Questioning the preparatory function of counterfactual thinking

Abstract

Why do individuals mentally modify reality (e.g., "If it hadn't rained, we would have won the game")? According to the dominant view, counterfactuals primarily serve to prepare future performance. In fact, individuals who have just failed a task tend to modify the uncontrollable features of their attempt (e.g., "If the rules of the game were different, I would have won it"), generating counterfactuals that are unlikely to play any preparatory role. By contrast, they generate prefactuals that focus on the controllable features of their ensuing behavior (e.g., "If I concentrate more, I will win the next game"). Here, we test whether this tendency is robust and general. Studies 1a and 1b replicate this tendency and show that it occurs regardless of whether individuals think about their failures or their successes. Study 2 shows that individuals generate relatively few controllable counterfactuals, unless explicitly prompted to do so. These results raise some questions regarding the generality of the dominant view according to which counterfactuals mainly serve a preparatory function.

Keywords: Counterfactuals, preparatory hypothesis, prefactuals, advice.

Questioning the preparatory function of counterfactual thinking

Individuals often mentally modify everyday occurrences (e.g., "If it hadn't rained, we would have won the game"), historical events (e.g., "If it hadn't rained during the night of Waterloo, the future of Europe would have been different"), and even fiction (e.g., "If it hadn't rained the night of the concert, maybe the protagonist of Joyce's "*The mother*" would not have ruined her daughter's career"). That is, individuals often think *counterfactually* (for a recent review, see Byrne, 2016). Why do they do it?

Although several functions have been suggested for the ability to produce counterfactual thoughts (see Byrne, 2016), the main hypothesis that purports to answer this question has been dubbed the "preparatory theory." It states that: "the primary function of counterfactual thinking centers on the management and coordination of ongoing behavior. Thinking about what might have been influences performance and facilitates improvement." Accordingly, "Counterfactual thoughts are typically activated by a failed goal, and they specify what one might have done to have achieved that goal" (Epstude & Roese, 2008, 169-170; see also Epstude & Roese, 2011, Kray, Galinsky & Markman, 2009; Markman, Gavanski, Sherman & McMullen, 1993; Roese, 1997). This prediction of the preparatory hypothesis is intuitively compelling. Indeed, if counterfactuals essentially serve to regulate behavior and improve performance, the typical condition in which they are produced should be a failure, and their typical content should be something that the agent of the failure might have done to avoid it, and will not do again in similar future endeavors. Is this prediction empirically supported?

2

As for the conditions that typically elicit counterfactual thoughts, evidence exists that individuals tend to produce more counterfactuals after a failure than after a success (e.g., Roese & Hur, 1997; Roese & Olson, 1995). As for the content of counterfactuals, the evidence is mixed. According to the preparatory hypothesis, individuals who think about their failures should generate relatively more modifications focusing on what they might have done (e.g., "If I had used a different strategy, I would have won the game") than modifications focusing on how the circumstances might have been different (e.g., "If the rules of the game were different, I would have won it"). In other words, they should produce *controllable* counterfactuals more often than *uncontrollable* ones. Here, we define controllable counterfactuals as counterfactuals that refer to factors that are under the participant's control, that is, factors that the participant could modify to improve performance in the next iteration of the task. Uncontrollable counterfactuals are defined as counterfactuals that refer to factors that are not under the participant's control, and thus could not be modified to improve performance in the next iteration of the task.

By this definition, it is clear that only controllable counterfactuals can play a useful preparatory role (e.g., individuals can try using a different strategy in the next attempt, but they cannot change the rules of the game). When individuals read a scenario about a protagonist fails who to reach her goal, they behave as predicted by the preparatory hypothesis: They modify the elements of the scenario that were under the protagonist's control rather than uncontrollable aspects of reality (e.g., Girotto, Legrenzi, & Rizzo, 1991; McCloy & Byrne, 2000). However, scenario readers do not have the opportunity to think about failures that happened to them. And so their counterfactuals do not obviously serve the primary purpose of improving their own future performance. In fact, if one wants to provide a direct test of the preparatory hypothe-

sis' predictions, one should investigate counterfactual thinking in individuals who have just experienced an actual failure. Evidence presently reviewed suggests that, unlike scenario readers, these individuals generate counterfactuals that are not easily reconciled with the preparatory hypothesis.

In a series of studies, participants were asked to solve a task (e.g., a mathematical puzzle) and, if they failed to solve it, to think about how things could have been better for them (Girotto, Ferrante, Pighin, & Gonzalez, 2007; Pighin, Byrne, Ferrante, Gonzalez, & Girotto, 2011). Contrary to the preparatory hypothesis' prediction, these participants modified the uncontrollable features that had constrained their failed attempt, including their permanent traits (e.g., "If I had better mathematical abilities … ") and the rules of the game (e.g., "If I had been able to use some paper and a pencil …"), rather than the features that were under their control, like the strategy they used (e.g., "If I had started by multiplying the tens ….") and their attention and concentration level (e.g., "If I had concentrated better …").

