LUNDS UNIVERSITET

INSTITUTIONEN FÖR PSYKOLOGI

Debiasing for everyone: Testing an educational intervention to reduce causal illusions in rural Kenya

Linnea Cederlund
Henrik Josander

Kandidatuppsats VT 2017

Handledare: Anna Kemdal Pho
Examinator: Roger Johansson

Abstract

It has been argued that cognitive biases are the source of a range of problems in modern society, from stereotype formation to belief in pseudoscience. One of those biases causes people to perceive causal relationships between unrelated events and is known as illusion of causality. Scientists are pointing to the importance of strategies to debias people on a global scale. However, research on effective debias techniques is lacking. The purpose of this study was to examine if results from previous studies on illusion of causality could be generalized to another context. Specifically, a debias intervention that had been found to be effective in Europe was examined in rural Kenya. To the best of our knowledge, this was the first study to examine the effectiveness of a debias intervention aimed to reduce illusions of causality, outside Western countries. The intervention consisted of a lesson in basic scientific thinking, which has been argued to be the best, if not the only, way to reduce or eliminate illusion of causality. A repeated measures design was used, and illusions of causality were assessed with a contingency judgment task. The results showed, in contrast to the main hypothesis, that illusions of causality persisted after the intervention. Further research is needed to assess if the educational intervention is a globally effective strategy to debias people from illusions of causality.

Keywords: illusion of causality, bias, debias, intervention, contingency judgment task, basic scientific thinking, Kenya

Introduction

Wilson and Brekke (1994) succinctly refer to cognitive biases as ‘mental contamination’, and to debiasing as ‘mental correction’. We are all vulnerable to ‘mental contamination’, and it strongly affects our judgments and our decision-making (Croskerry, Singhal & Mamede, 2013). The consequences of the contamination, the biases, can be severe, even harmful (e.g. Barberia, Blanco, Cubillas & Matute, 2013).

One specific bias that the human mind is evolutionary vulnerable to is illusion of causality. Illusion of causality is developed when people perceive that there is a causal relationship between events that are in fact unrelated (Barberia et al., 2013). Illusions of causality seem to have sparked a newfound interest among researchers in recent years (e.g. Barberia et al., 2013; Blanco, Matute & Vadillo, 2011; Yarritu, Matute & Vadillo, 2014). Barberia et al. (2013) successfully tested an intervention consisting of education in scientific thinking to reduce illusions of causality.

Regarding the mental correction, the debiasing, some researchers have been pessimistic about our capacity to correct our thinking (e.g. Willingham, 2008). Currently, researchers seem to have an optimistic perspective on debiasing. For example, Croskerry et al. (2013) argue that “..clearly people can change their minds and behaviours for the better.” (p. 23).

There seems to be a consensus among researchers in the area about the importance of designing a worldwide strategy to debias people (e.g. Barberia et al., 2013; Matute et al., 2015). Despite this, and to the best of our knowledge, studies on debiasing against causal illusions have only been conducted in Western countries. With the aim to fill a bit of the gap in this worldwide debiasing strategy, this study was conducted in rural Kenya to explore to what extent findings from previous research could be generalized to a different context. After all, as Lilienfeld, Ammirati and Landfield (2009) put it: “..a plausible case can be made that debiasing people against errors in thinking could be among psychology’s most enduring legacies to the promotion of human welfare.” (p. 391).

Cognitive biases

The main psychological theory that this thesis aims to investigate is that of cognitive biases, and techniques for reducing cognitive biases in people. In short, cognitive biases are byproducts of cognitive processing limitations. In other words, distortions to mental representations or cognitive processes in the human (and animal) mind (Trimmer, 2016).

Some cognitive biases may be viewed as design flaws of the mind when looked upon from one perspective, or as design features when looked upon from another perspective. From an evolutionary perspective most features of the human mind serve an adaptive function. These functions are mostly domain specific and might therefore not generalize well to different domains than the ones they were “designed” for. For that reason, these evolutionary adaptive functions are prone to produce biases when applied in the “wrong” context (Haselton, Nettle & Andrews, 2015).

