Social Science & Medicine 255 (2020) 113012

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Diseases that resolve spontaneously can increase the belief that ineffective treatments work

Fernando Blanco*, Helena Matute

University of Deusto, Spain

ARTICLE INFO

Keywords:
Cognitive bias
Treatment effectiveness
Pseudotherapy
Patients' beliefs
Causality bias

ABSTRACT

Rationale: Self-limited diseases resolve spontaneously without treatment or intervention. From the patient's viewpoint, this means experiencing an improvement of the symptoms with increasing probability over time. Previous studies suggest that the observation of this pattern could foster illusory beliefs of effectiveness, even if the treatment is completely ineffective. Therefore, self-limited diseases could provide an opportunity for pseudotherapies to appear as if they were effective.

Objective: In three computer-based experiments, we investigate how the beliefs of effectiveness of a pseudotherapy form and change when the disease disappears gradually regardless of the intervention.

Methods: Participants played the role of patients suffering from a fictitious disease, who were being treated with a fictitious medicine. The medicine was completely ineffective, because symptom occurrence was uncorrelated to medicine intake. However, in one of the groups the trials were arranged so that symptoms were less likely to appear at the end of the session, mimicking the experience of a self-limited disease. Except for this difference, both groups received similar information concerning treatment effectiveness.

Results: In Experiments 1 and 2, when the disease disappeared progressively during the session, the completely ineffective medicine was judged as more effective than when the same information was presented in a random fashion. Experiment 3 extended this finding to a new situation in which symptom improvement was also observed before the treatment started.

Conclusions: We conclude that self-limited diseases can produce strong overestimations of effectiveness for treatments that actually produce no effect. This has practical implications for preventative and primary health services. The data and materials that support these experiments are freely available at the Open Science Framework (https://osf.io/xt3z9/). https://bit.ly/2FMPrMi

1. Introduction

Self-limited diseases are those that resolve spontaneously, with or without specific treatment (Self-limited disease, 2012). Examples range from mild infections (e.g., common cold, flu, conjunctivitis) to more general symptoms (e.g., headache, back pain). In most cases, there is no need for a specific intervention to make the disease disappear. Regardless, people normally resort to medicines and drugs with the belief that they will increase their chances of overcoming the disease. As we will argue in this paper, self-limited diseases can be dangerous because they represent a good opportunity for pseudotherapies or bogus treatments to gain credibility, since they will appear to be effective when their use coincides with spontaneous improvement.

Pseudotherapies, or pseudomedicines, are medical interventions (in the form of drugs, manual interventions, psychological treatments, diets, etc.) that lack credible evidence supporting their effectiveness beyond being a placebo (Hellmuth et al., 2019). Famous examples include homeopathy (NHMRC, 2015) and chiropraxy (Proctor et al., 2006). Typically, pseudotherapies invoke mechanisms that are biologically implausible (e.g., the “reiki energy,” “memory” in water), and, when tested in properly controlled randomized trials, show no effectiveness compared to a placebo (Hellmuth et al., 2019). Pseudotherapies are widely used in modern society (Barnes et al., 2004; Ernst and Smith, 2018), despite the availability of scientific information. Although most pseudotherapies are offered as “complementary” therapeutical options that do not aim to replace conventional treatments, in fact there are many cases in which patients abandon their scientifically validated treatments in favor of pseudotherapy, with serious consequences (Freckelton, 2012; Johnson et al., 2018).

Generally, we could assume that one of the main reasons why

* Corresponding author. Departamento de Fundamentos y Métodos de la Psicología, Universidad de Deusto, 48007, Bilbao, Spain.
E-mail address: fernandoblanco@deusto.es (F. Blanco).

https://doi.org/10.1016/j.socscimed.2020.113012
Received 11 October 2019; Received in revised form 6 April 2020; Accepted 20 April 2020
Available online 25 April 2020

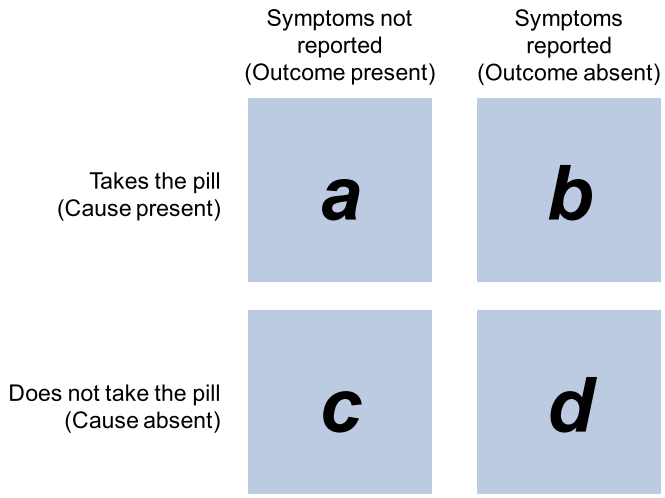

Fig. 1. 2 × 2 contingency table summarizing the information that can be observed during a treatment.

pseudotherapies are used is the presence of unsupported beliefs of effectiveness: if people did not hold such beliefs, they would not replace evidence-based treatments by unscientific approaches. But how are these beliefs of effectiveness formed? Most people cannot directly consult primary scientific sources, or interpret the results of clinical trials, but rather, they rely on their own experience. Surely, advertising campaigns and hearsay play an important role on the acquisition of treatment effectiveness beliefs (de Barra, 2017). Additionally, perhaps they can form gradually as people acquire experience with the treatment (Blanco et al., 2014; Rottman et al., 2017). This means such beliefs are in part the result of causal learning, the ability to acquire causal links between potential causes and their effects, in this case between the use of a treatment and a health improvement, as some authors have argued (Matute et al., 2019). For example, a patient who starts a pharmacological treatment will observe how often the symptoms improve after taking the pill, and compare this situation with those days in which the pill was not taken, to form an idea about how good the pill is for reducing the symptoms. The information of this example is summarized in a 2 × 2 table such as Fig. 1 illustrates. Note that, throughout this article, we focus on how people form beliefs of effectiveness informally, based on their experiences, rather than on actual efficacy, which requires a more sophisticated procedure (e.g., a control group or placebo condition), which is typically unavailable for individual patients. What underlies both concepts is the notion of causality (the treatment must be the cause of the remission).

Each cell, or "type of trial", in Fig. 1 can appear with various frequencies depending on the situation, thus leading to different levels of association, or contingency, between the cause (medicine) and the outcome (symptom remission). If the contingency is null, relief is equally likely to occur when one takes the pill as when one does not (Allan and Jenkins, 1983). This would be the case of a completely ineffective medicine that fails to produce even a placebo effect (which would entail a slight positive contingency).

There is robust evidence for biases in people's contingency estimations. In particular, the outcome-density bias (OD) appears when the outcome (e.g., symptom remission) occurs very often, that is, when the probability of the outcome, P(O), is high. In this situation, null contingencies are greatly overestimated, both in adult participants (Chow et al., 2019) and children (Moreno-Fernández et al., 2017). Importantly, the OD bias is suggested to underlie the beliefs of effectiveness of pseudotherapies (Lilienfeld et al., 2014; Macfarlane et al., 2020). Previous experiments have documented how patients tend to overestimate the effectiveness of a completely ineffective drug when they feel symptomatic relief often (but as frequently when taking the drug as when not taking it). Moreover, if the pseudotherapy is used with high frequency too (e.g., because it has no side effects), then the OD bias becomes stronger, resulting in a large overestimation of the causal relationship between taking the drug and feeling better (Blanco et al., 2014). Yet there is a potentially critical factor that has not yet been studied in this particular domain: What would happen to beliefs of effectiveness if symptoms improved gradually while taking the pseudotherapy, just by mere chance? This situation of experiencing increasing probabilities of remission could be in fact very common, as it is the natural course of self-limited diseases.

