



The more, the merrier: Treatment frequency influences effectiveness perception and further treatment choice

Itxaso Barberia¹ · Fernando Blanco² · Javier Rodríguez-Ferreiro³

Accepted: 15 October 2020 / Published online: 29 October 2020
© The Psychonomic Society, Inc. 2020

Abstract

Causal illusions have been postulated as cognitive mediators of pseudoscientific beliefs, which, in turn, might lead to the use of pseudomedicines. However, while the laboratory tasks aimed to explore causal illusions typically present participants with information regarding the consequences of administering a fictitious treatment versus not administering any treatment, real-life decisions frequently involve choosing between several alternative treatments. In order to mimic these realistic conditions, participants in two experiments received information regarding the rate of recovery when each of two different fictitious remedies were administered. The fictitious remedy that was more frequently administered was given higher effectiveness ratings than the low-frequency one, independent of the absence or presence of information about the spontaneous recovery rate. Crucially, we also introduced a novel dependent variable that involved imagining new occasions in which the ailment was present and asking participants to decide which treatment they would opt for. The inclusion of information about the base rate of recovery significantly influenced participants' choices. These results imply that the mere prevalence of popular treatments might make them seem particularly effective. It also suggests that effectiveness ratings should be interpreted with caution as they might not accurately reflect real treatment choices. Materials and datasets are available at the Open Science Framework [<https://osf.io/ctjts/>].

Keywords Contingency learning · Causal illusion · Illusory correlation · Illusion of causality · Cue-density effect · Pseudomedicines · Pseudoscience

Introduction

When faced with a health issue, individuals must choose a treatment approach. Previous studies have identified patients' values and belief systems, along with education level and health status, as significant predictors of treatment selection (Astin, 1998). However, cognitive phenomena underlying (inadequate) treatment choice have been frequently overlooked.

Causal illusions are defined as the “perception of a causal relationship between events that are actually unrelated” (Matute et al., 2015, p. 1), one of these events having the role of a potential cause and a second event referring to the outcome or potential effect. This phenomenon has been routinely studied through simple contingency learning computer tasks in which participants are asked to judge the causal relationship between two events. For instance, a frequent scenario comprises determining the relationship between an experimental drug and recovery from a disease (Blanco, Matute, & Vellido, 2013; Matute, Yarritu, & Vellido, 2011). During a typical experiment, participants receive correlational information (e.g., about the presence/absence of treatment and about the health outcome) in a sequential manner, and they have to report their impression regarding the effectiveness of the remedy. If the participants report that the remedy is effective, we conclude that they are victims of a causal illusion (Matute et al., 2015), given the null one-way contingency between the remedy and the recovery from the disease.

Causal illusions seem to proliferate if the candidate cause (e.g., the remedy) is present with a high probability (Allan &

✉ Javier Rodríguez-Ferreiro
rodriguezferreiro@ub.edu

¹ Departament de Cognició, Desenvolupament i Psicologia de l'Educació, Institut de Neurociències, Universitat de Barcelona, Passeig de la Vall d'Hebron, 171, Barcelona 08035, Spain

² Departamento de Psicología Social, Facultad de Psicología, Universidad de Granada, Campus Universitario de Cartuja, Granada 18071, Spain

³ Grup de Recerca en Cognició i Llenguatge, Departament de Cognició, Desenvolupament i Psicologia de l'Educació, Institut de Neurociències, Universitat de Barcelona, Passeig de la Vall d'Hebron, 171, Barcelona 08035, Spain

Jenkins, 1983; Blanco et al., 2013; Matute et al., 2011), a phenomenon often termed *cue-density* or *cause-density effect*. Consistent with this idea, Blanco et al. (2013) found that, particularly for a situation in which the rate of spontaneous recovery was high, the higher the proportion of patients who received the fictitious remedy, the stronger the causal illusion the participants developed at the end of the experiment, even if the actual contingency was null (i.e., the remedy did not work).

The importance of causal illusions is tightly connected with the proposal of it being the potential cognitive background for a variety of causal misbeliefs that invade our society, such as those favoring the success of paranormal or pseudoscientific beliefs (Blanco, Barberia, & Matute, 2014, 2015; Griffiths, Shehabi, Murphy, & Le Pelley, 2018; Matute et al., 2011, 2015). In this sense, a parallel has been established between the circumstances where pseudomedicines are typically used and the conditions under which causal illusions are observed in these laboratory tasks. For instance, as noted by Blanco et al. (2014), specific circumstances such as the lack of side effects might promote the use of pseudomedicines (e.g., homeopathy). In line with the cue-density effect, an increase in the pseudomedicine popularity would, in turn, intensify its perceived effectiveness.

However, in their attempt to capture the circumstances in which pseudomedicines proliferate, previous experiments might have bypassed an important aspect present in real contexts. In most cases, the consumer will more likely face the decision of selecting one treatment over other candidate treatments, rather than choosing between taking the candidate remedy versus not taking anything at all. In this sense, the typical laboratory experiments on causal illusions present participants with information of the recovery rate of patients both when they take the candidate remedy and when they do not take any remedy. In contrast, in real contexts, the information that people will frequently compare when deciding what treatment to choose will be the rate of recovery conditional on taking different alternative treatments (e.g., the rate of recovery when taking a pill vs. when taking a syrup).

Another possible limitation of prior research that we attempt to address here refers to the dependent variable used to measure the extent to which the participants have perceived a causal relation between the candidate remedy and recovery of the patients. Previous studies have measured causal illusions with effectiveness or causal questions along the lines of “To what extent do you think that the remedy has been effective in healing the ailment?”, which are usually required to be answered by providing a number on a 0 (not effective at all) to 100 (totally effective) numerical scale (e.g., Barberia, Vadillo, & Rodríguez-Ferreiro, 2019; Blanco et al., 2013). The comparison between results obtained with this customary measure and other probe questions has revealed interesting observations regarding how the wording of the question modulates the participants’ responses. For example, when the

question is formulated in purely causal terms “Does the medicine cause the healings?”, the illusion of causality seems smaller than when it is asked in a more natural way, in terms of effectiveness: “Is the medicine effective?” (Matute et al., 2011). Additionally, it is known that different question wordings foster the use of different pieces of information – whereas causal and preparatory judgments are sensitive to contingency, prediction questions are based on the conditional probability of the outcome (Blanco, Matute, & Vadillo, 2010).