Most biases are associated with heuristics, which are adaptive mechanisms that save effort and time in everyday life decision-making and problem solving. To think about every event and stimuli that occur on a daily basis in an analytical and focused manner would be impossible (Croskerry et al., 2013). So, we use mental shortcuts, heuristics (Tversky & Kahneman, 1974). Most of the time our minds are set to an intuitive mode in which one event or thought automatically triggers the next and so on. We tend to use heuristics often, and while they mainly work well they frequently result in errors in thinking and decision-making (Croskerry et al., 2013). These systematic errors are what we call biases (Kahneman, 2011). Heuristics are natural mechanisms in the human mind, and consequently, everyone is affected by biases (Lilienfeld, Ammirati & David, 2012). For several biases there are no correlations with intelligence, but for other biases there are modest correlations. However, as Stanovich (2010) puts it “Overall, the associations are surprisingly modest.” (p. 121).

Causes of cognitive biases

Arkes (1991) argues that a wide range of biases can be understood by three underlying causes, namely strategy-based judgment errors, association-based judgment errors and psychophysically based errors.

Firstly, the strategy-based judgment error occurs when a poor decision is made that may be thought to be beneficial in a larger sense. If a person uses a quick but not very thoroughly elaborated strategy to solve a problem it could be argued that the benefit of saving time and cognitive effort may be larger than the cost of errors that occurs because of the poorly chosen strategy. However, when errors occur from such a strategy, they are considered biases.

Secondly, the association-based judgment error is related to errors in the automatic processes that underlie the retrieval of information in memory. One bias caused by association-based errors is the bias of ignoring, which is closely related to illusion of causality, the main bias of this study. The bias of ignoring occurs when people take into account only the presence of a potential cause to predict its outcome, but ignores the instances when the outcome is present and the potential cause is absent.

Finally, the psychophysically based error occurs when people are presented with a reference point and make an estimate that is based more on the reference point than on other relevant data. Tversky & Kahneman (1974) illustrated a bias caused by this error in an experiment. In one test, subjects were asked to estimate the percentage of African countries in the UN. One group was presented with 10 as an initial value and had a median estimate of 25. Another group that was presented with 65 as an initial value had a median estimate of 45. Thus, people seemed to adjust their estimates based on the initial value that they were presented with.

The three underlying causes presented above are related to judgment behaviors and judgment- and decision biases. There are however other types of biases and some biases may have more than one single underlying cause. Still, categorization of different biases based on their underlying causes does hold a value when it comes to techniques for debiasing (Arkes, 1991).

The dual process theory is important when discussing the origins of biases, and the theory is heavily supported in research. According to it, there are two distinctive types of cognition. The two types have been named differently in various studies (Stanovich, 2010). We will call them type 1 and type 2 processes, since these names are the most neutral ones (Evans, 2008).

Type 1 processes are fast, intuitive and demand little, if anything, from the thinker. In fact, they are automatic and we are unaware of them. Type 2 processes are slower, analytical, more demanding and we are in control of these processes. When type 1 processes are in action we can make routine decisions, but when it becomes more advanced, type 2 processes are needed (Kahneman, 2011). Type 1 processes are known to make systematic errors, biases, and as a consequence result in irrational responses (Kahneman, 2011; Stanovich, 2010). This means that type 2 processes are crucial in overriding type 1 processes (Stanovich, 2010). Biases can occur in type 2 processes as well, but they are most often associated with type 1 processes. Decisions can either be made through type 1 processing or type 2 processing. However, the brain tends to make use of type 1 processing, prone to biases and more likely to fail, more than type 2 processing. In fact, we spend most of our time using type 1 processes. The override function of type 2 processes is vital for debiasing because only type 2 processing can correct biases (Croskerry et al., 2013).

Bias and rationality

Arnott (2006) writes that “One way of viewing cognitive biases is as predictable deviations from rationality” (p. 59). Stanovich (2010) discusses what he calls “the Great Rationality debate”. This debate regards human cognition and irrationality, and there are two main perspectives in this matter. Followers of the first perspective are called meliorists. They believe that there is a gap between on the one hand, normative models of rational responding and on the other hand, descriptive models of how people actually respond. Simply put, they believe that people act irrationally at times and that people need and can improve their thinking, for example through education (Stanovich, 2010). Researchers that study biases are in many cases meliorists (Stanovich, 1999, cited in Stanovich, 2010).

