusions. Additionally, we sought to qualify this effect in a second study by manipulating the repeatability of the failure situation. We expected that upward counterfactuals in response to a goal pursuit failure would reduce illusions of agency, especially if people expect to encounter a similar event again in the future.

Study 3.1

In Study 3.1, we asked all participants to recall a situation of goal pursuit failure (cf. Moskowitz, 2002) and subsequently to generate upward or downward counterfactuals about that situation. We predicted that giving participants the possibility to generate upward counterfactuals can reduce illusions of agency in the wheel-of-fortune paradigm (Aarts et al., 2005), in contrast to generating downward counterfactuals (Epstude & Roese, 2008).

Method

Participants and Design. Participants were 45 undergraduates from the University of Amsterdam (30 females, $M_{age} = 22$) who were randomly assigned to either the upward or the downward counterfactuals condition.

Materials and Procedure. In the first part of the study participants received the recall task with instructions to “recall and write down a situation where something negative happened because you wanted to reach a certain goal but that did not happen, due to your own fault or because you made a mistake”. After writing about the situation for 2 minutes, participants received instructions to generate counterfactual thoughts about the goal pursuit failure. Specifically, they were asked “what could you have done differently such that”, depending on the condition, “the situation might have turned out better” (upward counterfactual) or “the situation might have turned out even worse” (downward counterfactual). To make sure that upward counterfactuals lead to more planning of concrete future actions than downward counterfactuals, we subsequently asked participants “If you ever encounter a similar situation again, would you do anything different?” (1-not at all; 7-very much). Lastly, partici-
pants marked a continuous scale to answer the questions “How do you feel right now” (1-very bad; 100-very good) and “How motivated are you to perform the following task?” (1-not at all motivated; 100-very much motivated), to make sure these aspects did not influence the predicted effects.

In the second part of the experiment, participants worked on the ostensibly unrelated wheel-of-fortune task that was identical to the one used by Aarts and colleagues (Aarts et al., 2005, Experiment 3), as explained in the introduction. The participant and the computer both rotated a differently shaded gray square across a path of white tiles on the computer screen. When asked to stop the rotation of their square by pressing a designated stop-key, participants saw that one of the white tiles was colored black (see Figure 3.1). This black tile could possibly represent the stop location of their square or that of the computer. Experiences of agency over this ambiguous event were measured by asking, after each of 16 trials, “To what extent do you think you let the square stop on the indicated location?” (1-not at all me; 10-absolutely me). To prompt the false representation of this future outcome, and thereby trigger agency illusions, in 8 of the 16 trials the stop-location was primed subliminally. This priming was realized by having the stop position light up for 34 ms in the gray-shade of the participant’s square before the stop-signal was displayed.

![Figure 3.1. Illustration of the wheel-of-fortune paradigm (adopted from Aarts et al., 2005)](image)

Funnel debriefing in all studies confirmed that participants were not aware of the priming or the purpose of the study. Unlike in previous research (see Jones, de-Wit, Fernyhough, & Meins, 2008), we did not find an effect of gender on illusions of agency in any of the reported studies.
Results and Discussion

As expected, participants who generated upward counterfactuals were more likely to report wanting to do something different the next time they encounter a similar situation ($M = 5.90$, $SD = 1.12$), compared to those who generated downward counterfactuals ($M = 4.96$, $SD = 1.79$), $t(43) = 2.05$, $p < .05$, $d = .63$. This indicates that people in the upward counterfactuals condition were considering, and maybe even planning, more concrete goal-conducive actions relative to people in the downward counterfactuals condition.Reported mood and motivation did not differ between the two recall conditions, $ps > .20$, or change the effect of counterfactual thoughts on experienced agency.