Nevertheless, requiring the volunteers to translate their causal impression into a numerical scale, irrespective of the specific question employed, might not be the best way to study real-life implications of causal detection. As noted by Perales, Catena, Cándido, and Maldonado (2017, p. 39), “people are very rarely asked to make [causal] judgments in real life; more often, they are asked to make choices or to identify where to intervene in their environments in order to obtain a desired result.” Moreover, it is not clear that causal estimations result in behavior change, although most previous research using numerical causal judgments rests on this assumption when extending findings to real-life consequences. It is, thus, possible that causal estimations inform or constrain the decision-making process, so that judgment and decision align. But it is also possible that they do not coincide, or that decisions are independent of causal estimations. Taking all this into account, in addition to using the customary effectiveness questions, in our experiments we explicitly asked our participants to select which of the treatment options they would use if faced with the pertinent health condition (*action question*). We consider this measure to be less ambiguous and more informative for our purposes, as it directly refers to the ultimately relevant issue of treatment choice. Materials and datasets for our experiments are available at the Open Science Framework (OSF) [<https://osf.io/ftjts/>].

Experiment 1

Participants in this experiment were exposed to a series of several fictitious patients who could take one of two remedies. Both remedies produced recovery with the same probability but one of them was more frequently administered than the other. Note that, unlike the typical experiment on causal illusions, here participants did not receive any information about the baseline rate of recovery without taking any remedy, and, therefore, in this context there is no clear answer as to what the absolute effectiveness of each treatment is. Importantly, since both remedies produced recovery with the same probability (i.e., $P(\text{recovery} | \text{High Frequency Remedy}) = P(\text{recovery} | \text{Low Frequency Remedy})$), the perceived effectiveness of the two treatments, whatever its absolute value, should have not differed if participants rely on these conditional probabilities when forming their impressions. However,

our prediction is that, when the task involves comparing two different treatments, and analogous to previous results on the cue-density effect, participants would tend to evaluate the more frequent treatment as more effective than the less frequent one, even when both of them are associated with the same rate of improvement in the observed patients. Following our hypothesis regarding perceived effectiveness, we expected our participants to select the most frequent treatment, even though the recovery probabilities of patients using the frequent and infrequent treatments were the same.

Method

Participants A power analysis conducted in GPower (Erdfelder, Faul, Buchner, & Lang, 2009) indicated that a minimum of 34 participants were required to detect a medium-size effect of $d = 0.5$ with a statistical power of 0.8. Forty psychology undergraduate students took part in the study in exchange for course credits (37 females, mean age = 19.73 years, $SD = 2.37$). Participants provided signed informed consent before their participation. The study protocols were approved by the ethics committee of the University of Barcelona.

Design and procedure Table 1 shows the specific frequencies of combinations of the two events used in this experiment. Thirty-six patients took one of the remedies, out of which 27 recovered and nine did not, and 12 patients took the other remedy, out of which nine experienced recovery, whereas the remaining three did not. Therefore, 75% of the patients experienced recovery, irrespective of the treatment received, but one treatment was used more often (75% of the trials) than the other (25% of the trials).

The task was an adaptation of the typical one employed in the study of causal illusions (Blanco et al., 2013; Matute et al., 2011). The initial screen provided participants with the task instructions (see Appendix). In short, participants were asked to imagine that they were exploring the effectiveness of two herbal infusions from the Amazon against headaches. It was further explained to them that they would have at their

disposal medical records of patients who suffered from a headache episode and took either herbal infusion A or herbal infusion B. For each medical record, they would have to decide whether they thought the patient's headache would disappear within the next 2 h or not. After their prediction, they would be informed whether each patient overcame the headache or not. They were further told that, once they observed several patients, we would ask them some questions, and that their goal was to determine to what extent each of the infusions was effective. Participants observed a total of 48 patients, from which 36 took one of the infusions (*high-frequency infusion*) and the remaining 12 took the other one (*low-frequency infusion*). Both infusions were associated with a probability of overcoming the headache of .75 (as shown in Table 1). Which herb, A or B, was the high-frequency or the low-frequency herbal infusion was balanced across participants.

After observing all 48 patients, participants were presented with three questions (see Appendix). They were asked to assess the extent to which each of the herbal infusions was effective against headaches (*effectiveness questions*). Possible answers ranged on a scale from 0 (not effective at all) to 100 (totally effective). They were also required to indicate which of the two infusions they would prefer to take in case of having a headache episode (*action question*). About half of the participants answered the causal questions first and then the action question, whereas for the other half we reversed this order. Moreover, the order of the two effectiveness questions was also counterbalanced across participants.¹

Results

The ratings provided by the volunteers in response to the effectiveness questions (see Fig. 1a) were significantly higher for the high-frequency infusion (mean = 66.88, $SD = 15.59$) compared to the low-frequency infusion (mean = 58.25, $SD = 17.76$), $t(39) = 2.408$, $p = .021$, $d = .381$.

Regarding the action judgment (Fig. 1b), the proportion of participants choosing the high- (0.675) and low- (0.325) frequency remedies were significantly different from each other, $\chi^2(1) = 4.900$, $p = .027$. Therefore, participants preferred the high-frequency over the low-frequency remedy when imagining future occasions in which they had to deal with headaches.

Table 1 Frequencies of the different trial types in Experiments 1 and 2

	Patients who recovered	Patients who did not recover
Patients who took the high-frequency remedy	27	9
Patients who took the low-frequency remedy	9	3
Patients who did not take any remedy (only in the <i>With Context</i> condition, Experiment 2)	18	6

¹ For exploratory reasons, and only in Experiment 1, after the effectiveness and action questions were completed, participants were also asked to report retrospectively the number of patients who recovered during training, among both those who took the high- and those who took the low-frequency remedies (*frequency questions*, see Appendix). The responses to these questions (available at the OSF [<https://osf.io/ftjs/>]) were consistent with causal impressions as measured by the effectiveness and action questions (i.e., higher proportion of recovery reported for the high- vs. the low-frequency remedy) but, for the sake of brevity, they are not further discussed in the paper.