Studying the hypothetical thoughts of individuals who have recently experienced a failure allows one to test a further prediction of the preparatory hypothesis: If counterfactual thinking is primarily aimed at improving the future, then when controllable alternatives are available, people should tend to use them. Our prefactual condition provides a good test of the availability of controllable alternatives and therefore if the preparatory theory is correct, people should focus on controllable events to a similar extent in the counterfactual and prefactual conditions. This prediction is clearly made in a recent article defending the preparatory function of counterfactuals:

> To the extent that an action fails to yield the desired and expected outcome, counterfactuals are activated relatively automatically and center on alternative actions that might well have brought about the desired outcome (Epstude &

Roese, 2011). Action phases situate within a regulatory loop, such that failed goal pursuit loops from the postactional phase back to a new instantiation of the predecisional phase. Prefactuals may then become part of the predecisional phase and feed into the formation of a decision (which goal to pursue) and intention (how to pursue that goal). (Epstude, Scholl, & Roese, 2016, p.49)

In other words, given the same experienced failure, imagining a better past should not differ from imagining a better future. Ferrante, Girotto, Stragà, and Walsh (2013) compared the number of thoughts participants generated that focused on controllable features of a task (e.g. how much the participant concentrates on the task) and uncontrollable features of a task (e.g. the imparted time to complete it). Among participants who had just failed a task, those who were asked to produce counterfactual thoughts (i.e. "Things would have been better if…"), tended to produce significantly fewer controllable thoughts (43%) than those who were asked to produce prefactual thoughts (i.e. "Things will be better for me in the next game if…"; 78%). In sum, compared to prefactual thinking, counterfactual thinking seems less likely to fulfill preparatory goals. Taken together, these and other results (e.g., McCrea, 2008; Petrocelli & Harris, 2011), which were obtained for different tasks and conditions, seriously challenge the preparatory hypothesis, suggesting that the primary function of counterfactuals may be other than the preparatory one.

The goal of this paper is twofold. The first goal is to provide more evidence that the relative lack of controllable counterfactual thoughts, compared to prefactual thoughts, is robust. Studies 1a and 1b aim to replicate and generalize the experiments documenting that actual failures do not elicit controllable counterfactuals. Whereas previous studies have investigated only *upward counterfactuals*, that is, hypothetical

thoughts about how things could have been better, Studies 1a and 1b also investigated *downward counterfactuals,* that is, hypothetical thoughts about how things could have been worse. According to the preparatory hypothesis, downward counterfactuals may also serve a preparatory function: In some cases, "downward counterfactual thinking ("I feel bad when I focus on how I might have done worse") motivated participants to try harder on a subsequent [...] task" (Epstude & Roese, 2008, p. 176). It is thus important to test whether or not individuals also tend to produce uncontrollable downward counterfactuals.

The second goal of this paper is to test an implicit prediction of the preparatory hypothesis. If the main function of counterfactual thinking is to prepare individuals for better performance, individuals who have experienced a failure should produce counterfactuals of the controllable sort whether or not they are explicitly prompted to generate useful thoughts. Contrary to this prediction, we posit that generating counterfactuals with the explicit aim of improving future performance is close to generating prefactuals: In both cases, individuals cannot focus on features that could not change in the future. Accordingly, in both cases individuals should produce controllable alternatives more often than when they receive the standard instruction to generate counterfactuals. To test these diverging possibilities, we compared the hypothetical thoughts of participants who had just failed a task and who received the standard generic request to imagine some modifications with those of participants who received a specific request to imagine modifications that would be useful for their own or other individuals' future performance on a similar task.

Study 1a

The aim of Studies 1a and 1b was to establish whether individuals tend to generate relatively more controllable modifications when they create prefactuals than

when they create counterfactuals, regardless of the direction of their thoughts (i.e., upward vs. downward). In Study 1a, individuals who had just failed or succeeded in solving a task were required to imagine better or worse modifications in an ensuing endeavor or in the past attempt.

Method

Participants

One hundred twenty-six undergraduate students (71 female; mean age = 24 y)[1] from the University IUAV of Venice participated in exchange for raffle tickets to win photocopy cards.