Illusions of Agency. In order to examine whether upward counterfactuals reduced illusions of agency compared to downward counterfactuals, we first averaged agency self-reports over the 8 trials with an outcome prime and the 8 trials without an outcome prime (Aarts et al., 2005). A repeated-measures ANOVA on those scores revealed that participants claimed somewhat more agency when an outcome prime was present ($M = 4.40$, $SD = 1.76$) than when it was absent ($M = 3.88$, $SD = 1.89$), $F(1,43) = 3.05$, $p = .088$, $\eta_p^2 = .07$. Importantly, this marginally significant main effect was qualified by a significant interaction effect with counterfactual thought, $F(1,43) = 7.10$, $p = .01$, $\eta_p^2 = .14$. Planned contrasts showed that, in line with predictions, participants who generated upward counterfactuals in response to their goal-failure did not claim significantly more agency after the presentation of an outcome prime ($M = 3.75$, $SD = 1.62$) relative to trials where no prime was presented ($M = 3.98$, $SD = 2.00$), $F(1,43) = 0.38$, $p = .54$, $\eta_p^2 = .009$. In contrast, participants who generated downward counterfactuals reported significantly more agency on trials with an outcome prime ($M = 4.92$, $SD = 1.72$) than on trials without an outcome prime ($M = 3.81$, $SD = 1.82$), $F(1,43) = 10.94$, $p < .01$, $\eta_p^2 = .20$.

Study 3.1 thus provides the initial demonstration that upward counterfactuals can reduce illusions of agency. After generating upward counterfactuals, participants reported wanting to act differently in a similar future situation. Presumably, this resulted in a focus on specific goal-conducive future actions that decreased illusions of agency. In Study 3.2 we tested whether it really is the future-oriented preparatory function of upward counterfactuals that reduces agency illusions. To do that, we manipulated the repeatability of the failure situation to which upward counterfactuals were generated.
Study 3.2

In Study 3.1 we found preliminary support for the assumption that upward counterfactual thinking lets people focus on specific behaviors that could potentially lead to the successful attainment of a goal in the future. Behaviors that may reoccur in the future are more probable and consequently elicit more low-level or concrete construals (Wakslak et al., 2006). For instance, it is reasonable to assume that someone who recalls failing in an exam and expects to take another exam in his life might generate a more specific upward counterfactual such as “If only I had studied every morning before class”. If it was however a once in a lifetime situation, a less concrete counterfactual thought such as “If only I studied more” is likely to result. Thus, together with others, we assume that part of the behavior-regulative function of counterfactuals lies in the anticipation of being able to set things right the next time one gets a chance to pursue one’s goal (Epstude & Roese, 2008, 2011; Markman et al., 1993). If this opportunity of repeatability is not given, the planning of actions should be less concrete and illusions of agency might still result.

To test this, in Study 3.2 we had participants do a flanker task in the laboratory and then gave them negative feedback on their performance. We assumed that in general, participants would be unlikely to expect such a situation again in the future if not explicitly told so. Therefore, after giving the failure feedback, we told half of them (repeaters) that they would be able to perform the flanker task again at the end of the experiment, whereas the other half (no-repeaters) did not receive any such information. Then all participants generated upward counterfactuals about the failure situation. We expected that illusions of agency in the wheel-of-fortune task after upward counterfactual thinking should be more reduced for repeaters than for no-repeaters, because only in this case would the focus shift away from mere outcomes and towards concrete action planning.

Method

Participants and Design. Participants were 30 undergraduates from the University of Amsterdam (10 men, $M_{age}$ = 20 years) who took part in the experiment and were randomly assigned to either the repeaters or no-repeaters condition.

Materials and Procedure. Participants were informed that they would take part in two unrelated studies investigating executive functioning and human computer interaction. First, they received an arrow-version of the Eriksen Flanker
Task (Eriksen & Eriksen, 1974). Their goal was to respond as quickly and as accurately as possible to a target arrow pointing left or right in the center of the screen that was flanked by two or more distractor arrows. Each trial began with a fixation cross in the middle of the screen (500 ms), followed by distractor arrows (100 ms) that were presented left, right, above, and/or below the fixation cross and could either have the same (congruent) or the opposite (incongruent) orientation as the target. Lastly, the target arrow was presented in the middle of the screen (50 ms). Then participants saw a blank screen until they made a response. In total, participants worked on 76 trials (26 congruent, 50 incongruent).

Before the beginning of the task, participants could make some bogus choices as to whether they wanted big versus small arrows, slow versus fast presentation, short versus long duration of the task, and the color of the background (black, red, or purple). These choices were framed as possibly having up- or downsides for performance (e.g. they were told that choosing a short task would result in a less accurate measurement of their abilities, but would cost less effort). In fact, all participants received the same type of task (except for the color of the background, which was adjusted to participants’ choices), such that duration, size and speed of the arrows were identical. We believed that by giving them the impression of being able to exert some influence on the make-up of the task, the generation of upward counterfactuals would be easier and more productive (e.g. “If only I had chosen larger arrows”).