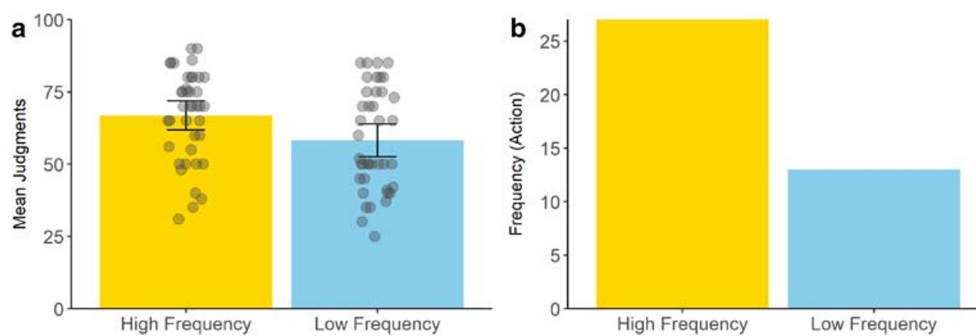


Fig. 1 Panel **a** Mean effectiveness judgments and jittered data points in Experiment 1. Error bars depict 95% confidence intervals for the mean. Panel **b** Frequency of choice for each remedy in the Action judgment

Experiment 2

In the previous experiment, our participants considered the most frequent treatment as more effective and they tended to select it when asked which of the two remedies they would choose in case they experienced the pertinent health problem. We can draw a parallel between this result and the cue-density effect observed in relation to causal illusions as, in both cases, higher treatment frequency leads to higher perceived effectiveness. However, a fundamental difference between the two findings lies in the fact that the source of the effect in our participants is related to relative preference for a high-density treatment over a low-density one, whereas in the standard causal illusion paradigm a high-frequency treatment is compared to no treatment at all.

Crucially, preference for the most frequent treatment in Experiment 1 cannot be considered a causal illusion per se, because the volunteers were not presented with information of the spontaneous recovery rate (probability of recovery without any treatment or base rate of recovery) and, therefore, it was impossible for them to ascertain whether the two treatments were effective or not with the information presented. In any case, note that the mean effectiveness ratings provided by the volunteers are at least as high as those obtained in previous causal illusion studies using the same recovery proportions (Barberia et al., 2019), suggesting that lacking access to the spontaneous recovery rate did not diminish the tendency to give high-effectiveness judgments.

In order to further explore if the present findings would be sustained when participants are provided the opportunity to gather information about the spontaneous recovery rate, in a second experiment we replicated Experiment 1, but including an additional condition in which volunteers received information about two remedies, plus information about patients who did not take any of the remedies (i.e., information about the base rate of recovery). Note that, in this condition, the probability of recovery was the same among patients taking either of the remedies and among patients not taking either remedy. Therefore, this situation resembles the typical causal illusion experiment in which the contingency between a remedy and

recovery is zero, and (unlike in Experiment 1) this time the effectiveness can be truly assessed by comparing the recovery rates to the baseline without treatment. Our hypothesis was that participants in this condition would form a causal illusion, misbelieving that the remedies were effective, and that this would be especially true for the high-frequency one (Blanco et al., 2013). Consequently, they should judge this frequent remedy as more effective than the low-frequency one, and select it when given the option to choose in the action judgment.

Method

Participants A power analysis conducted in GPower (Erdfeiler et al., 2009) indicated that a minimum of 57 participants were required to detect the effect observed in the previous experiment, $d = 0.381$, with a statistical power of 0.8. One hundred anonymous Internet users took part in the study in exchange for a small reward through the Prolific Academic service (Peer, Brandimarte, Samat, & Acquisti, 2017). Data corresponding to four participants were not recorded due to technical reasons, leading to a final sample of 96 volunteers (43 male, 51 female, two other; mean age = 30.6 years, $SD = 10.5$). The Ethical Review Board of the University of Deusto reviewed and approved the protocol, and the experiment was then conducted according to the approved plan.

Design and procedure Participants in Experiment 2 were randomly assigned to one of two possible conditions. One of the conditions (henceforth *Without Context* condition, $n = 52$) was analogous to Experiment 1, as participants observed patients who could take one of two possible remedies against headaches, and they never observed patients who had not taken any of the remedies. The other condition (henceforth *With Context* condition, $n = 44$) also presented participants with patients who could take one of two different remedies, but these were intermixed with 24 patients who did not take either remedy. As shown in Table 1, the number of patients taking either remedy was equal to that in Experiment 1, and both remedies were associated with a probability of recovery of

.75. Moreover, in the case of participants in the *With Context* condition, the probability of recovery among patients not taking any remedy was also .75 (i.e., 18/24), therefore resulting in a zero contingency between taking either treatment and recovery.

The task was, again, an adaptation of the usual task employed when studying causal illusions (Blanco et al., 2013; Matute et al., 2011). The cover story was slightly modified in this second experiment, making the two potential remedies to be two different drugs, instead of being herbs brought from the Amazon. Similar cover stories involving drugs have frequently been used in the past in the literature of causal illusions (Blanco et al., 2013; Yarritu, Matute, & Vadillo, 2014). The use of a different cover story in this new experiment would allow us to evaluate the generalizability of our results. Participants were instructed to imagine that they were medical researchers interested in the extent to which two drugs, Batatrim and Dugetil, were able to reduce headaches (the instructions are available in the Appendix). They were told that they would observe several medical records corresponding to patients suffering intense headaches. For each record, they would know whether the patient took either treatment (or nothing, in the case of the participants in the *With Context* condition). They would be asked to predict whether the patient would recover from the headache within the next 2 h and they would receive feedback afterwards. The instructions finished noting that, after seeing all the information available, they would have to evaluate how effective each of the two treatments was.