Materials and procedure

After providing informed consent, participants sat in front of a personal computer to complete a computer-based word-search puzzle. It was explained to them that the purpose of the game was to find as many words as possible in an 11-by-11 sized letter grid. Participants were shown a labeled screen shot of an example grid with accompanying instructions. The instructions explained that the hidden words could appear in any direction (vertical, horizontal, or diagonal) and that some letters may belong to more than one word. Participants typed the words into an empty box that appeared below the grid. If an entered word was one of the hidden words in the grid then clicking an "OK" button added the word to a list of discovered words that appeared to the right of the grid. A timer appeared below the list of discovered words showing the time remaining. Further instructions explained the rules of the puzzle game: that all words must be in Italian, contain five or more letters and be common nouns (i.e., not adjectives, verbs, or proper names). Points were awarded according to

---

[1] One of our aims was to replicate Ferrante *et al.*'s (2013) studies whose group size ranged between about 40 (Experiment 1) and 30 (Experiment 2) respondents. Accordingly, we have tested about 30 respondents per group in Study 1a, and about 40 respondents per group in Study 1b and Study2.

the orientation of the words in the grid, such that vertical words earned more points than horizontal ones, words with letters in reverse order earned more points than ones in the normal order, and diagonal words earned more points than horizontal and vertical words. Participants were given three minutes to find as many words as possible. They were informed that their performance on the task would determine whether they received 20 or 60 raffle tickets to win photocopy cards (value of each card = 1€, about $1.10). Independent of their performance, the number of tickets participants received was determined by whether they had been randomly assigned to a success ($n = 63$; winning 60 tickets) or failure ($n = 63$; winning 20 tickets) condition.

After the task and following the provision of failure or success feedback (i.e., participants were told that their score was below or above the reference score and, as a consequence, they won 20 or 60 tickets, respectively), participants were informed that they would shortly complete a similar word-search puzzle, and were first asked to think about their performance. They were randomly assigned to a counterfactual or a prefactual condition following their previous failure (counterfactual, $n = 32$; prefactual, $n = 31$) or success (counterfactual, $n = 33$; prefactual, $n = 30$). Those who received failure feedback were given the following instructions (text in brackets refers to the prefactual condition): "Things would have been better for me [Things will be better for me in the next game], if … Please, write at least one way in which you would complete this sentence." Those who received success feedback were given the following instructions (text in brackets refers to the prefactual condition): "Things would have been worse for me [Things will be worse for me in the next game], if … Please, write at least one way in which you would complete this sentence."

Results

Responses to the counterfactual / prefactual sentences included modifications of the characteristics of the task (e.g., "If I had more time"/ "If the problem is easier"), the participants' psychophysical status (e.g., "If I had slept more" / none given), their stable traits (e.g., "If I were good at this sort of game" / "If my verbal intelligence was trained"), the traits that they could not improve before the following task (e.g., "If I had trained more in this sort of quiz"/ "If I can train"), and contextual factors (e.g., "If I had an exam later", "If something distracts me"). These responses were coded as *uncontrollable* modifications (see Supplementary Materials for an alternative classification). Other responses included alterations to participants' strategic approach (e.g., "If I had used a better strategy" / "If I use a different tactic"), and their level of attention and concentration (e.g., "If I had paid more attention"/ "If I concentrate better"). These responses were coded as *controllable* modifications. Two independent judges, unaware of the hypotheses, classified the responses. Their agreement rate was 91%, Cohen's $k = .83$, $p < .001$. Disagreements were resolved by discussion. We discarded the data of five participants because the response they gave was ambiguous or non-informative (three participants in the success counterfactual, and one in the failure prefactual condition) or because they failed to complete the task (one participant in the failure prefactual condition). The remaining participants were distributed as follows: failure prefactual, $n = 30$; failure counterfactual, $n = 31$; success prefactual, $n = 30$; success counterfactual, $n = 30$.

Participants discovered on average 7.06 words (min = 1, max = 14) in the word-search task. A two-way between-subjects analysis of variance showed that the number of discovered words did not differ in the success ($M = 7.02$, $SD = 2.30$) vs. failure ($M = 7.10$, $SD = 2.58$; $F(1,117) = 0.04$, $p = .85$) conditions, nor in the counter-

factual (*M*= 7.10, *SD* = 2.51) vs. prefactual (*M* = 7.02, *SD* = 2.38; $F(1,117) = 0.04$, *p* = .85) conditions.

The four conditions elicited a similar mean number of modifications (see Table 1). In all studies, we report the analyses of the first responses generated on the hypothetical thinking task and of all responses elicited in each condition (i.e., proportion of controllable modifications on all the modifications generated by each participant). Following failure, respondents generated significantly more controllable thoughts in the prefactual condition than in the counterfactual one, both when we considered the first modifications, (77% vs. 35% respectively), $\chi^2(1, N = 61) = 10.48$, *p* = .001, φ = .41, and all the modifications (78% vs. 33% respectively), Mann-Whitney U = 226.5, *p* < .001, *r* = .48 (see Table 1). Likewise, following success, respondents produced significantly more controllable thoughts in the prefactual condition than in the counterfactual one, both when we considered the first modifications, (60% vs. 20% respectively), $\chi^2(1, N = 60) = 10.0$, *p* = .002, φ = .41, and all the modifications (52% vs. 24% respectively), Mann-Whitney U = 263.5, *p* = .002, *r* = .40. [2]

Moreover, in each condition, most response patterns were consistent, that is, they contained only uncontrollable or only controllable mutations (failure prefactual: 87%; failure counterfactual: 81%; success prefactual: 77%; success counterfactual: 93%).