Only a few experiments have conducted this type of manipulation (Ejova et al., 2013; Langer and Roth, 1975; Matute, 1995). Ejova et al. (2013) reported a study in which participants predicted the outcome of a sequence of shots in a soccer-simulation game presented as a gambling activity, similar to a slot-machine. The feedback (correct/incorrect) was predefined, and the sequence of trials was manipulated so that in the "ascending" condition the predictions turned out correct with increasing probability as the session progressed. That is, the probability of the outcome, P(O), increased during the experiment. Consistent with previous findings (Matute, 1995), participants were more likely to overestimate their ability to predict the outcomes in the "ascending" condition than in a "descending" condition (which showed the opposite progression, with fewer correctly predicted outcomes at the end of the sequence), even though both conditions presented exactly the same total number of right and wrong predictions. According to the authors, the increasing proportion of successes created in the participants the impression that they were learning the right strategy, thus reinforcing their sense of control over the outcome. On the contrary, other studies found the opposite effect (Langer and Roth, 1975) when conducting a similar manipulation, which indicates that the effect of increasing schedules of P(O) in contingency learning experiments is not completely clear.

When perceiving the effectiveness of a treatment, the experience of increasing probabilities of symptom relief can be highly relevant, because it seemingly indicates an improvement of health condition. As self-limited diseases evolve favorably even without treatment, they imply the presentation of outcomes (i.e., symptomatic relief) with increasing probabilities, in a similar way to the contingency-learning experiments described above (Ejova et al., 2013; Langer and Roth, 1975; Matute, 1995). If increasing patterns of P(O) lead to stronger overestimation of causality (and hence of effectiveness), even when the treatment is completely useless, then patients who treat a self-limited disease with a pseudotherapy could develop an illusory belief of effectiveness, because they would incorrectly attribute the improvement in symptoms to the effect of the treatment, rather than to the natural course of the disease. This means that self-limited diseases could be a particularly dangerous and prevalent opportunity for pseudotherapies to acquire undeserved trust and be used in the future for more serious health problems.

1.1. Experiments 1 and 2

Experiments 1 and 2 used a contingency learning task framed in a medical scenario. In a computer-game format, participants imagined that they were treating their symptoms with a fictitious drug, *Dugetil*. The medicine was completely ineffective (e.g., the improvements were pre-programmed to occur during the session, so that the contingency between the drug and the improvements was null). Thus, Dugetil was conceived as a model for pseudotherapy in the context of our experiments. As participants can see the frequency of occurrence of the symptoms both in the presence and in the absence of treatment, they in principle should be able to combine these pieces of information and correctly detect the absence of effectiveness. However, in the self-limited group, the symptoms were programmed to remit more often as trials progressed, that is, P(O) increased during the session, while in the Control group the probability of symptomatic remission remained fixed

through the session. We expected that the pattern of progressive improvement in the symptoms in the self-limited group would increase the illusory belief of effectiveness compared to the control group. Our aim was to document how beliefs of effectiveness evolve when symptoms improve spontaneously, as it happens in self-limited diseases.

2. Method

2.1. Ethics statement

The Ethical Review Board of the University of Deusto reviewed and approved the methodology reported in this article, and the experiments were conducted according to the approved guidelines.

2.2. Participants and apparatus

The experiments were programmed in *JavaScript*, and made available online through the Prolific Academic website (Palan and Schitter, 2018). Participants (anonymous Internet users) were paid 1£ for about 10 min of work. They logged in on the platform using a temporary identifier generated by the system, which is also used to handle the micro-transaction. This ensures that a participant does not take place twice in a same experiment while keeping the whole process anonymous to the researcher. We planned a sample of 100 participants for each experiment (which allows detecting medium effect sizes with 80% power), but some data were not recorded due to technical issues. In Experiment 1, the final sample consisted of 99 participants (51 female; M_{age} = 30.30 years; SD = 9.18). The computer program randomly assigned 51 participants to the Self-Limited group, and 48 to the Control group. In Experiment 2, 97 participants formed the sample (44 female; M_{age} = 29.60 years; SD = 10.30): Forty-two in the Self-Limited group, and 55 in the Control group.

2.3. Procedure

We used a standard contingency learning task (Matute et al., 2019). Participants imagined that they were suffering from a fictitious disease called "Hamkaoman Syndrome," which produces a variety of symptoms (pain, sickness, fever …) that appear from time to time (see the full instructions in the supplementary file S1). Their doctor prescribed the drug "Dugetil", which they may forget to take on some days. We used fictitious diseases and drugs to prevent contamination due to previous knowledge or attitudes, as we want to study how people form beliefs of effectiveness on the basis of their observations. The experimental session consisted of a series of trials corresponding to consecutive days. Fig. 2 shows schematically the procedure for one trial. On each day, the screen showed whether Dugetil was taken ("You took Dugetil/You did not take Dugetil"), and whether symptoms were experienced ("You reported symptoms/You did not report symptoms"). After seeing this information, participants had to rate the perceived effectiveness of Dugetil in each trial (trial-by-trial judgment). Thus, they were expected to adjust their judgment of the effectiveness of Dugetil as they gathered more information during the phase. Once they answered, the task proceeded to the next trial. There were 4 consecutive blocks of 16 trials, making a total of 64 trials in each group.

The trial frequencies were chosen so that the contingency between using the drug and feeling relief was null for all participants (Fig. 3). That is, the probability of symptomatic relief was identical when taking the medicine as when not taking the medicine. As the actual contingency was null, the treatment is completely ineffective. Thus, any judgment of effectiveness above zero can be treated as an overestimation of the actual effectiveness.

Now we describe our manipulation: in the Self-Limited group, the probability of symptomatic remission was 0.50 on average for the whole session, but was not fixed across blocks. Rather, it increased progressively during the session, with the aim of modeling the natural course of a self-limited disease. In the Control group, by contrast, the probability of relief was always 0.50. The overall amount of trials of each type, the overall P(O), and the contingency were identical between the groups: only the order in which the trials were presented varied (with the different types of trials randomly arranged within each block).

At the end of the training session, we asked several complementary questions: First, a global judgment of effectiveness was collected in the same format and wording as the trial-by-trial judgments, but after encouraging participants to take into account and integrate all the information from the session. We expected that this instruction would help to reduce the illusory beliefs of effectiveness in the global judgment, compared to the trial-by-trial judgments (Matute et al., 2002). Second, participants were also asked to judge the extent to which they would attribute the remission of the symptoms to any unspecified factor different from Dugetil, on a scale from 0 to 10 (Alternative Cause judgment: "On those days when you felt relieved of your symptoms, would you attribute the improvement to Dugetil, or to other factors? 0: To other factors/5: I don't know/10: To Dugetil"). Then, two more questions assessed the perceived conditional probabilities: P(O|C) (probability of relief conditional on having taken the drug), and P(O|~C) (probability of relief conditional on not having taken the drug). Both questions were worded in frequency terms rather than on probabilities, because they are easier to interpret and use by participants (Gigerenzer and Hoffrage, 1995). That is, "Imagine a different person who suffers from the Hamkaoman Syndrome. This person takes Dugetil on 100 consecutive days. Out of these 100 days in which the person takes Dugetil, on how many of them will the person report that the symptoms improved?". Finally, participants indicated whether they had perceived any change in effectiveness during the experimental session (Yes/No dichotomous response) and the direction of the perceived change (Improving/Worsening).