Participants in the *Without Context* condition observed a total of 48 patients, of whom 36 took one of the drugs (*high-frequency remedy*) and the remaining 12 took the other one (*low-frequency remedy*). Both drugs were associated with a probability of overcoming the headache of .75 (as shown in Table 1). Participants in the *With Context* condition observed, apart from the 48 patients just described for the *Without Context* condition, 24 more patients, who did not receive any of the drugs. The probability of overcoming the headache was also .75 among these 24 patients (see Table 1). Which of the two drugs, Batatrim or Dugetil, was the high-frequency or the low-frequency drug was randomly decided for each participant.

After observing all the patients, the participants were asked to answer three questions. Similar to Experiment 1, participants had to judge the effectiveness of each of the drugs (*effectiveness questions*) on a scale from 0, “Not effective at all,” to 100, “Completely effective” (with 50 labeled as “Moderately effective”), although in this case the two drugs were judged simultaneously on the same screen (which drug appeared in the upper or bottom part of the screen was randomly decided for each participant). Finally, they were also asked to imagine that they had a headache, and were required to choose what option they would prefer (*action question*, in which participants could

choose “I would take Dugetil,” “I would take Batatrim,” or “I would take nothing”). Note that the structure of the action question was slightly different from that used in Experiment 1, because in this case participants could choose not only between the two treatments, but they could also choose not to take anything. This change was included in order to determine if, when given the option, participants in the *With Context* would take into account the high spontaneous recovery rate and, therefore, would prefer not to take any of the drugs. Which type of dependent measure, the *effectiveness questions* or the *action question*, was requested first was randomly decided for each participant.

Results

Figure 2a shows the mean effectiveness ratings for each of the remedies given by the participants of each of the two conditions. Participants appeared to perceive the high-frequency remedy as more effective than the low-frequency one. An ANOVA with Condition (*With context* vs. *Without context*) as a between-subjects factor and Remedy (*Low frequency* vs. *High frequency*) as a within-subjects factor returned a significant main effect of Remedy, $F(1, 94) = 32.06, p < .001, \eta_p^2 = .254$. Neither the main effect of Condition, $F(1, 94) = 3.39, p = .069, \eta_p^2 = .035$, nor the interaction, $F(1, 94) = 2.52, p = .116, \eta_p^2 = .026$, reached significance.

Regarding the action judgment, the distribution of responses was significantly different between the two conditions, $\chi^2(2) = 23.566, p < .001$ (see Fig. 2b). In the *With Context* condition, further analyses indicated significant differences between the percentages of selection of the high- and low-frequency remedies, $\chi^2(1) = 8.167, p = .004$, as well as between the low-frequency remedy and the no-remedy option, $\chi^2(1) = 9.000, p = .003$, while the percentages of selection of the high-frequency remedy and the no-remedy option were not significantly different from each other, $\chi^2(1) = 0.26, p = .873$. Regarding the *Without Context* condition, the percentages of selection of the high-frequency remedy, the low-frequency remedy, and the no-remedy option were all significantly different to each other ($ps < .005$). To sum up, whereas in the *Without Context* condition we replicated the preference for the high-frequency candidate cause over the low-density candidate cause, with only 4% of participants choosing the new “I would take nothing” option, the pattern of choices in the *With Context* condition was substantially different. In this condition, even though the high-frequency candidate cause was still preferred to the low-frequency one (43% vs. 11% of participants chose each of these options), the percentage of participants choosing not to take any of the remedies was similar to those choosing the preferred high-frequency remedy (45% vs. 43%, respectively).

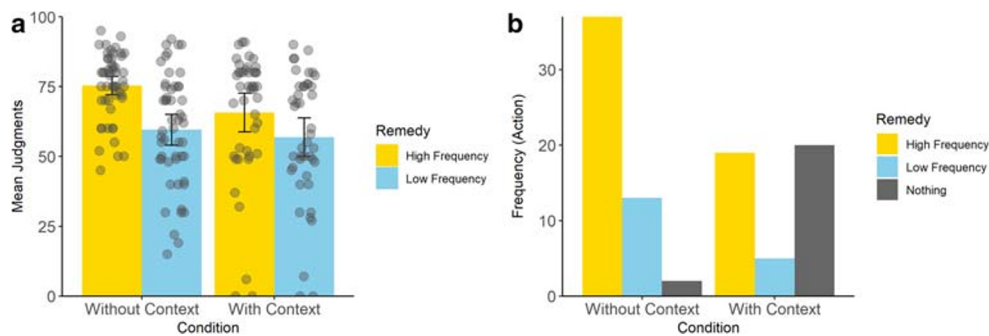


Fig. 2 Panel **a** Mean effectiveness judgments and jittered data points in Experiment 2. Error bars depict 95% confidence intervals for the mean. Panel **b** Frequency of choice for each remedy (including the option of choosing “nothing”) in the Action judgment

In order to explore to what extent both dependent variables (effectiveness ratings and action judgments) converged, we divided participants in three subgroups depending on their choices in the action judgment, and analyzed the effectiveness ratings for each of these subgroups. Figure 3 shows the mean effectiveness ratings given by participants of each subgroup. The analysis concerning participants who chose the high-frequency remedy in the action question indicated that the volunteers produced significantly higher effectiveness ratings to the higher-frequency remedy compared to the low-frequency one both in the *With Context*, $t(18) = 6.461$, $p < .001$, $d = 1.482$, and in the *Without Context*, $t(36) = 8.554$, $p < .001$, $d = 1.406$, conditions. On the other hand, those participants choosing the low-frequency remedy also gave consistent effectiveness ratings, i.e., lower ratings for the high-frequency than for the low-frequency remedy in both the *With Context*, $t(4) = -3.574$, $p = .023$, $d = -1.599$, and *Without Context*, $t(12) = -5.290$, $p < .001$, $d = -1.467$, conditions. Finally, those participants choosing “nothing” in the action question did not show differential effectiveness ratings for high versus low remedies, neither in the *With Context* condition, $t(19) = 1.042$, $p = .311$, $d = 0.233$, nor in the *Without Context* condition, $t(1) = 2.429$, $p = .249$, $d = 1.717$ (note that the later result in the *Without Context* condition is hardly meaningful given that only two participants chose this option).