These results replicate the results obtained when participants reflected on their failures in Ferrante *et al.* (2013), and extend them to situations in which participants reflect on their successes. In both cases, participants tend to produce more uncontrol-

---

[2] The minor difference in the framing of the counterfactual and prefactual question could have focused the prefactual question ("things will be better for me in the next game if") more on the task and the counterfactual question ("things would have been better for me if") more on the participant's experience. However, we found that even in the counterfactual condition most thoughts focused on the task. Hence, this potential confound is unlikely to have substantially affected the results.

lable modifications when they generate counterfactuals and more controllable modifications when they generate prefactuals. Only a third of the counterfactuals generated were controllable, a result that would not be predicted if counterfactuals had a preparatory function. Furthermore, individuals knew that they were about to receive another similar task to do and so their motivation to think of how they could improve should be strong (see Markman *et al.*, 1993). Yet, participants spontaneously produced few controllable counterfactuals.

Table 1. Mean number of modifications, and the percentage of first and all modifications that were controllable in the four conditions of Studies 1a and 1b.

Condition		Study 1a			Study 1b		
		Mean (SD)	First	All	Mean (SD)	First	all
Failure	Prefactual	1.80 (.76)	77	78	1.09 (.34)	61	58
	Counterfactual	1.74 (.63)	35	33	1.29 (.58)	9	8
Success	Prefactual	1.80 (.85)	60	52	1.48 (.57)	45	43
	Counterfactual	1.97 (.85)	20	24	2.07 (.77)	11	16

Study 1b

In Study 1a, participants received feedback that was unrelated to their actual performance. Thus, for some participants the feedback might have seemed out of touch with their performance. This possibility opens the question of whether the reported difference in the generation of future vs. past modifications might also occur when participants are given accurate feedback on their past performance. To answer this question, in Study 1b participants were required to produce modifications about their own actual performance on a given task. A secondary aim of Study 1b was to

generalize the results of Study 1a using a different type of task (a syllogistic task rather than a word-search game) and a different population (on-line survey respondents rather than undergraduates).

Method

Participants

Participants were 199 U.S. residents (mean age: 30 y; age range 18-67 y; 71 women) recruited using the Amazon Mechanical Turk platform. They were paid $1 to complete a syllogistic task in a limited time.

Materials and procedure

The task asked the participants to order five individuals by height on the basis of four statements describing the height relations between pairs of individuals (e.g. "Matt is shorter than Paul"; see Supplemental Materials for details). An example was provided and the participants had to complete a simple task to learn how to use the interface to order the individuals. Once they had completed this trivial task, participants were reminded of how the actual task would work, and that they would only have 20 seconds to complete it. After the task, participants were provided with accurate failure or success feedback (i.e. depending on whether they solved the task in the allotted time) and were asked to think about their performance. Participants were randomly assigned to a counterfactual or prefactual condition following their failure (prefactual, $n = 64$; counterfactual, $n = 68$) or success (prefactual, $n = 36$; counterfactual, $n = 31$). In the prefactual condition, they were informed that they would shortly complete another similar syllogistic task. They were given the same instructions as in the corresponding conditions of Study 1a.

Results

The responses were similar to those obtained in Study 1a, and were coded by

two independent judges, blind to the hypotheses, who used the same coding criterion as in Study 1a. Their agreement rate was 94%, Cohen's $k = .87$, $p < .001$. Disagreements were resolved by discussion. We discarded the data of six participants who failed to complete the task and ten participants who produced only ambiguous responses. The remaining participants were distributed as follows: failure prefactual, $n = 57$, failure counterfactual, $n = 65$, success prefactual, $n = 33$, success counterfactual, $n = 28$.

Participants generated a significantly higher number of modifications in the success conditions ($M = 1.75$, $SD = 0.72$) than in the failure conditions ($M = 1.20$, $SD = 0.49$), $F(1, 179) = 46.26$, $p < .001$, $\eta^2_p = .21$ (see Table 1). Moreover, in the success conditions, they generated a significantly higher number of counterfactual ($M = 2.07$, $SD = 0.77$) than prefactual modifications ($M = 1.48$, $SD = 0.57$), $F(1,179) = 4.88$, $p = .03$, $\eta^2_p = .03$. We have no plausible explanation for these differences. They did not emerge in the other two studies (see Table 1 and Table 2).