The two experiments were identical in their procedure, except for one aspect: in Experiment 1, we used a bidirectional scale (from −100 to +100) to collect trial-by-trial judgments and global judgments. Thus, in this experiment, negative numbers indicate that the drug worsens the symptoms, positive numbers indicate that the drug improves the symptoms, and zero indicates no effect. Given that overestimations of contingency have been more extensively studied with unidirectional scales, and that participants could find it hard to use and understand negative values in this medical context (i.e., "how could a medicine worsen the symptoms?"), in Experiment 2, we repeated the experiment using a unidirectional scale, from 0 (no effect) to 100 (improves the symptoms), expecting the same results.

3. Results and discussion

3.1. Trial-by-trial judgments

Fig. 4 shows the mean trial-by-trial judgments for each block of 16 trials and per group, in the two experiments. A mixed 2 (Group) × 4 (Block) ANOVA was conducted on these data. The results were similar in both experiments, revealing a Group × Block interaction: $F(3, 291) = 19.93$, $p < 0.0001$, $\eta_p^2 = 0.170$ (Experiment 1); and $F(3, 285) = 21.30$, $p < 0.0001$, $\eta_p^2 = 0.183$ (Experiment 2). Judgments increased during the session in the self-limited group, but not in the control group. In the last block, the self-limited group produced higher judgments than the control group (see Table 2 with the contrasts). That is, despite the drug was equally non-effective in both groups and that the only difference between groups was the trial order, when the pattern of symptom remissions was increasing (which mirrors that of a self-limited disease), the belief that the drug was actually effective was stronger.

3.2. Additional judgments

Next, we analyzed the judgments collected after the sequence of

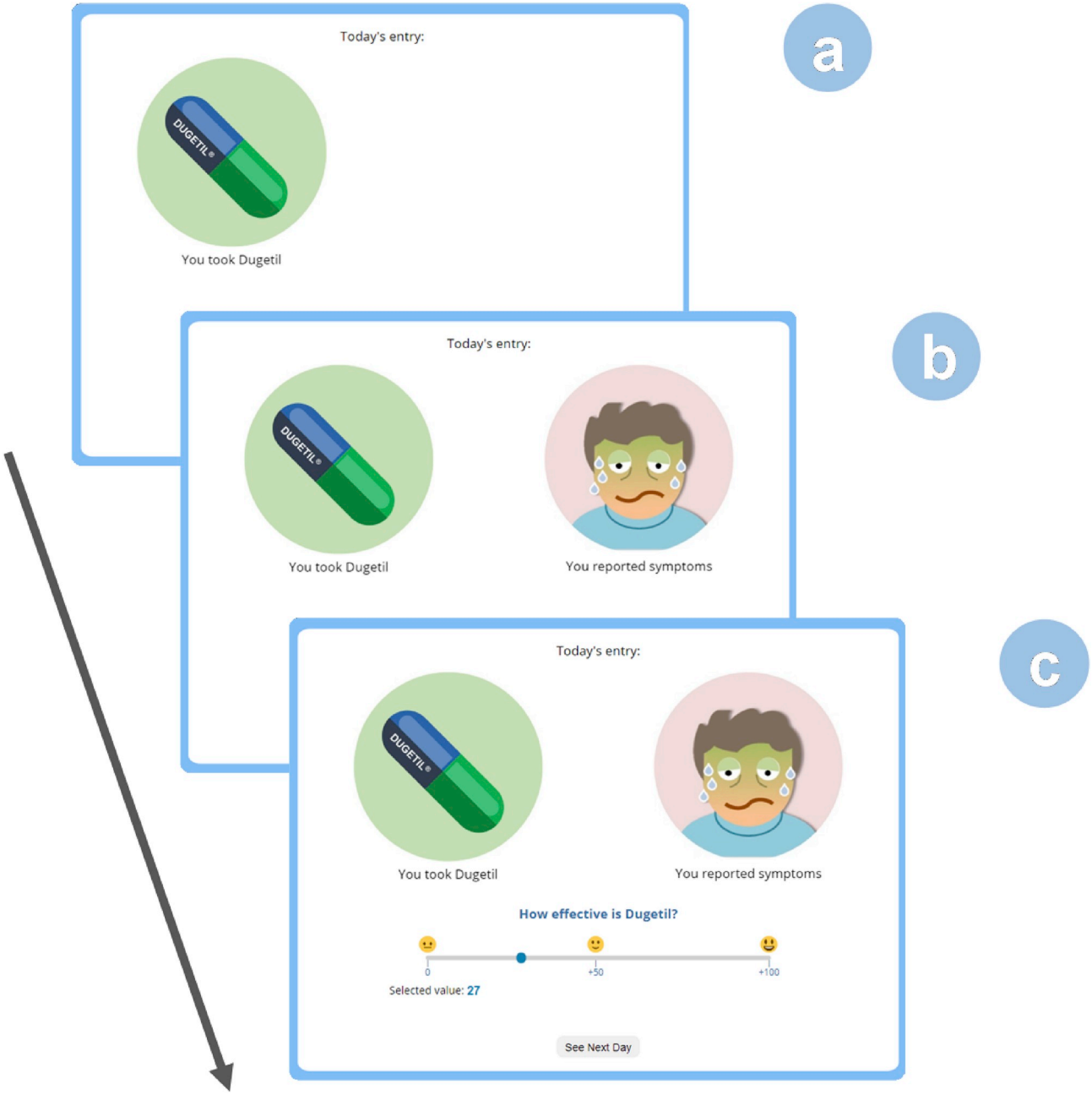

Fig. 2. Schematic depiction of one trial in the task: (a) first, the participant knows whether the medicine was taken; (b) then, the information about the outcome (symptom remission) is displayed; and (c) a judgment about the treatment effectiveness is requested.

trials (Table 1). Global judgments were higher in the self-limited group than in the control group: in Experiment 1, $t(97) = 1.922$, $p = 0.057$, $d = 0.387$; and in Experiment 2, $t(95) = 2.042$, $p = 0.044$, $d = 0.418$. This aligns with the results from the trial-by-trial judgments, and highlights the impact of the manipulation on the effectiveness beliefs: even when participants were explicitly encouraged to weight and take into account all the information given during the session, they still overestimated the pseudotherapy when the probability of remissions increased versus when it remained fixed.

The Alternative Causes Judgment revealed no significant differences between groups in any experiment (both $ps > 0.229$), although the tendency was consistent with the global judgment: Dugetil was identified as the cause for symptomatic remissions more often in the self-limited group than in the control group.