General discussion

In two experiments, we found that, when training two potential remedies (i.e., potential causes) as treatments for an ailment (i.e., the outcome), the remedy that was more frequently administered was given higher effectiveness ratings, even though the rate of recovery was identical for both high-frequency and low-frequency remedies. Experiments 1 and 2 (*Without Context* condition) demonstrated this effect when participants did not receive any information about the spontaneous recovery rate of the ailment, and, therefore, there was no available contingency information that could be used to determine if any of the remedies was actually increasing the chances of recovery. The same pattern of results emerged when this baseline information (i.e., how many patients recover without treatment) was provided (Experiment 2, *With Context* condition). These results extend previous observations of cue density effects (Blanco et al., 2013; Matute et al., 2011) obtained with typical causal illusion paradigms (i.e., experiments comparing one target remedy vs. no-remedy), to situations in which two remedies are compared to each other. Even though the recovery rates are the same, the more frequent remedy is perceived to be more effective, irrespective of the availability of base-rate information.

Our experiments can, to some extent, be parallel to illusory correlation studies on stereotype formation. In these studies, participants tend to favor a more numerous fictitious social group (majority group) compared to a less numerous one

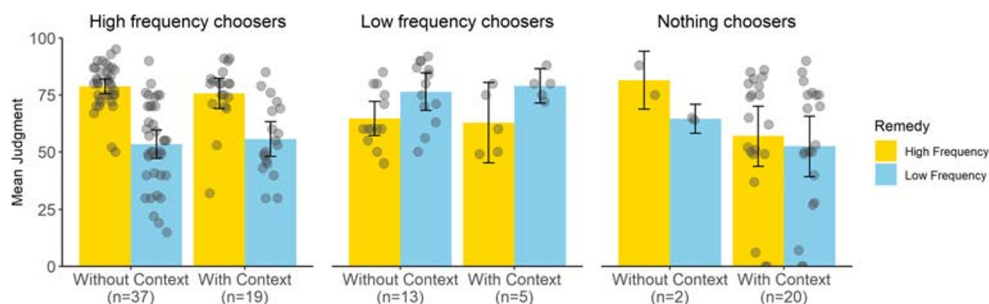


Fig. 3 Mean effectiveness judgments and jittered data points in all subgroups depending on participant response to the action question in Experiment 2. Error bars depict 95% confidence intervals for the mean

(minority group), when positive behaviors are more frequent than negative ones and both majority and minority groups show an identical proportion of positive-to-negative behaviors (Hamilton & Gifford, 1976; Murphy, Schmeier, Vallée-Tourangeau, Mondragón, & Hilton, 2011; Rodríguez-Ferreiro & Barberia, 2017). Note the similarities with our results, in which the more prevalent remedy is perceived as more effective than the less prevalent one, even when the recovery rate is kept constant (and high) among the two treatments. Typically, illusory correlation procedures do not include conditions in which baseline information is available (like our *With Context* condition), so participants are bounded to just form an attitude towards the groups in relative, rather than absolute, terms. Nevertheless, as our Experiment 2 indicates, this additional information does not seem to abolish the tendency created by the density manipulation. In any case, it is worth commenting on the potential explanations that have been proposed for the illusory correlation effect, and how they could apply to our design.

Initial accounts of the illusory correlation effect were based on a memory bias toward infrequent events (Hamilton & Gifford, 1976): the combination of minority group-uncommon trait is relatively rare, and thus becomes more salient and is better remembered. The same could occur in our experiments when the infrequent cue is followed by no healing (i.e., the infrequent cue could be associated with the infrequent outcome, no healing, to a greater extent than the frequent cue does, thus leading to the conclusion that the infrequent cue is not as effective as the frequent one). However, empirical support for a memory bias of this kind is scarce (Fiedler, Russer, & Gramm, 1993).

Another common explanation for the illusory correlation effect is the “incomplete learning hypothesis” (Kutzner & Fiedler, 2015). This hypothesis proposes that, because there are more trials corresponding to the majority group than to the minority group, the association of the former is learned sooner, or more readily, than the latter. Thus, when tested, the majority group reveals stronger associations with the common (positive) traits than does the minority group. This explanation could be easily applied to our experiments, especially in the *Without Context condition*: if the frequent cue receives more training, then it will be learned sooner than the infrequent cue, thus creating a stronger association with the outcome. Such an account assumes that the random order of trials during the training warrants faster exposure to the frequent cue since the beginning of the session. A consequence of the incomplete learning explanation is that the advantage of the frequent cue over the infrequent cue may in fact not be considered a bias per se, because the different responses also reflect the different levels of uncertainty associated to each representation. Interestingly, the same concept of “incomplete learning” can be captured by certain associative learning models initially developed to study animal conditioning, to provide an

associative explanation for both the illusory correlation effect (Murphy et al., 2011) and our results.

In most associative learning models, the process of learning is conceptualized as an error-correction mechanism (e.g. Rescorla & Wagner, 1972) that is updated on a trial-by-trial basis. As has been found in computer simulations (Matute, Blanco, & Díaz-Lago, 2019), when a single cue is trained and the contingency is null, it is common to observe a positive association at the beginning of the training session that eventually wears off as the algorithm converges, which means that the illusions predicted by these models are preasymptotic. This initial spurious association is produced by the accumulation of cause- and outcome-present trials. When cause-outcome co-occurrences are frequent, the model produces larger preasymptotic overestimations than when these co-occurrences are rare. On the other hand, when the base-rate of the outcome is not trained, as occurs in the illusory correlation paradigm and in our study (i.e., Experiment 1 and *Without Context* condition in Experiment 2), the outcome base-rate is unknown and the model does not converge to zero (Murphy et al., 2011).

To illustrate this point, we here report two simulations of the Rescorla-Wagner model representing the conditions studied in our two experiments (see Fig. 4). When setting up the simulation, we assumed that a contextual cue X was present in all trials (i.e., when remedy A was presented, the trial included the cue A and the context X), as is common practice. The parameters used for these simulations are based on common assumptions about the learning process, and they do not affect the ordinal results shown here, only the shape of the learning curves.