After a failure, the prefactual condition elicited significantly more controllable modifications than did the counterfactual condition, both when we considered the first modifications (61% vs. 9%, respectively), $\chi^2(1, N = 122) = 37.05$, $p < .001$, $\varphi = .55$, and all modifications (58% vs. 8% respectively), Mann-Whitney $U = 864.0$, $p < .001$, $r = .55$ (see Table 1). Likewise, after a success, the prefactual condition elicited significantly more controllable modifications than did the counterfactual condition, both when we considered the first modifications (45% vs. 11% respectively), $\chi^2(1, N = 61) = 8.79$, $p = .003$, $\varphi = .38$, and all modifications (43% vs. 16%, respectively), Mann-Whitney $U = 287.5$, $p = .003$, $r = .38$. Failure and success feedback elicited a similar rate of controllable counterfactuals, both when we considered the first modifications (9% vs. 11%, respectively), $\chi^2(1, N = 93) = 0.05$, $p = .82$, $\varphi = .02$, and all

modifications (8% vs. 16%, respectively), Mann-Whitney U = 850.5, p = .39, r = .09. Likewise, the two sorts of feedback elicited a similar rate of controllable prefactuals, both when we considered the first modifications (61% vs. 45% of controllable modifications, respectively), χ^2 (1, N = 90) = 2.15, p = .14, φ = .15, and all modifications (58% vs. 43%, respectively), Mann-Whitney U = 766.0, p = .10, r = .17.

As in Study 1a, most response patterns were consistent both in the failure (prefactual = 98%; counterfactual = 95%) and success (prefactual = 85%; counterfactual = 86%) conditions.

Study 1b replicates the findings of Study 1a. Participants generated more controllable modifications in the prefactual condition than in the counterfactual condition. They did so both when they had successfully solved the task and when they had failed it. The very low rate of controllable counterfactuals (12% across conditions) clearly conflicts with the predictions of the preparatory hypothesis.

Study 2

Following the preparatory hypothesis, individuals who have experienced a failure should generate controllable counterfactuals regardless of whether they are explicitly prompted to do so. The results reported by Ferrante *et al.* (2013) as well as those obtained in Studies 1a and 1b suggest a different prediction. Individuals who have experienced a failure are not likely to generate controllable counterfactuals. Controllable thoughts, however, are available to these individuals, as proved by their tendency to generate controllable prefactuals. What if these individuals are explicitly required to generate counterfactuals that could be useful for their own or other individuals' future performance on a similar task? Their situation is close to the one of individuals who must generate prefactuals: They cannot focus on the constraints that have governed their past attempt (e.g., the rules of the task) because these features

14

will not change in the following attempt. Hence, they should focus on controllable features (e.g., the strategy they used in their past attempt). In sum, individuals who have to generate counterfactuals that could be useful in an ensuing endeavor should generate controllable thoughts more often than individuals who receive the standard instruction to generate counterfactuals.

In Study 2, we tested these diverging predictions by comparing the hypothetical thoughts of participants who had just failed a task and who received four different sorts of instructions. In two conditions, they received the standard instructions to produce counterfactual or prefactual modifications. In two other conditions, they were requested to generate counterfactual modifications that would be useful for their own or other individuals' future performance on a similar task. If the main function of counterfactual thinking were to improve future performance, then participants should produce similar modifications in the four conditions. In particular, they should produce a similar rate of controllable counterfactuals when they are asked to give a piece of advice (to themselves or to other individuals) and when they receive no explicit instruction to this effect. By contrast, if the main function of counterfactual thinking is not a preparatory one, then participants should produce a higher rate of controllable thoughts in the prefactual and in the advice conditions than in the standard counterfactual one.

Method

Participants

A sample of 181 U.S. residents (mean age: 30 y; age range 18-61 y; 51 women) was recruited using Amazon Mechanical Turk. They were paid $1 to complete the syllogistic task introduced in Study 2 and to complete a hypothetical sentence.

Materials and procedure

The method was similar to that of Study 1b (see Supplemental Materials for details). Participants were randomly assigned to one of four conditions: prefactual, n = 43, counterfactual advice-self, n = 46, counterfactual advice-other, n = 43, and counterfactual, n = 49. They were asked to solve a syllogistic task. Study 2 differs from Study 1b in the following ways. 1) Participants only had 15s (in contrast with 20s) to complete the task. This change guaranteed that most participants would fail, since we were only interested in modifications following failures. 2) Accordingly, we did not ask the few participants who succeeded to produce any modifications. 3) In addition to the prefactual and counterfactual conditions, participants could also be prompted to produce counterfactual thoughts that would help them improve their per- formance on a repetition of the task (counterfactual advice-self condition), or that would help another participant improve his or her performance on the same task (see Supplemental Materials for details of the requests).

Results

Participants successfully completed the task at similar rates across conditions (prefactual = 23%; counterfactual self-advice = 15%; counterfactual other-advice = 25%; counterfactual = 18%). The participants who succeeded were not asked to pro- duce a statement. The statements of those who had failed were coded by two inde- pendent judges, blind to the hypotheses. Their agreement rate was 93.6%, Cohen's k = .88, p < .001. Disagreements were resolved by discussion. We discarded the data of four respondents in the prefactual condition and one respondent in the other-advice condition because the only response each one produced was ambiguous. The remain- ing participants were distributed as follows: prefactual, n = 29; counterfactual self- advice, n = 39; counterfactual other-advice, n = 31; counterfactual, n = 40.