Then, the conditional probability judgments were analyzed by means of a mixed 2 (Group) × 2 (Probability: conditional on the cause presence vs. conditional on the cause absence) ANOVA, producing a significant main effect of Probability: P(O|C) was judged higher than P(O|~C) in both groups, in the two experiments: $F(1, 97) = 28.05$, $p < 0.001$, $\eta_p^2 = 0.224$ (Experiment 1), and $F(1, 95) = 25.17$, $p < 0.001$, $\eta_p^2 = 0.209$ (Experiment 2). That is, participants tended to overestimate P(O|C) over P(O|~C), which is consistent with the generalized overestimation of effectiveness seen in the judgments. In fact, the difference between these two probability estimation judgments (which is analogous to asking about contingency) correlated significantly with global effectiveness judgments in all groups: all $ps < 0.01$. Those participants who more clearly overestimated P(O|C) above P(O|~C) consistently gave stronger global judgments.

Finally, when asked after the training session, most participants in the self-limited group reported perceiving a change in the effectiveness

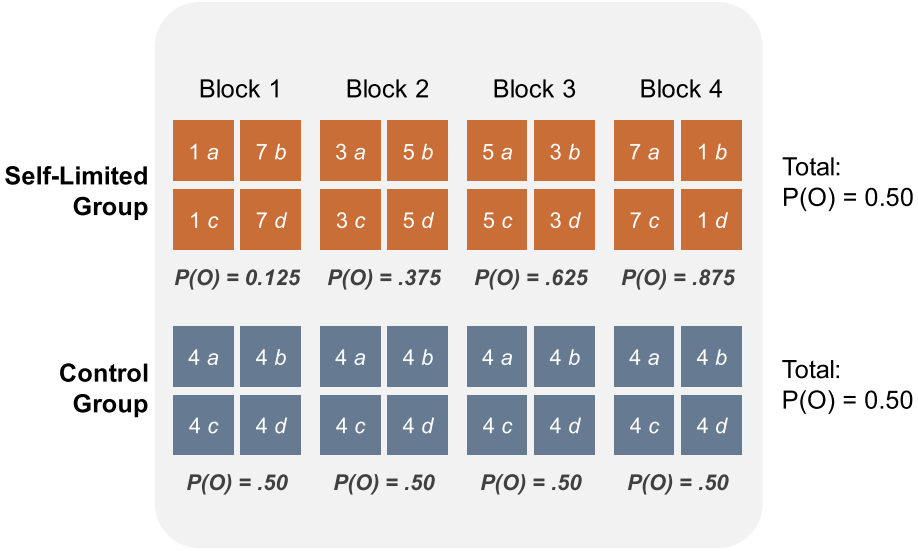

Fig. 3. Design of experiments 1 and 2.

of the medicine (64.70% in Experiment 1, 59.52% in Experiment 2), and the majority of them (about 84%) were able to correctly express the direction of the change (i.e., an improvement). This indicates that they actually detected the increase in P(O) during the task, and attributed it to a change in effectiveness.

3.3. Experiment 3

The previous experiments aimed to model a real situation (the experience of using a pseudotherapy for treating a self-limited disease) in a fictitious and simple setting. Crucially, because the scenario is very simple, it is possible that participants make additional assumptions that were not directly presented in the task. In particular, we did not provide any information about the time-window in which the medicine was expected to make an effect. Rather, our experiments were programmed under the assumption that trials are independent from each other, which means that taking the medicine in one day should not affect the probability of feeling better some days later. This belief could be true of many real treatments (e.g., painkillers usually have effects that are limited in time). However, it could be an unnatural assumption in many real-life scenarios, in which it is correct to infer that medicines make an effect on the long run and, thus, that trials are not independent on each

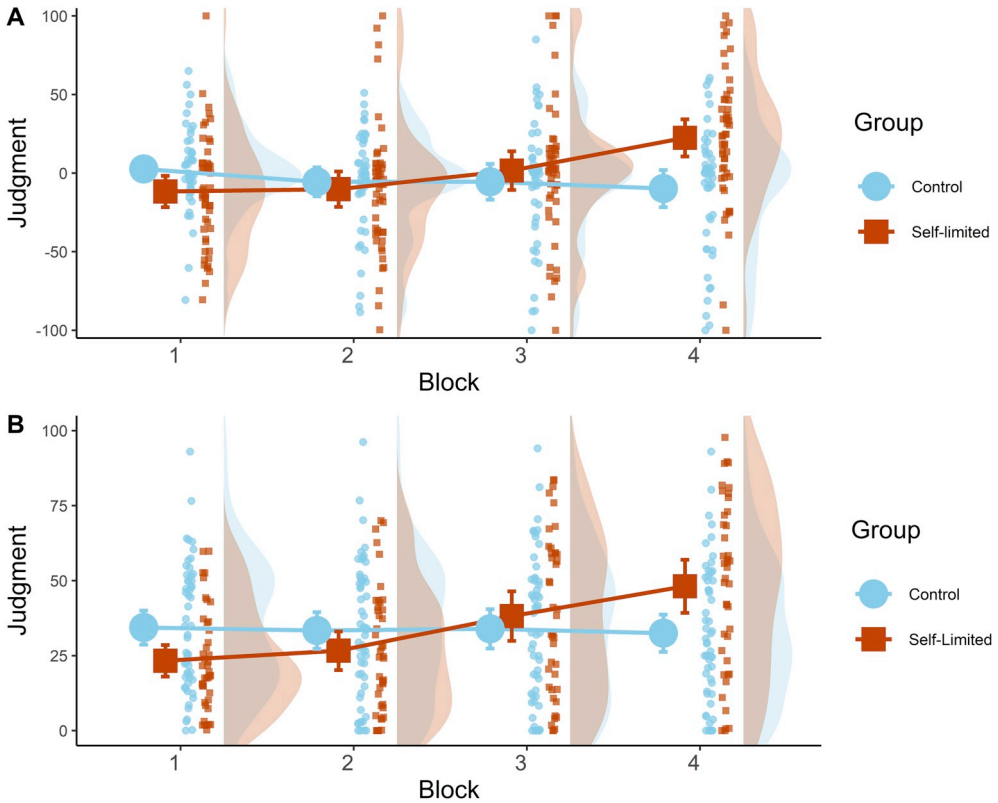

Fig. 4. Trial-by-trial judgments by Block and Group, in the Experiments 1 (panel A) and 2 (panel B). Experiment 1 uses a scale from −100 to +100, while Experiment 2 uses a scale from 0 to 100. Error bars depict 95% CIs for the mean.

Table 1
Descriptive statistics for the four complementary types of judgment collected after the training session (Experiment 1 and 2).

Experiment	Group	Global Judgment		Alternative Causes Judgment		P(O|C) Judgment		P(O|~C) Judgment	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment 1	Self-Limited	9.73	44.80	6.14	2.99	51.10	19.90	32.50	17.70
	Control	−7.13	42.30	6.83	2.72	45.80	22.10	35.70	18.50
Experiment 2	Self-Limited	44.30	26.70	6.36	2.43	46.50	24.30	29.80	18.60
	Control	33.30	26.00	6.55	3.15	51.20	20.60	39.20	16.40

Note. Global judgments were collected on a scale from −100 to +100 in Experiment 1, and on a scale from 0 to 100 in Experiment 2.

Table 2
Between-group contrasts within training blocks (Experiments 1 and 2).