Upon completion of the task, all participants were given bogus feedback explaining that they did not reach their performance goal and scored among the worst 15% of the students we tested so far. We checked whether participants understood the feedback by asking them to answer “Did you perform worse or better than most people on the task?” by putting a mark on a continuous line from 1 (worse) to 100 (better). Together with the bogus feedback, we also manipulated the repeatability of the situation. Repeaters were told that they would get a chance to do the flanker task again at the end of the experiment, while no-repeaters did not receive any further information. After the negative feedback, we asked participants to generate upward counterfactuals (what they could have done to improve their performance), as in Study 3.1. Then, also like in Study 3.1, participants answered questions concerning their mood and motivation, followed by the wheel-of-fortune task.

Results and Discussion
Fooling the Feeling of Doing

Repeaters and no-repeaters rated their performance level equally \( (M = 21.45, SD = 21.83) \), \( p > .83, ns. \) Importantly, choice of background color did not significantly influence performance on the flanker task. Also, performance did not differ between the two repeatability conditions; neither did mood and motivation \( (ps > .15) \), and these variables did not change the effect on reported agency.

**Illusions of Agency.** Reported experiences of agency were computed in the same way as in Study 3.1. As expected, experiences of agency were heightened for trials that included an outcome prime \( (M = 4.93, SD = 1.75) \) compared to trials without an outcome prime \( (M = 3.77, SD = 1.44) \), \( F(1,28) = 9.29, p < .01, \eta_p^2 = .25. \) Importantly, this effect differed significantly between conditions, \( F(1,28) = 4.63, \ p = .04, \eta_p^2 = .14. \) Planned contrasts showed that after the generation of upward counterfactuals, repeaters claimed the same amount of agency for trials with an outcome prime \( (M = 4.50, SD = 1.39) \) and without an outcome prime \( (M = 4.14, SD = 1.41) \), \( F(1,28) = 0.43, \ p = .52, \eta_p^2 = .02. \) In contrast, illusions of agency remained prominent for no-repeaters in that they claimed significantly more agency for trials with an outcome prime \( (M = 5.43, SD = 2.03) \) than without an outcome prime \( (M = 3.35, SD = 1.42) \), \( F(1,28) = 12.67, \ p < .01, \eta_p^2 = .31. \) In sum, Study 3.2 demonstrated that upward counterfactuals only reduced illusions of agency if participants expected to encounter the failure situation again. When they did not expect to re-encounter the same situation, agency illusions were prominent, despite the generation of upward counterfactuals.

**General Discussion**

Illusions of agency occur when people are prepared to process mental representations of future outcomes (Van der Weiden et al., 2010). The representation of outcomes is reduced when people move from a high-level behavioral focus on outcomes to a lower level of abstraction focusing on specific actions (Vallacher & Wegner, 1987). In the present research we drew on previous findings showing that upward counterfactuals let people focus relatively more on the actions required to reach a goal rather than on the outcomes (Epstude & Roese, 2011; Smallman & Roese, 2009). Connecting the two largely separated research areas of agency illusions and counterfactual thinking, we hypothesized that upward counterfactuals can reduce the susceptibility to illusions of agency.

Our hypothesis was confirmed in two studies. Thinking about what might have been, and thus which actions might lead to the attainment of an unfulfilled
goal, reduced illusions of agency in the wheel-of-fortune paradigm (Study 3.1). Importantly, we could qualify this effect by showing that these upward counterfactuals primarily reduce illusory agency if people expect to be able to set things right in a similar future situation (Study 3.2). In the personal goal-failure situation of Study 3.1, most participants presumably expected a future opportunity to reach their goal. However when participants fail in a situation that they are unlikely to encounter again, as in the flanker task in Study 3.2, manipulated repeatability of the situation seems to be crucial for the generation of specific action plans or behavioral intentions. Therefore, only when people expect a repetition of the goal-failure situation do upward counterfactuals reduce the excessive claiming of agency after outcome priming. We could furthermore rule out that general mood or general motivation influenced the found effects.