Participants generated a similar mean number of modifications across conditions (see Table 2). The self-advice and other-advice counterfactual conditions elicited similar rates of controllable modifications both when we considered the first modifications (72% vs. 68%, respectively), $\chi^2(1, N = 70) = 0.14$, $p = .71$, $\varphi = .04$ and all modifications (69% vs. 60% respectively), Mann-Whitney $U = 595.5$, $p = .90$, $r = .02$ (see Table 2). Likewise, rates of controllable modifications in the two advice conditions did not differ significantly from prefactual ones: self-advice vs. prefactual first modifications (72% vs. 79%, respectively), $\chi^2(1, N = 68) = 0.50$, $p = .48$, $\varphi = .09$, all modifications (69% vs. 80%, respectively), Mann-Whitney $U = 488.5$, $p = .23$, $r = .14$); other-advice vs. prefactual first modifications (68% vs. 79%, respectively), $\chi^2(1, N = 60) = 1.03$, $p = .31$, $\varphi = .13$, all modifications (60% vs. 80%, respectively), Mann-Whitney $U = 398.5$, $p = .33$, $r = .13$.

By contrast, the prefactual condition elicited a significantly higher rate of controllable modifications than the counterfactual condition both when we considered the first modifications (79% vs. 15%, respectively), $\chi^2(1, N = 69) = 28.54$, $p < .001$, $\varphi = .64$, and all modifications (80% vs. 12%, respectively), Mann-Whitney $U = 178.5$, $p < .001$, $r = .68$. Similarly, both advice conditions elicited a significantly higher rate of controllable modifications compared to the counterfactual condition: self-advice vs. counterfactual first modifications (72% vs. 15%, respectively), $\chi^2(1, N = 79) = 25.98$, $p < .001$, $\varphi = .57$, all modifications (69% vs. 12%, respectively), Mann-Whitney $U = 297.0$, $p < .001$, $r = .60$; other-advice vs. counterfactual first modifications (68% vs. 15%, respectively), $\chi^2(1, N = 71) = 20.61$, $p < .001$, $\varphi = .54$, all modifications (60% vs. 12%, respectively) Mann-Whitney $U = 248.0$, $p < .001$, $r = .59$. In all conditions, most response patterns were consistent (prefactual = 97%; counterfactual self-advice

= 87%; counterfactual other-advice = 94%; counterfactual = 98%).[3]

As predicted, we found that participants generated many more controllable modifications not only when asked for prefactuals, but also when asked to generate counterfactuals that could be useful for their future selves or for others. This finding suggests that the failure of participants to generate such controllable modifications in the counterfactual condition does not stem from a difficulty in generating controllable counterfactuals in this task. In sum, contrary to the implicit prediction of the preparatory hypothesis, participants who were asked to give advice (to themselves or to other individuals) produced a significantly higher rate of controllable counterfactuals than participants who received no explicit instruction to this effect.

Table 2. Mean number of modifications, and the percentages of first and all modifications that were controllable in the four conditions of Study 2

Condition	Mean (SD)	Percentage	
		First	All
Prefactual	1.38 (.68)	79	80
Counterfactual self-advice	1.56 (.88)	72	69
Counterfactual other-advice	1.45 (.81)	68	60
Counterfactual	1.23 (.48)	15	12

General Discussion

In the reported studies, participants who had just completed a task imagined a different outcome to their past attempt (counterfactual condition) or to a following at-

---

[3] Because the design was not fully symmetrical (i.e. there was only one prefactual condition), we could not perform a log-linear analysis.

tempt (prefactual condition). Compared to participants who had to think about a different future, those who had to think about a different past were more likely to mentally modify uncontrollable features of their attempt. This tendency occurred both when participants had failed the task ("Things would have been better if the allocated time were longer") and when they had successfully solved it ("Things would have been worse if the allocated time were shorter"). This difference was observed both when participants received mock (Study 1a) and veridical feedback on their past performance (Study 1b). In Study 2, participants who generated counterfactuals in the standard way without an explicit purpose produced fewer controllable thoughts than those who were explicitly prompted to generate counterfactuals that would be useful either to themselves or to others. In the three studies, the effect sizes for comparisons between counterfactual conditions and other conditions were medium to large (ranging .38 to .68; see Coolican, 2009).

These results confirm and extend those obtained in earlier studies (e.g., Ferrante *et al*., 2013; Girotto *et al*., 2007; Pighin *et al*., 2011), and conflict with the predictions of the preparatory hypothesis (e.g., Epstude & Roese, 2008, 2011; Markman *et al*., 1993, Roese, 1997). If counterfactuals mainly serve to prepare future performance, they should be of the controllable sort, they should not differ from prefactuals, and they should not be affected by external prompts. Contrary to these predictions, our results show that unless participants are explicitly prompted to generate useful thoughts they often fail to generate controllable counterfactuals. By contrast, they spontaneously generate a significantly higher rate of controllable prefactuals.