The simulations show that the high-frequency remedy acquires more associative strength than does the low-frequency remedy, just as our participants showed in their causal judgments, but this difference tends to disappear with further training (i.e., it is preasymptotic). Eventually, both remedies converge on similar values. Additionally, when the context is trained separately (*With Context* condition), the algorithm converges on the true contingency, zero, if given enough training. However, when the context is not trained separately (*Without Context* condition), the model does not converge to zero (see also Murphy et al., 2011). Importantly, in both conditions (with and without contextual training), we observe that the high-frequency remedy acquires more association than the low-frequency remedy, and that this difference between the two remedies tends to vanish as more training is given.

Another fundamental contribution of our study is related to the use of a novel measure: the *action question*. While effectiveness ratings were not significantly affected by the inclusion of spontaneous recovery rate information, action judgments clearly were. Mirroring the responses to the *effectiveness questions*, participants who had no access to information about the base rate of recovery preferred to choose the high-

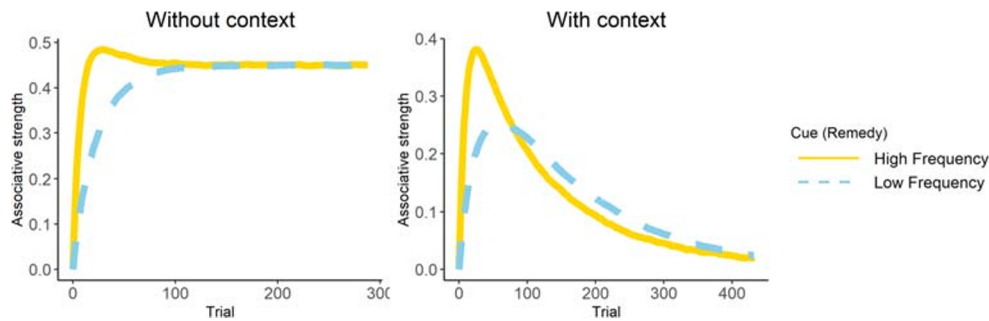


Fig. 4 Simulations with the Rescorla-Wagner model, representing the two scenarios presented in our two experiments: high- and low-frequency remedies – without context (as in Experiment 1 and in the *Without Context* condition in Experiment 2), and high- and low-frequency remedies – with context (as in the *With Context* condition in

Experiment 2). The number of trials that we used in the experiments have been multiplied by six to ensure that the curves reach the asymptote in these simulations. The parameters for these simulations are: $\alpha_{HF} = \alpha_{LF} = 0.3$; $\alpha_{context} = 0.1$; $\beta_{Outcome} = \beta_{-Outcome} = 0.4$

frequency remedy compared to either the low-frequency remedy or neither of them. In contrast, in the condition in which this information was available, a substantial number of participants decided not to take either of the remedies. The absence of influence of the base rate information (i.e., spontaneous recovery rate in our medical scenario) over the effectiveness ratings could lead us to conclude that participants' causal impressions are barely affected by this information. However, the additional dependent variable (action question) introduced in this work points out otherwise. Even though the inclusion of information regarding the base rate of recovery did not explicitly affect causal impressions as measured by our effectiveness ratings, it did impact participants' decisions regarding whether they would choose taking the remedy or not. The conclusion when focusing on these choices is more optimistic, as a substantial number of participants chose not to take any of the remedies when spontaneous recovery rate information was available to them.

Although the general results could be interpreted as proof of a dissociation between our two dependent variables, further analyses indicated otherwise. The study of the effectiveness ratings provided by the volunteers choosing one of the two treatments in the action question showed that the responses to these two questions converged. In general, more participants selected the higher-frequency remedy over the lower-frequency one and, congruently, overall effectiveness means favored the high-frequency remedy. However, a closer inspection of the data revealed that this pattern emerged only in the subgroup of participants selecting the high-frequency remedy. In contrast, the less numerous group of choosers of the low-frequency one produced higher effectiveness judgments for this later remedy.

With regard to the subgroup of volunteers who chose neither of the remedies, the convergence of the responses to the effectiveness and action questions is not that clear. These participants perceived that both remedies were, indeed, equivalent regarding their possible effectiveness. More importantly, even though they chose neither of the remedies, they provided

effectiveness ratings that indicated, at least, moderate effectiveness for both of them (as can be seen in Fig. 3, participants in this subgroup produced effectiveness ratings well above zero and around the intermediate point of the scale, which was labelled as “Moderately effective”). This is especially relevant for those participants who were provided with base-rate information because in this case the contingency between remedy intake and recovery is computable and null. Previous research (e.g. Blanco, Matute, & Vadillo, 2011; Matute et al., 2011; Torres, Barberia, & Rodríguez-Ferreiro, 2020) has usually considered responses to the effectiveness question significantly above 0 in null contingency situations to reflect a causal illusion. This implies the assumption that the effectiveness response reflects perceived contingency. Nevertheless, the fact that these participants selected neither of the remedies even though they reported above zero effectiveness ratings indicates that the absolute values provided in response to this scale should be interpreted with caution, i.e., *effectiveness question* routinely applied in causal illusion literature might fail to capture the real causal impressions that would subsequently drive treatment choice (see Perales et al., 2017).

An interesting possibility is that the responses to this effectiveness question reflect, or are affected by, the level of confidence in the information perceived (Buehner, Cheng, & Clifford, 2003; Liljeholm & Cheng, 2009). For instance, and as pointed out by Perales and Shanks (2008, p. 1487), an intermediate judgment of 50 might equally reflect that participants are sure that an intermediate causal connection exists between a potential cause and an outcome, or that they are moderately confident about the existence of a strong connection between them. Taking this into account, we could interpret that participants choosing either of the remedies did so because they were more confident (or, from a different perspective, less uncertain; see Curley, Yates, & Abrams, 1986) about it being connected to the recovery (as opposed to believing that it was more effective). In the same vein, we could interpret that the participants who refrained from selecting any of the remedies did so because they were not

sure whether there was a relationship between remedy intake and recovery (and this was reflected by their intermediate responses in the numerical scale).