	Experiment 1	Experiment 2
Block 1	$t(97) = -2.272$, $p = 0.025$	$t(95) = -2.796$, $p = 0.006$
Block 2	$t(97) = -0.646$, $p = 0.520$	$t(95) = -1.507$, $p = 0.135$
Block 3	$t(97) = 0.853$, $p = 0.396$	$t(95) = 0.821$, $p = 0.414$
Block 4	$t(97) = 3.881$, $p < 0.001$	$t(95) = 2.992$, $p = 0.004$

other.

The implications of this assumption are important. First, the assumption of non-independence could allow the participant to conclude that the drug is actually effective in the self-limited groups of Experiments 1 and 2. That is, if the assumption is correct, then a medicine that does work to reduce the symptoms could present the pattern of symptom remission that we showed in these experiments (e.g., increasing chances of improving symptoms, both when the medicine is taken and *when it is not*). Second, because of this, contingency (as previously defined) is no longer a suitable index to assess effectiveness when trials are not independent from each other (see Christensen and Bedrick, 1997). Instead, if trials are dependent on each other, one needs to compare the rate of improvement in the symptoms before and during the treatment. Only when these rates are similar one can conclude that the medicine has no effect. Since we did not show participants the baseline before initiating the treatment in Experiment 1 and 2, those participants who assumed that trials were dependent on each other could have concluded that the medicine was working on a deferred basis, and in principle they could be correct.

Experiment 3 tries to address these limitations by instructing participants about the baseline behavior before the treatment starts. Now, even if participants assume dependency between trials, the information will objectively indicate that the medicine is not working (as the rate of improvement in the symptoms will be the same before and during treatment). The goal is to test whether a self-limited disease creates the conditions to overestimate the effectiveness, even under these non-independence assumptions.

4. Method

4.1. Participants and apparatus

We recruited 100 participants through the Prolific Academic website, but due to technical errors, data from only 95 participants were recorded (44 female, 50 male, and 1 did not answer; $M_{age} = 30.6$ years; $SD = 9.70$). The computer program randomly assigned 47 participants to the self-limited group, and 48 to the control group.

4.2. Procedure

The procedure was similar to that of Experiment 2, except for the presentation of two phases during the training. Phase 1 (48 trials) corresponds to the time before any treatment is given (e.g., it contains only c and d trials), thus representing the base-rate of the outcome without medicine. Phase 2 (36 trials) contains trials with and without medicine. As in Experiments 1 and 2, once the treatment starts, the contingency between taking the medicine and symptom relief was null: that is, the chances of improvement are the same in medicine-present trials as in medicine-absent trials. In the self-limited group, symptoms start improving from the beginning (e.g., even before the treatment starts), and continue at the same pace during the whole session, irrespective of whether the medicine is taken or not. Therefore, participants have enough information to conclude that the medicine is not working, but can still be biased by the positive progression in P(O). In contrast, the control group shows symptom relief occurrences with a fixed probability, also non-contingent with the medicine (see the design in Fig. 5). Both groups model real situations: a pseudomedicine used on a

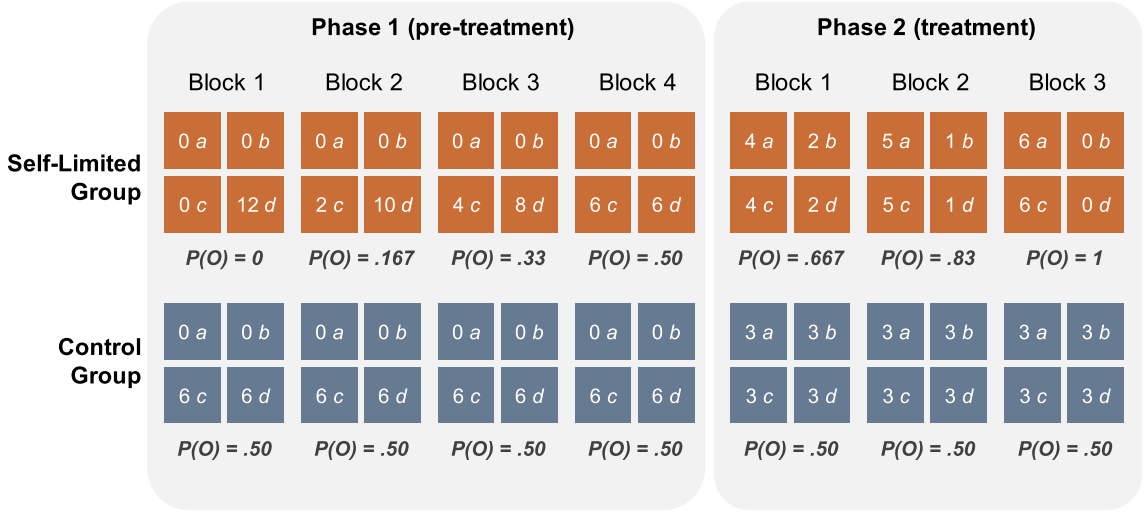

Fig. 5. Design of experiment 3.

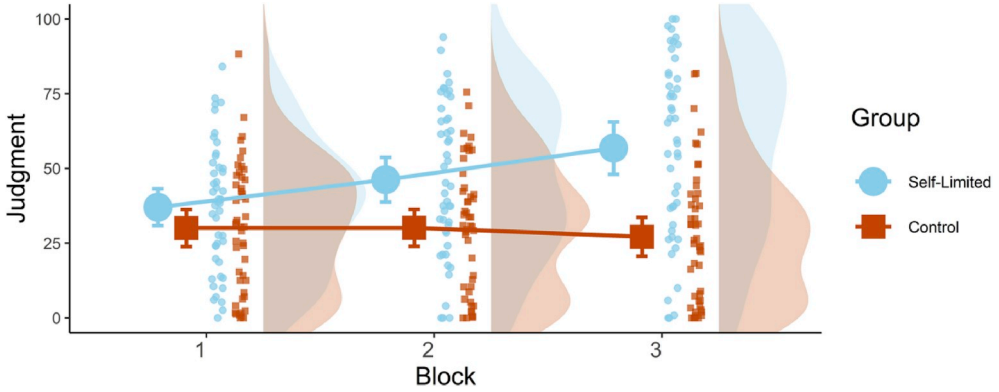

Fig. 6. Trial-by-trial judgments by Block and Group, in the Experiment 3. Error bars depict 95% CIs for the mean.

self-limited disease, and on a stable disease. Trial-by-trial judgments were collected only during Phase 2 (as the medicine was never taken before the treatment starts). Finally, note that both groups end Phase 1 in an identical way (with a probability of remission of 0.50), thus contributing to make their expectations similar before the treatment.

5. Results and discussion

5.1. Trial-by-trial judgments

The results are summarized in Fig. 6. A mixed 2 (Group) × 3 (Block) ANOVA revealed a Group × Block interaction: $F(2, 186) = 22.10$, $p < 0.0001$, $\eta_p^2 = 0.192$. Judgments increased during Phase 2 in the Self-Limited group, but not in the Control group. The groups differed significantly in the second and last blocks (see Table 3).

5.2. Additional judgments

A summary of these data appears in Table 4. Global judgments aligned with the previous results, showing a significant advantage of the Self-Limited group, $t(93) = 5.06$, $p < 0.001$, $d = 1.039$. In this case, we also found differences in the Alternative Causes judgments: Dugetil was identified as the cause for symptomatic remissions more often in the self-limited group than in the control group, $t(93) = 4.27$, $p < 0.001$, $d = 0.876$, despite the many occurrences of symptom remission without taking the medicine in the self-limited group.