Recently, Epstude and Roese (2011) have refined the preparatory function hypothesis by distinguishing between a content-neutral pathway and a content-specific pathway. The content neutral pathway corresponds to a general increase of motivation

after the generation of counterfactual thoughts. It would be very difficult to show that counterfactuals thoughts, even uncontrollable ones, do not serve such a general function. By contrast, the content-specific pathway "embodies the transmission of particular semantic information from the counterfactual to a behavioral intention to an action." (2011, p.21). Our results clearly argue against this content-specific pathway, in particular since our participants were in the ideal condition to generate controllable hypothetical thoughts—that is, thoughts that would fit with the definition of the content-specific pathway. First, participants who think about the outcome of a scenario may be uncertain about the elements that its characters might or might not control. Yet, they produce controllable counterfactuals (e.g., Pighin *et al.*, 2011). Our participants, however, completed the task themselves and thought about their own performance. Therefore, they had direct experience of the elements that were and were not under their control. Second, the participants in our experiments completed a task whose outcome largely depended on their attention and concentration level, and on the strategies they employed, instead of luck for instance. Despite these favorable conditions, our participants generated a very low rate of controllable counterfactuals (16% across studies), and did so both when they reasoned about a failure or about a success.

One potential criticism of our studies is that for participants the stakes were relatively low. Participants stood to earn (or fail to earn) only small rewards. However, in Experiment 1a, participants were incentivized by the possibility of winning photocopy cards. In fact, experiments offered in support of the preparatory hypothesis have typically used low stakes. Moreover, it is unclear why low stakes would particularly affect the preparatory function and not all potential functions of counterfactuals.

Indeed, several functions other than the preparatory one have been suggested for counterfactual thoughts: "They explain the past, prepare for the future, modulate emotional experience, and support moral judgments" (Byrne, 2016, p.136). The list of counterfactuals reported at the beginning of the paper might be used to illustrate some of these functions. In the present case, one interpretation of the data is that many of the counterfactuals generated by our participants could help them to save face. Following a failure, individuals often explain it away (e.g., Gilovich, 1983; for a review, see Tavris & Aronson, 2007). Accordingly, the typical counterfactuals produced by our participants (e.g., "Things would have been better if the allocated time were longer") could be considered as potential excuses of their failure, because they suggest that it was due to factors outside of participants' control. Such counterfactuals would belong to the family of motivated reasoning: reasoning that does not aim at accuracy, but at defending a pre-established point of view (Kunda, 1990; Mercier & Sperber, 2011). In the present case, the participants would be defending their competence in the face of failure.

Following a failure, our participants modified uncontrollable features of their attempt. Uncontrollable counterfactuals, however, are not the only sort of counterfactuals that can play a self-defensive role. In our studies, we have used tasks that did not involve any training or practice session. When tasks involve such sessions, individuals who have experienced a failure generate counterfactuals that do focus on their preparatory effort. For example, McCrea (2008) found that undergraduates who had failed an exam and had reported a lack of study effort tended to produce counterfactual thoughts about studying (e.g., "If I had studied more I could have done better"). Importantly, these undergraduates experienced an increase in self-esteem as a result of generating such counterfactuals, but a decrease in motivation to adequately prepare

for the next exam in the class. This finding shows that, along with uncontrollable counterfactuals, controllable counterfactuals can help save face following negative performance. This finding also shows that counterfactuals, including controllable ones, might have a detrimental effect on learning.

A series of studies by Petrocelli and colleagues (e.g., Petrocelli & Harris, 2011; Petrocelli, Seta & Seta, 2013; Petrocelli, Seta, Seta, & Prince, 2012; see also Kruger, Wirtz, & Miller, 2005) confirmed this finding. For example, students who generated counterfactuals about a failed item of a multiple-choice practice exam (e.g., "If I had read the answer choices more thoroughly…") were subsequently less likely to study exam topics related to that item than exam topics related to items for which they had not generated counterfactuals (Petrocelli *et al.*, 2012). In other words, counterfactuals, including controllable ones, may provide an erroneous sense of competence (e.g., "I mastered the topic but I did not read the question carefully") which in turn may hinder efforts toward improvement. Results of this sort corroborate the view that counterfactuals may improve future performance only to the extent that they indicate the correct causal antecedent to the negative outcome (e.g., lack of knowledge rather than a simple oversight of an item's answers), and that individuals have the ability and motivation to change their behavior in the direction prescribed by counterfactual modifications (see Petrocelli & Harris, 2011). Along with the strikingly low rate of controllable counterfactuals reported in the present studies, the finding that counterfactuals are often dysfunctional are difficult to reconcile with the preparatory function hypothesis.