Followers of the second perspective are called Panglossians. According to them, there is no gap between the normative and the descriptive. They believe that the way that people respond is the normative way. If a heuristics and bias researcher works with a model that fails to predict people's responses the model should be changed because the people are the norm, the model is not. Panglossians do not believe that people are irrational (Stanovich, 2010).

Illusion of causality

Illusion of causality, or causal illusion, is a bias that occurs when we develop the belief that one event causes another, when in fact there is no relationship between the two events (Matute et al., 2015). The events are independent of each other, but due to chance they appear together (Barberia et al., 2013). Illusions of causality "..occur because of the way the human mind has evolved: It extracts causality from coincidences." (Matute et al., 2015, p. 3). It is common knowledge that we have developed abilities through natural selection that help us achieve the wanted, and to avoid the unwanted (Blanco et al., 2011). By having developed the ability to identify causal relationships, we can predict events and prepare ourselves for them, which means that we can influence our environment (Greville & Buehner, 2010). However, the tendency to detect causal relationships is so strong that we identify causal relationships even when they do not exist (Barberia et al., 2013). Causality itself cannot be observed, it has to be inferred from evidence that can be observed (Hume, 1739/1888, cited in Greville & Buehner, 2010). The evidence that indicates a causal relationship is contingency and contiguity. Contingency, or regularity, is necessary to infer causality because it means that events reliably and regularly follow each other. Contiguity refers to the closeness in time between events (Buehner, 2005).

Illusion of causality is regarded as the basis of beliefs in pseudoscience (Matute et al., 2011) and may be the basis of for example stereotypes (Hamilton & Rose, 1980), racism and economical collapses (Barberia et al., 2013).

According to Barberia et al. (2013), what some studies call "illusion of control" (e.g. Blanco et al., 2011) is simply a specific case of illusion of causality. Therefore, Barberia et al. (2013) include illusion of control in the more general term illusion of causality. This study will follow that approach.

How to assess illusions of causality

The contingency judgment task, or contingency learning task, has often been used as a way to assess illusion of causality (Blanco et al., 2011; Crump et al., 2007; Barberia et al., 2013).

Research on contingency judgments was initiated in the 1960s, when researchers started questioning early learning theorists' notions about Pavlovian conditioning. It was believed that temporal contiguity alone was enough to elicit conditioning between a conditioned stimulus (CS) and an unconditioned stimulus (US) in animal subjects (Allan, 1993). In contrast to earlier notions, the new idea of contingency took into account how often events are not paired as well as how often events are paired (Allan, 1993). The difference of the probability of the US in the presence and in the absence of the CS is contrasted to give the contingency between events. In accordance with these new theories, Rescorla (1968) conducted two experiments to test the ideas. Results from these experiments confirmed that both contingency and temporal contiguity were needed to elicit conditioning. Although these experiments were conducted to assess Pavlovian conditioning, the procedural similarities between human contingency learning and conditioning are striking. Research like that of Rescorla (1968) sparked an interest in human contingency learning within experimental psychology. The literature on human contingency learning has since then developed and several theoretical models of explanation have been introduced (for a review, see De Houwer & Beckers, 2002).

In a contingency judgment task, two variables are paired over several trials. Then, the strength of the relationship, the contingency, is rated. People tend to overestimate the contingency between the variables. The overestimation occurs due to illusion of causality (Blanco et al., 2011). In the task, a potential cause is either presented or not presented, followed by either the presence or the absence of an outcome. The cause and the outcome are often illustrated with examples, such as rain (the cause) that is presented or not presented, followed by plant growth (the outcome) that occurs or does not occur. This creates four possible cause-outcome pairings that can have different frequencies (see table 1). The frequencies of the cause-outcome pairings can be manipulated to change the contingency between the cause and the outcome (Blanco et al., 2011). The potential cause has also been called "cue" (e.g. Blanco et al., 2011). In this study, the term cause or potential cause will be used, to emphasize that it is potentially causing the outcome.