Conditional probability judgments were analyzed by means of a mixed 2 (Group) × 2 (Probability: conditional on the cause presence vs. conditional on the cause absence) ANOVA, revealing an interaction, $F(1, 93) = 19.3$, $p < 0.001$, $\eta_p^2 = 0.172$. P(O|C) was judged higher than P(O|~C), but only in the self-limited group. This is consistent with the higher global judgments in this group. The difference between the two probability estimations correlated significantly with effectiveness judgments in both groups: all $ps < 0.001$.

Finally, most participants in the Self-Limited group detected a change in the effectiveness of the medicine (68.1%), and 90.6% of them identified it as an improvement.

Table 3
Between-group contrasts within training blocks (Experiment 3).

	Experiment 1
Block 1	$t(93) = 1.60$, $p = 0.114$
Block 2	$t(93) = 3.35$, $p = 0.001$
Block 3	$t(93) = 5.47$, $p < 0.001$

5.3. General discussion

Pseudotherapies are drugs or other types of treatment that produce no beneficial effect on the likelihood of improving from a disease, such that the contingency between using the pseudotherapy and observing an improvement is null, once potential placebo effects are discounted (Hellmuth et al., 2019). Nevertheless, certain situations can induce illusory beliefs of effectiveness, even for pseudotherapies that are completely ineffective. In particular, diseases with high likelihood of spontaneous remission, P(O), can result in strong overestimations of effectiveness (Matute et al., 2019). Yet in most experiments the P(O) is set to a fixed level during the whole session (Blanco et al., 2014; Chow et al., 2019), with only a few studies investigating the effect of changing the P(O) levels as the session proceeds. In this context, self-limited diseases are of interest because, given their natural evolution, they usually produce a pattern of steady increasing probability of remission. According to previous research, an increasing pattern of P(O) could in principle promote strong illusions of effectiveness in a null contingency setting (Ejova et al., 2013; Matute, 1995), though sometimes the opposite result (e.g., accurate estimations of non-effectiveness) has been reported (Langer and Roth, 1975). Additionally, none of these previous studies investigated the effect of increasing patterns of P(O) in a medical context. Testing the effect of increasing P(O) using a medical context is important, because this pattern, which is common in self-limited diseases, could represent one of the gaps through which pseudotherapies infiltrate our societies. Given that self-limited diseases are common, if they do favor the use of pseudotherapy by promoting the illusory belief in their effectiveness, then many people could turn to these pseudotherapies uncritically whenever they suffer a more serious health problem.

Our three behavioral experiments align in documenting evidence for the overestimation of effectiveness in self-limited diseases. In this regard, they coincide with results reported in non-medical scenarios (Ejova et al., 2013; Matute, 1995), and contradict a classic study (Langer and Roth, 1975). Nevertheless, we note that Langer and Roth's study tapped on a chance situation (coin toss) in which no improvement should be expected with time or practice, which could help to interpret the discrepancy. Additionally, in Ejova et al.'s study, the interpretation of the situation as chance-based was measured and controlled for in the analysis.

In our experiments, trial-by-trial judgments increased gradually with the P(O) in the self-limited group (see Figs. 4 and 6). As a result, the participants in this group ended their training session with a strong belief of effectiveness. This contrasts with participants in the control group, who showed a significantly smaller illusion of effectiveness. A possible interpretation is that people assume that patterns of gradual improvement are the typically expected ones when a treatment is actually working. That is, people rarely attribute these gradually improving patterns to the natural course of a self-limited disease, and

Table 4
Descriptive statistics for the four complementary types of judgment collected after the training session (Experiment 3).

Experiment	Group	Global Judgment		Alternative Causes Judgment		P(O|C) Judgment		P(O|~C) Judgment	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment 3	Self-Limited	60.1	30.2	4.09	3.16	61.8	24.7	29.7	21.7
	Control	30.2	27.3	6.56	2.46	42.0	20.5	38.9	18.1

instead they look for a potential explanation, which in this case is the treatment.

Experiment 2 replicated the results of Experiment 1 (Fig. 4), differing only on the type of response scale used (unidirectional, instead of bidirectional). Although contingency can take on either positive or negative values, it is not obvious how to interpret a negative value in this particular context — a treatment for a disease. Consequently, most previous studies used a unidirectional scale (from 0, ineffective, to 100, perfectly effective). Here we obtain the same results irrespective of the scale. This speaks of the robustness of the general result: regardless of the type of response scale, the tendency to overestimate the effectiveness was higher in the group that mirrored the self-limited disease, compared with the control group.

In Experiment 3, we modeled a self-limited disease in a different way: in addition to the null contingency between taking the treatment and observing a remission of the symptoms, we showed the symptom-occurrence baseline before the treatment starts. This should allow participants to fairly compare the P(O) before and after the treatment, even under the assumption that trials are dependent on each other — that taking the drug on one day can help to improve health days later. Even with this information available, participants overestimated the effectiveness of the pseudomedicine when the pattern of P(O) was increasing, as compared to the control group (Fig. 6). Thus, a different design, resting on different assumptions than those in Experiments 1 and 2, still produced similar results.

Experiment 3 presents an interesting question, as it is designed to work under the assumption of non-independence between trials. Most contingency learning experiments describe trials that can be naturally interpreted as mutually independent: for example, it is common to present a sequence of trials arranged in random order, each one corresponding to a different individual patient (Blanco et al., 2014; Blanco and Matute, 2019). Thus, it is unlikely that participants assume that treating one patient could affect subsequent patients. This was probably the case in Ejova et al. (2013) experiments, and it certainly was in the case of Matute (1995) and Langer and Roth (1975) studies. Under the assumption of trial independence, it is easy to reproduce the overestimation of effectiveness when a disease has a high probability of spontaneous remission (Blanco et al., 2014; Blanco and Matute, 2019). However, in our current procedure, it is possible to interpret the training as a time-series with dependencies between trials: if the patient takes the medicine today, the effects can be seen some days later. As we have argued, assuming that trials can be mutually dependent in this fashion has important consequences, the main one being that contingency, computed as the difference between P(O|C) and P(O|~C), stops being a useful index. This is a possibility that has been largely overlooked in most research in human contingency learning, with few exceptions (Bramley et al., 2015). In Experiment 3, we replicated the overestimation of effectiveness in these conditions, which extends the results of Ejova et al. (2013) to a new scenario. Future studies could deepen further in the implications of assuming dependency between observations.