The reported tendencies indicate that future research should pay more attention to the potential social functions of counterfactual thought. In any case, it is important to bear in mind that testing functional hypotheses takes more than demonstrat-

ing that a given cognitive mechanism (here the mechanisms that generate counterfactual thoughts) has a given effect (preparing for the future or explaining away the past, say). Any adaptation is bound to have a multitude of effects as by-products, such as the noise our heart produces when beating. Functional hypotheses must be supported instead by evidence of a particularly good match between the hypothesized function of a mechanism and its working (see Williams, 1966, and regarding functionalism in general, Elster, 1989). For instance, more exigent tests of the functional hypothesis might involve showing that engaging in counterfactual thought better prepares individuals for future actions than other cognitive activities would, or that the mechanisms that allow counterfactual thought have features that are explained better by the preparatory function hypothesis than by other functional hypotheses. Obviously, the hypothesis that counterfactuals serve a social function would require the same type of evidence. At the moment, no theory of counterfactual thought seems to have enough arguments to support a functional hypothesis.

Acknowledgments

References

Byrne, R. M. J. (2016). Counterfactual thought. *Annual Review of Psychology*, in press.

Coolican, H. (2014). *Research methods and statistics in psychology*. London: Psychology Press.

Elster, J. (1989). *The cement of society.* Cambridge: Cambridge University Press.

Epstude, K., Scholl, A., & Roese, N. J. (2016). Prefactual thoughts: Mental simulations about what might happen. *Review of General Psychology*, 20(1), 48-56.

Epstude, K. & Roese, N. J. (2008). The functional theory of counterfactual thinking. *Personality and Social Psychology Review*, *12*, 168-192

Epstude, K., & Roese, N. J. (2011). When goal pursuit fails. *Social Psychology*, 42(1), 19–27.

Ferrante, D., Girotto, V., Stragà, M., & Walsh, C. (2013). Improving the past and the future: A temporal asymmetry in hypothetical thinking. *Journal of Experimental Psychology: General*, 143, 23-27

Gilovich, T. (1983). Biased evaluation and persistence in gambling. *Journal of Personality and Social Psychology*, 44, 1110–1126.

Girotto, V., Ferrante, D., Pighin, S., & Gonzalez, M. (2007). Post-decisional counterfactual thinking by actors and readers. *Psychological Science*, 18, 510-515.

Girotto, V., Legrenzi, P., & Rizzo, A. (1991). Counterfactual thinking: The role of events controllability. *Acta Psychologica*, *78*, 111-133.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93(2), 136–153.

Kray, L., Galinsky, A., & Markman, K. (2009). Counterfactual structure and learning from experience in negotiations. *Journal of Experimental Social Psychology*, 45, 979-982.

Kruger, J., Wirtz, D., & Miller, D. (2005). Counterfactual thinking and the first instinct fallacy. *Journal of Personality and Social Psychology, 88*, 725–735.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480–498.

Markman, K. D., Gavanski, I., Sherman, S. J., & McMullen, M. N. (1993). The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology, 29,* 87-109.

McCloy, R. & Byrne, R.M.J. (2000). Counterfactual thinking about controllable actions. *Memory & Cognition, 28*, 1071–1078.

McCrea, S. M. (2008). Self-handicapping, excuse making, and counterfactual thinking: Consequences for self-esteem and future motivation. *Journal of Personality and Social Psychology, 95*, 274-292.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.

Morris, M. W., & Moore, P. C. (2000). The lessons we (don't) learn: Counterfactual thinking and organizational accountability after a close call. *Administrative Science Quarterly*, 45, 737–765.

Petrocelli, J. V., & Harris, A. K. (2011). Learning inhibition in the Monty Hall Problem: The role of dysfunctional counterfactual prescriptions. *Personality and Social Psychology Bulletin*, 37, 1297–1311.

Petrocelli, J. V., Seta, C. E., & Seta, J. J. (2013). Dysfunctional counterfactual thinking: When simulating alternatives to reality impedes experiential learning. *Thinking and Reasoning*, 19, 205- 230.

Petrocelli, J. V., Seta, C. E., Seta, J. J., & Prince, L. B. (2012). "If only I could stop generating counterfactual thoughts": When counterfactual thinking interferes with academic performance. *Journal of Experimental Social Psychology*, 48, 1117-1123.

Pighin, S., Byrne, R.M., Ferrante, D., Gonzalez, M., & Girotto, V. (2011) Counterfactual thoughts by experienced, observed and narrated events, *Thinking and Reasoning,* 17, 197-211.

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

Roese, N.J. & Hur, T. (1997). Affective determinants of counterfactual thinking. *Social Cognition*, *15*, 274-290.

Roese, N.J. & Olson (1995). Outcome controllability and counterfactual thinking. *Personality and Social Psychology Bulletin, 21*, 620-628.

Tavris, C. & Aronson, E. (2007). *Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions, and hurtful actions*. Boston, MA: Houghton Mifflin.

Williams, G. C. (1966). *Adaptation and natural selection*. Princeton: Princeton University Press.