Our findings suggest that cognitive consequences of counterfactual thinking can directly influence people’s susceptibility to agency illusions. Concern with low-level actions lets people dismiss concern with their overarching goal and sustains focus on behaviors that make successful goal pursuit possible in the future. This focus on behavioral intentions and actions results from the generation of upward counterfactuals. These thoughts appear to be incompatible with the primed representation of future outcomes in the wheel-of-fortune paradigm. This incompatibility can reduce the susceptibility to illusions of agency over external events. Future research may investigate other factors that increase a focus on actions instead of outcomes, such as implementation intentions (Gollwitzer & Brandstätter, 1997), and their further influence on illusions of agency.

Overall, we believe that the present findings have important applications for the research fields on agency or authorship and counterfactual thinking. Firstly, the evidence obtained in the two studies might help to explain why these illusions occur. Secondly, they imply why the reduced susceptibility to illusions of agency following upward counterfactuals might be functional for effective behavior regulation.

First, we will consider the potential implications for the field on agency. On the one hand, one could view the present findings as a mere cognitive side-effect of upward counterfactuals that trigger a lower level of behavioral abstraction and then influence experiences of agency. However, the general prominence of agency illusions might be an intriguing phenomenon in itself. Perceiving and integrating agency cues, even if this means that one is likely to
err on the illusory side of personal causation, can sustain people’s feeling of competence as agents in the world (DeCharms, 1968). Such a feeling of competence is desirable, as it is an important determinant of intrinsic motivation (Vallerand & Reid, 1984; Wegner, 2004). Overall, people may be motivated to identify the self as an intentional and controlling agent, hence the common over-sensitivity to cues signaling personal causation or agency in the external world (cf. Deci & Ryan, 1980; White, 1959). One interpretation of the present results may be that the spontaneous planning of goal-conducive future behavior already yields a feeling of control. The counterfactual thoughts emphasize that one could have done otherwise. When this experience is not given, people seem to be in need of perceiving control over actions and events in the world, as reflected in the excessive claiming of agency.

Thus, it might be that illusions of agency are generated as a means to compensate for an obstructed sense of personal causation, authorship and control. Related to this notion, a feeling of personal causation is particularly vital after a self-inflicted goal-failure, as tested in the present studies. In this case, upward counterfactuals can help people regain a feeling of control over their own outcomes, making illusory perception of causation dispensable. Goal-failure can, however, also be due to other people’s actions or to uncontrollable factors (Girotto, Legrenzi, & Rizzo, 1991; Roese & Olson, 1995). Then the influence of upward counterfactuals on illusions of agency is likely to change. One could expect that illusions of agency are prominent no matter what type of counterfactual thought follows an uncontrollable goal-failure. Here too, enhanced experiences of agency might help people regain a sense of purposiveness over their environment. However, the opposite prediction is also possible. The experience of uncontrollability might dampen people’s overall ability to perceive effective action and hence result in an overall reduction in the claiming of authorship (e.g. learned helplessness; see Maier & Seligman, 1976). Whereas the present findings do not permit any specific predictions, such an extension of the present experimental investigation certainly has the potential to yield interesting insights for theory and research.

Lastly, we want to speculate on why it might be functional that upward counterfactuals reduce illusions of agency. Consider, for example, the case of a person who misses a plane and then counterfactually thinks: “If only I had taken the train to the airport, then I would have not been stuck in traffic for an hour”. Counterfactual thoughts are functional if they help the person to identify instrumental behavior (e.g. taking the train) to attain the goal in a similar future situation. Being susceptible to illusions of agency in the causation of
events might be harmful in such cases. For instance, people might not follow the identified strategy if illusions of agency make them erroneously conclude that they have control over an uncontrollable event (e.g. holding tight their lucky charm to make the traffic jam go away; see Pronin, Wegner, McCarthy, & Rodriguez, 2006). Yet, when a counterfactual thought does not engender an identifiable strategy to attain one’s goal in the future, having illusions of agency might be beneficial for motivation and overall well-being (cf. Seligman, 2006).

In sum, in the present research we aimed to identify circumstances that give rise to or diminish illusions of agency. Counterfactual thoughts not only have a significant impact on behavior regulation, but they also influence people’s experience of agency. Although it remains to be investigated whether reductions in agency illusions are a cognitive byproduct of behavior regulation - or whether they alternatively fulfill an important motivational function in its own right - we can conclude that counterfactual thought and illusions of agency are inextricably intertwined.