Our research is an important first step in the study of the conditions influencing treatment choice when several alternative treatments are available. We show that given the same recovery rates, treatments with a relatively higher prevalence are preferred over less frequent ones. This result is observed even when base recovery rates are available and indicate that both treatments are, actually, ineffective. This condition is analogous to situations in which an individual must decide between two equally ineffective treatments. It would be interesting for future studies to investigate whether the preference for a high-frequency remedy remains even when the low-frequency one is associated with a relative increase in the recovery rate. This situation would mirror those real-life situations in which an individual might tend to choose an ineffective but popular pseudomedicine over a lesser known, but actually effective, option. There are in fact results in the area of reinforcement learning that point in this direction. Don, Otto, Cornwall, Davis, and Worthy (2019) explored choice behavior when confronting two alternatives, one of them being associated with a higher probability of reward, the other being connected with greater absolute amount of reward. For instance, participants learned to choose between options A (65% reward) or B (35% reward) in 100 trials, and C (75% reward) or D (25% reward) in 50 trials. Subsequently, the volunteers were faced with the choice between A and C options. Even though A was actually associated to lower reward rates than C during training, participants consistently tended to select option A over C. This suggested that the choice was driven by cumulative instances of reward (how often each option has yielded reward in previous experience) more than the average reward (the probability of reward associated to each alternative). Future studies should explore whether analogous effects are found with regard to causal detection, using causal learning tasks such as the one in the present work. For example, participants might be presented with a disease showing an intermediate level of spontaneous recovery (e.g., 50% base recovery rate), and two potential remedies: an infrequent but effective one (e.g., nine out of 12 patients recovering after taking this remedy) and a frequent but ineffective one (e.g., 18 out of 36 patients recovering after taking this remedy).

All in all, our results call for the employment of dependent variables directly reflecting relevant behavior such as choice-related measures over more difficult to interpret judgments. Moreover, this study is relevant for ecological contexts related to self-medication and inadequate treatment selection. We show that individuals tend to choose more frequent treatments over less frequent ones even when they are both associated to the same recovery proportions. This effect could contribute to a vicious circle in which the popularity of a given pseudomedicine perpetuates its selection. On the positive side, when given

exposure to patients who do not receive any treatment, the number of participants choosing not to initiate any treatment increases.

Author note This study was supported by grant PSI2016-75776-R (AEI/FEDER, UE) to IB and grant PSI2017-83196-R (AEI/FEDER, UE) to FB, both from the Agencia Estatal de Investigación of the Spanish government and the European Regional Development Fund. We thank Marta N. Torres for testing the computer task used in Experiment 1 and for running that experiment.

Open practices statements The data and materials for the two experiments are available at the Open Science Framework repository (<https://osf.io/ctjs>). None of the experiments was preregistered.

Appendix

Experiment 1

Task instructions (English translation) “Imagine you are a researcher exploring to what extent two herbal infusions brought from the Amazon work as a treatment for headaches (to simplify, we will call them infusions A and B).

To do this, you have a series of medical records corresponding to patients suffering from headaches. In each record you will observe a patient during one of the headache episodes. We will tell you if the patient has taken infusion A or B during the episode. You will have to decide whether you think the patient's headache will disappear within the next two hours or not. Then, we will tell you if the pain disappeared or not, and you will go ahead to observe the next patient.

After observing several patients, we will ask you some questions. Remember your goal is to determine to what extent each of the infusions is effective.

You can click on the CONTINUE button to start.”

Final questions (English translation) Causal questions: “To what extent do you think the herbal infusion A [B] is effective against headaches? Please use the slider to respond and then click OK. You can select any value between 0 and 100. A value of 0 means that the herbal infusion A [B] is not effective at all and a value of 100 means that is totally effective.”

Action question: “If you had a headache episode, which of the two infusions would you prefer to take? Click on the image of the infusion of your choice.”

Frequency questions: “You have observed 36 [12] medical records in which the patients had a headache and took the A [B] herbal infusion. Could you remember in how many of these patients the headache disappeared within two hours? Please use the slider to respond and then click OK. You can select any value between 0 and 36 [12].”

Task instructions (original in Spanish) “Imagina que eres un/a investigador/a que está explorando hasta qué punto dos infusiones de hierbas traídas del Amazonas funcionan como tratamiento contra el dolor de cabeza (las llamaremos infusiones A y B para simplificar).

Para ello, dispones de una serie de fichas médicas correspondientes a pacientes que sufren dolores de cabeza. En cada ficha podrás observar a un paciente durante uno de sus episodios de dolor de cabeza. Te indicaremos qué infusión, A o B, ha tomado el paciente durante el episodio. Tú deberás indicar si piensas que el dolor de cabeza del paciente desaparecerá en las siguientes dos horas o no. A continuación te diremos si el dolor desapareció o no, y pasarás a observar el siguiente paciente.

Después de observar varios pacientes te haremos algunas preguntas. Recuerda que tu objetivo es determinar hasta qué punto cada una de las infusiones es efectiva.

Puedes hacer clic en el botón de CONTINUAR para comenzar.”

Final questions (original in Spanish) Causal questions: “¿Hasta qué punto crees que la infusión de hierbas A [B] es efectiva contra el dolor de cabeza? Por favor, utiliza la escala móvil para responder y luego haz clic en OK. Puedes seleccionar cualquier valor entre 0 y 100. Un valor de 0 significa que la infusión de hierbas A [B] no es nada efectiva y un valor de 100 que es totalmente efectiva.”

Action question: “Si tuvieses un episodio de dolor de cabeza, ¿cuál de las dos infusiones preferirías tomar? Haz clic en la imagen de la infusión que escojas.”

Frequency questions: “Has observado 36 [12] fichas médicas en las que los pacientes tenían dolor de cabeza y tomaban la infusión de hierbas A [B]. ¿Podrías recordar en cuántos de estos pacientes el dolor de cabeza desapareció en las siguientes dos horas? Por favor, utiliza la escala móvil para responder y luego haz clic en OK. Puedes seleccionar cualquier valor entre 0 y 36 [12]”.

Experiment 2

Task instructions Imagine that you are a medical researcher who is interested in how two drugs, Batatrim and Dugetil, are able to reduce headaches. In this study, you will observe a series of medical records corresponding to patients who suffer from intense headaches. On each record, you will know which treatment the patient took: either Batatrim, or Dugetil (or nothing). Then, you will have to predict whether you think that the headache stopped within the next two hours. Immediately after, you will be informed about whether the headache stopped. This will proceed to the next patient. Once you see all the information available, you will be asked about how effective you think the two treatments are. Remember that in

this study you cannot jot down notes: rather, you should rely on your memory and intuition abilities. Good luck!