Additionally, we collected information about the perceived effectiveness in two moments: on each trial (trial-by-trial judgments) and after the whole session (global judgment). It is well known that trial-by-trial judgments typically show recency effects: they are heavily affected by the information presented in the current or immediately previous trials (Matute et al., 2002). In global judgments, we asked participants to take into account all the information they had seen from the beginning of the session, a strategy that often results in the participants being able to integrate the information from the different phases (Matute et al., 2002). It would be useful to know whether people can be capable of integrating all the information and overcome their biases just by asking them to do so. Thus, we expected that global judgments in the self-limited group would not be so strongly influenced by the last part of the training session, which featured a very high P(O). This should probably reduce the illusion and blur the differences between the groups, given that they received exactly the same frequencies of trials when considering the overall session. However, global judgments in our experiments showed a familiar overestimation trend, with the self-limited group reporting stronger beliefs of effectiveness than the control group, although the effect was less pronounced than in trial-by-trial judgments (in Experiment 1, the difference between groups in the global judgments was just marginal). Therefore, this result is in line with previous studies showing that participants can better integrate all the information seen during the training at the end of the session if requested to do so, while trial-by-trial judgments are strongly affected by the most recent piece of information (Matute et al., 2002). Yet, note that the effect of increasing the P(O) during the training session survives despite the explicit instruction to try to avoid being too biased by the last part of the session. Consequently, the result highlights the view that self-limited diseases can potentially produce strong and resistant beliefs of effectiveness.

There were additional judgments in our task. First, through the Alternative Cause judgments, participants could report whether they felt that outcome occurrences (e.g., symptom remissions) could be attributed to causes other than the pseudomedicine. In Experiments 1 and 2, the results do not indicate any significant difference between groups. People do not tend to attribute health improvement instances to any unspecified alternative cause, irrespective of the group (average answers around 6 out of 10 points). In Experiment 3, in contrast, we found that participants in the self-limited group tended to judge that symptom remissions were due to the treatment more often than participants in the control group, which is in line with the predictions.

Second, our participants estimated the two conditional probabilities, P(O|C) and P(O|~C) — probability of symptom remission when taking the medicine versus when not taking the medicine. In Experiments 1 and 2, P(O|C) was systematically more overestimated than P(O|~C). The latter result is compatible with the presence of an illusion of causality, or overestimation of the effectiveness of the drug (Blanco and Matute, 2019), that exists in all groups, since participants seem to judge that improvements are generally more likely when one takes the medicine than when one does not. In fact, the difference between these two probability estimations (which is analogous to computing an estimated contingency) correlates significantly with effectiveness ratings in all six groups across the three experiments. In addition, in Experiment 3 not only this tendency to overestimate P(O|C) was found, but it was significantly stronger in the self-limited group, where effectiveness was judged higher, which also reinforces the idea that people in this group developed strong beliefs of effectiveness.

We can only speculate as to why the additional judgments (global judgment, Alternative Cause judgment, conditional probability judgments) seem to have captured stronger effects in Experiment 3 than in

Experiments 1 and 2. One of the most evident reasons is that, by separating the outcome base-rate training in one phase, instead of presenting it in intermixed trials, Experiment 3 made it much easier to grasp the base-rate to which compare the P(O) once the treatment starts. Another, somewhat more trivial, reason is that we started and ended the training using more extreme values of P(O), namely 0 and 1, which might create the impression that the medicine works at the end of the session.

When interpreting Experiment 3, one should be aware that the design does not perfectly control all potential extraneous variables. Of special relevance is the difference in the P(O) during Phase 2 between the two groups: that is, the probability of symptom remission is overall higher in this phase for participants in the self-limited group, P (O) = 0.83, compared to participants in the control group, P (O) = 0.50. This difference could contribute to the effect, as we know from previous literature that higher P(O) levels typically produce higher judgments, such as with outcome-density bias (Chow et al., 2019; Musca et al., 2010). Future studies could try to address this problem by including additional controls. For instance, it would be possible to add a stable, high P(O) control condition, in which symptom remissions appear as often as in the last part of the self-limited group. We note, however, that such control would still not be perfect, as it is impossible to fix all relevant parameters (outcome probability, contingency, number of trials, etc.) while one of the groups remains stable and the other increases in P(O).

These experiments can also help us shed some light on the mechanisms underlying the overestimation of causality more generally. Traditionally, these illusory beliefs have been explained by invoking differential "cell-weighting" processes (Kao and Wasserman, 1993). That is, of the four cells in the contingency matrix of Fig. 1, it is possible that people pay more attention to those events that are more salient — typically, type a trials, or medicine-remission events — or give them more importance when forming the belief. This proposal could explain both the present experiments as well as other instances of overestimation of effectiveness: even if the number of type a trials (medicine-remission) is the same as type c trials (no medicine-remission), people would weigh the former more heavily, thus strengthening the belief of effectiveness beyond the provided data. Concerning this possibility, we collected conditional probability judgments in addition to effectiveness judgments, and through the three experiments, these conditional probability estimations seem to show a tendency to overestimate the probability of remission conditional on the medicine, compared to when no medicine was taken, such that $P(O|C) > P(O|\sim C)$, especially in those individuals who developed stronger beliefs of effectiveness. This could be interpreted as an increased attention/importance given to medicine-present trials, which is in line with previous experiments investigating cell-weighting mechanisms (Kao and Wasserman, 1993).

5.4. Limitations

One obvious shortcoming of our procedure is the limited ecological validity, given that the situation is artificial, and participants must imagine how they would think in real life contexts. On the other hand, it seems difficult to experimentally study these processes of change in the beliefs of effectiveness with real treatments and outcomes. We believe that it would not be feasible for practical and ethical reasons, so our computer-based approach remains a valid candidate.

Another limitation of this research is that only two conditions were compared in each experiment: an increasing pattern of P(O) versus a stable pattern of P(O). Thus, we are not examining the potential role of consistency, but we are always comparing one group in which the outcome changes with another group in which it does not change. A more systematic approach would have included additional conditions: decreasing pattern of P(O), U-shaped pattern, etc. Some of these conditions are studied in previous experiments (Ejova et al., 2013; Matute, 1995), but, as they do not correspond well to any meaningful real situation in actual medicine usage, we decided to keep our experiments as simple as possible, by just comparing self-limited diseases with stable diseases. Future studies could put to test additional conditions.

In addition, this research rests on the assumption that the formation of effectiveness beliefs during the experimental session is due to basic processes (e.g., contingency learning) that are general-purpose, and work in similar ways for all people. Thus, we have not considered the potential impact of individual differences based on educational level, age, or attitudes, for example. This could be an interesting field of study for future research, as some studies have suggested that prior beliefs and attitudes can in fact modulate the formation of causal knowledge, leading to overestimations of causality similar to those reported here (Blanco et al., 2018). Moreover, it would be highly interesting to investigate whether unsupported beliefs of effectiveness actually translate into differences in treatment use, and whether this effect is further modulated by individual factors that could help us identify particularly vulnerable individuals or situations.

6. Conclusions

Previous research documents the risks of resorting to pseudotherapies, instead of using scientifically supported treatments, even for mild diseases. Since the use of pseudomedicine is at least partly driven by unsupported beliefs of effectiveness (Lilienfeld et al., 2014), studying how such beliefs develop is key to improve health and well-being. In our three experiments, we have showed how an ineffective medicine can produce inaccurate beliefs of effectiveness, especially when the disease tends to resolve spontaneously after some time (e.g., self-limited diseases). This finding could be relevant in real life, as those patients who incorrectly believe in the healing potential of a pseudotherapy are at risk of replacing scientifically valid treatments by other alternatives that lack evidential support (Macfarlane et al., 2020; Yarritu et al., 2015). Our results seem robust to variations in certain assumptions about the treatment course, such as the belief that treatments can have delayed effects, as we show in Experiment 3.