Final questions Causal questions: How effective is Dugetil? / How effective is Batatrim?

Answer on the scale. Any number between 0 and 100 is valid:

- 0: Not effective at all
- 50: Moderately effective
- 100: Completely effective

Action question: What would you DO in case you have a headache?

- a) I would take Dugetil
- b) I would take Batatrim
- c) I would take nothing

References

- Allan, L. G., & Jenkins, H. M. (1983). The effect of representations of binary variables on judgment of influence. *Learning and Motivation, 14*, 381–405.
- Astin, J. A. (1998). Why patients use alternative medicine. *JAMA, 279*(19), 1548. <https://doi.org/10.1001/jama.279.19.1548>
- Barberia, I., Vadillo, M. A., & Rodríguez-Ferreiro, J. (2019). Persistence of causal illusions after extensive training. *Frontiers in Psychology, 10*, 24. <https://doi.org/10.3389/fpsyg.2019.00024>
- Blanco, F., Barberia, I., & Matute, H. (2014). The lack of side effects of an ineffective treatment facilitates the development of a belief in its effectiveness. *PLoS ONE, 9*(1), e84084. <https://doi.org/10.1371/journal.pone.0084084>
- Blanco, F., Barberia, I., & Matute, H. (2015). Individuals who believe in the paranormal expose themselves to biased information and develop more causal illusions than nonbelievers in the laboratory. *PLoS ONE, 10*(7), e0131378. <https://doi.org/10.1371/journal.pone.0131378>
- Blanco, F., Matute, H., & Vadillo, M. A. (2010). Contingency is used to prepare for outcomes: Implications for a functional analysis of learning. *Psychonomic Bulletin and Review, 17*(1), 117–121.
- Blanco, F., Matute, H., & Vadillo, M. A. (2011). Making the uncontrollable seem controllable: The role of action in the illusion of control. *Quarterly Journal of Experimental Psychology, 64*(7), 1290–1304. <https://doi.org/10.1080/17470218.2011.552727>
- Blanco, F., Matute, H., & Vadillo, M. A. (2013). Interactive effects of the probability of the cue and the probability of the outcome on the overestimation of null contingency. *Learning & Behavior, 41*(4), 333–340. <https://doi.org/10.3758/s13420-013-0108-8>
- Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1119–1140.
- Curley, S. P., Yates, J. F., & Abrams, R. A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes, 38*(2), 230–256. [https://doi.org/10.1016/0749-5978\(86\)90018-X](https://doi.org/10.1016/0749-5978(86)90018-X)

- Don, H. J., Otto, A. R., Cornwall, A. C., Davis, T., & Worthy, D. A. (2019). Learning reward frequency over reward probability: A tale of two learning rules. *Cognition*, *193*, 104042. <https://doi.org/10.1016/j.cognition.2019.104042>
- Erdfelder, E., Faul, F., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fiedler, K., Russer, S., & Gramm, K. (1993). Illusory correlations and memory performance. *Journal of Experimental Social Psychology*, *29*(2), 111–136.
- Griffiths, O., Shehabi, N., Murphy, R. A., & Le Pelley, M. E. (2018). Superstition predicts perception of illusory control. *British Journal of Psychology*, *110*(3), 499–518. <https://doi.org/10.1111/bjop.12344>
- Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*, 392–407.
- Kutzner, F. L., & Fiedler, K. (2015). No correlation, no evidence for attention shift in category learning: Different mechanisms behind illusory correlations and the inverse base-rate effect. *Journal of Experimental Psychology: General*, *144*(1), 58–75. <https://doi.org/10.1037/a0038462>
- Liljeholm, M., & Cheng, P. W. (2009). The influence of virtual sample size on confidence and causal-strength judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 157–172. <https://doi.org/10.1037/a0013972>
- Matute, H., Blanco, F., & Díaz-Lago, M. (2019). Learning mechanisms underlying accurate and biased contingency judgments. *Journal of Experimental Psychology: Animal Learning and Cognition*, *45*(4), 373–389.
- Matute, H., Blanco, F., Yarritu, I., Díaz-Lago, M., Vadillo, M. A. & Barberia, I. (2015). Illusions of causality: How they bias our everyday thinking and how they could be reduced. *Frontiers in Psychology*, *6*, 888. <https://doi.org/10.3389/fpsyg.2015.00888>
- Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology*, *102*(3), 392–405. <https://doi.org/10.1348/000712610X532210>
- Murphy, R. A., Schmeier, S., Vallée-Tourangeau, F., Mondragón, E., & Hilton, D. (2011). Making the illusory correlation effect appear and then disappear: The effects of increased learning. *Quarterly Journal of Experimental Psychology*, *64*(1), 24–40. <https://doi.org/10.1080/17470218.2010.493615>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Perales, J. C., Catena, A., Cándido, A., & Maldonado, A. (2017). Rules of causal judgment: Mapping statistical information onto causal beliefs. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning*. New York, NY: Oxford University Press.
- Perales, J. C., & Shanks, D. R. (2008). Driven by power? Probe question and presentation format effects on causal judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1482–1494. <https://doi.org/10.1037/a0013509>
- Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York, NY: Appleton-Century Crofts.
- Rodríguez-Ferreiro, J., & Barberia, I. (2017). The moral foundations of illusory correlation. *PLoS ONE*, *12*(10), e0185758. <https://doi.org/10.1371/journal.pone.0185758>
- Torres, M. N., Barberia, I., & Rodríguez-Ferreiro, J. (2020). Causal illusion as a cognitive basis of pseudoscientific beliefs. *British Journal of Psychology*. <https://doi.org/10.1111/bjop.12441>
- Yarritu, I., Matute, H., & Vadillo, M. A. (2014). Illusion of control: The role of personal involvement. *Experimental Psychology*, *61*(1), 38–47. <https://doi.org/10.1027/1618-3169/a000225>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.