Additionally, our study presents implications for medical practice. In light of our results, self-limited diseases can be particularly dangerous, because they can induce beliefs of effectiveness in completely ineffective pseudotherapies. Moreover, with additional factors at play in real life (e.g., placebo effects that produce slight positive contingencies), the risk of effectiveness overestimation would be even higher. Therefore, it would be advisable that physicians and practitioners supervise patients suffering from these typically benign conditions, so that they can prevent any mistaken belief that could lead to dangerous actions in the future, such as replacing a valid treatment by a pseudotherapy in case a more serious disease appears.

CRediT authorship contribution statement

Fernando Blanco: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Visualization, Writing - original draft, Writing - review & editing. **Helena Matute:** Conceptualization, Funding acquisition, Writing - review & editing.

Acknowledgment

Agencia Estatal de Investigación of the Spanish Government (AEI) and European Regional Development Fund (FEDER), Grant PSI2017-83196-R. asque Government, Grant IT955-16.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2020.113012.

References

Allan, L.G., Jenkins, H.M., 1983. The effect of representations of binary variables on judgment of influence. Learn. Motiv. 381–405.

Barnes, P., Powellgriner, E., Mcfann, K., Nahin, R., 2004. Complementary and alternative medicine use among adults: United States, 2002. Semin. Integr. Med. 2, 54–71. https://doi.org/10.1016/j.sigm.2004.07.003.

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, e84084. https://doi.org/10.1371/journal.pone.0084084.

Blanco, F., Gómez-Fortes, B., Matute, H., 2018. Causal illusions in the service of political attitudes in Spain and the United Kingdom. Front. Psychol. 9, 1033. https://doi.org/10.3389/fpsyg.2018.01033.

Blanco, F., Matute, H., 2019. Base-rate expectations modulate the causal illusion. PloS One 14, 1–25. https://doi.org/10.1371/journal.pone.0212615.

Bramley, N.R., Dayan, P., Lagnado, D.A., 2015. Staying afloat on Neurath's boat – heuristics for sequential causal learning. Proc. 37th Annu. Conf. Cogn. Sci. Soc. 1, 262–267.

Chow, J.Y.L., Colagiuri, B., Livesey, E.J., 2019. Bridging the divide between causal illusions in the laboratory and the real world: the effects of outcome density with a variable continuous outcome. Cogn. Res. Princ. Implic. 4, 1–15. https://doi.org/10.1186/s41235-018-0149-9.

Christensen, R., Bedrick, E.J., 1997. Testing the independence assumption in linear models. J. Am. Stat. Assoc. 92, 1006. https://doi.org/10.2307/2965565.

de Barra, M., 2017. Reporting bias inflates the reputation of medical treatments: a comparison of outcomes in clinical trials and online product reviews. Soc. Sci. Med. 177, 248–255. https://doi.org/10.1016/j.socscimed.2017.01.033.

Ejova, A., Navarro, D.J., Delfabbro, P.H., 2013. Success-slope effects on the illusion of control and on remembered success-frequency. Judgm. Decis. Mak. 8, 498–511.

Ernst, E., Smith, K., 2018. More harm than good? The Moral Maze of Complementary & Alternative Medicine. Springer, Cham, Switzerland.

Freckelton, I., 2012. Death by homeopathy: issues for civil, criminal and coronial law and for health service policy. J. Law Med. 19, 454–478.

Gigerenzer, G., Hoffrage, U., 1995. How to improve Bayesian reasoning without instruction: frequency formats. Psychol. Rev. 102, 684–704.

Hellmuth, J., Rabinovici, G.D., Miller, B.L., 2019. The rise of pseudomedicine for dementia and brain health. J. Am. Med. Assoc. 321, 543–544. https://doi.org/10.1001/jama.2018.21560.

Johnson, S.B., Park, H.S., Gross, C.P., Yu, J.B., 2018. Complementary medicine, refusal of conventional cancer therapy, and survival among patients with curable cancers. JAMA Oncol. 4, 1375–1381. https://doi.org/10.1001/jamaoncol.2018.2487.

Kao, S., Wasserman, E.A., 1993. Assessment of an information integration account of contingency judgment with examination of subjective cell importance and method of information presentation. J. Exp. Psychol. Learn. Mem. Cogn. 19, 1363–1386.

Langer, E.J., Roth, J., 1975. Heads I win, tails it's chance: the illusion of control as a function of the sequence of outcomes in a purely chance task. J. Pers. Soc. Psychol. 32, 951–955.

Lilienfeld, S.O., Ritschel, L.a., Lynn, S.J., Cautin, R.L., Latzman, R.D., 2014. Why ineffective psychotherapies appear to work: a taxonomy of causes of spurious therapeutic effectiveness. Perspect. Psychol. Sci. 9, 355–387. https://doi.org/10.1177/1745691614535216.

Macfarlane, D., Hurlstone, M.J., Ecker, U.K.H., 2020. Protecting consumers from fraudulent health claims: a taxonomy of psychological drivers, interventions, barriers, and treatments. Soc. Sci. Med. 112790. https://doi.org/10.1016/j.socscimed.2020.112790.

Matute, H., 1995. Human reactions to uncontrollable outcomes: further evidence for superstitions rather than helplessness. Q. J. Exp. Psychol. Sect. B 48, 142–157. https://doi.org/10.1080/14640749508401444.

Matute, H., Blanco, F., Díaz-Lago, M., 2019. Learning mechanisms underlying accurate and biased contingency judgments. J. Exp. Psychol. Anim. Learn. Cogn. 45, 373–389. https://doi.org/10.1037/xan0000222.

Matute, H., Vegas, S., De Marez, P.-J., 2002. Flexible use of recent information in causal and predictive judgments. J. Exp. Psychol. Learn. Mem. Cogn. 28, 714–725.

Moreno-Fernández, M.M., Blanco, F., Matute, H., 2017. Causal illusions in children when the outcome is frequent. PloS One 12. https://doi.org/10.1371/journal.pone.0184707.

Musca, S.C., Vadillo, M.A., Blanco, F., Matute, H., 2010. The role of cue information in the outcome-density effect: evidence from neural network simulations and a causal learning experiment. Connect. Sci. 22, 177–192. https://doi.org/10.1080/09540091003623797.

NHMRC, 2015. Statement on Homeopathy. National Health and Medical Research Council - Australia, Sydney, Australia.

Palan, S., Schitter, C., 2018. Prolific.ac—a subject pool for online experiments. J. Behav. Exp. Financ. 17, 22–27. https://doi.org/10.1016/J.JBEF.2017.12.004.

Proctor, M., Hing, W., Johnson, T.C., Murphy, P.A., Brown, J., 2006. Spinal Manipulation for Dysmenorrhoea. Cochrane Database Syst. Rev. https://doi.org/10.1002/14651858.cd002119.pub3.

Rottman, B.M., Marcum, Z.A., Thorpe, C.T., Gellad, W.F., 2017. Medication adherence as a learning process: insights from cognitive psychology. Health Psychol. Rev. 11, 17–32. https://doi.org/10.1080/17437199.2016.1240624.

Self-limited disease, 2012. Farlex Partn. Med. Dict.

Yarritu, I., Matute, H., Luque, D., 2015. The dark side of cognitive illusions: when an illusory belief interferes with the acquisition of evidence-based knowledge. Br. J. Psychol. 1–12. https://doi.org/10.1111/bjop.12119.