Table 1. *A 2x2 matrix showing the four possible cause-outcome pairings.*

	Outcome present	Outcome absent
Cause present	A	B
Cause absent	C	D

There are generally two different types of contingency judgment tasks. The first type is a passive task with a preprogrammed sequence of cause-outcome pairings that the participant merely observes before rating the contingency. The second type is an active task in which participants are allowed to decide when the cause should be presented and when it should not. The active version of the task can have a preprogrammed and fixed order of when the outcome is present, so that the participants' decision to present the cause or not doesn't have any effect on the presence or absence of the outcome. In this case, it is a zero contingency and the outcome is independent of the cause. Consequently, the contingency should be rated as 0 to be correct. The active version of the task can also be programmed to show the outcome in a preprogrammed percentage of cause-present trials. This means that if the participant decides to present the cause, there is, for example, 75% chance that the outcome occurs (Matute et al., 2015).

After several cause-outcome pairings have been presented, the participant is asked to rate the strength of the contingency between the cause and the outcome on a scale (Crump et al. 2007). Some studies use a unidirectional scale, in which the scale ranges from 0 to a positive value (usually 100) representing a positive contingency. Other studies use a bidirectional scale, in which the scale goes from a negative value (usually -100) representing a negative contingency, to a positive value representing a positive contingency. A rating of 50 would mean a positive contingency of average strength. In addition, how often the participant has presented the cause, i.e. the response ratio, is of interest. It is called the probability of the potential cause, P(cause), and the value expressed in percent reveal how often the participant presented the cause divided on the total amount of trials (Allan, 1993).

The appropriate way of measuring the dependency of one variable on another is the ΔP index (Allan, 1993). The ΔP index is a normative measure of contingency (Matute et al., 2015), produced by the difference between the probability of the outcome given the cause, and the probability of the outcome in the absence of the cause: $\Delta P = A/(A+B) - C/(C+D)$ (Jenkins & Ward, 1965). ΔP values can range from -1, representing a negative contingency, to 1, representing a positive contingency. A ΔP value of 0 means that there is no contingency between the variables. A positive ΔP value means that the probability that the outcome occurs is higher in the presence of the potential cause, than in its absence. The cause is therefore potentially contributing to producing the outcome. A negative ΔP value means that the probability that the outcome occurs is higher in the absence of the potential cause, than in its presence. In this case, the potential cause is not producing the outcome but rather preventing it.

The ΔP index is used to calculate participants' actual contingencies in the contingency judgment task (Matute et al., 2015). The actual contingency is the contingency that the participants experience based on when they decide to present the potential cause in an active contingency judgment task. That is, even if the preprogrammed contingency is zero, participants might experience a contingency higher than zero if they happen to present the cause on those precise trials when the outcome has been preprogrammed to be present (Blanco et al., 2011). Some research do show that people are sensitive to actual contingencies (e.g. Shanks & Dickinson, 1987), but much research show that people are biased in their judgments of contingencies (Barberia et al., 2013).

Increasing the risk of illusions

Several variables are known to increase the likelihood of illusion of causality. The first variable is the probability of the outcome. When the outcome occurs often, people are more likely to overestimate the contingency. This is called the outcome-density bias (Matute et al., 2015). The second variable is the probability of the cause. If the cause is presented often, this will increase the likelihood of causal illusions. This effect is called the cause-density or cause-frequency bias (Matute et al., 2015). Blanco et al. (2013) calls this variable the probability of responding, P(R), and distinguishes a difference between the former and the latter. Namely, the effect is called cause-density effect when the cause is external, as in a passive task, and the probability of responding, P(R), effect when it appears in an active task, when the participant controls when to respond. In the present study, this effect will be called the probability of the potential cause, P(cause), as it is called in Barberia et al. (2013).

In active tasks that let the participants decide when to present the cause, participants tend to present the cause often. This seems connected to a general hypothesis testing strategy called positive testing strategy (Barberia et al., 2013). Using a positive testing strategy when testing a hypothesis means that you focus on the cases that can confirm the hypothesis, not on the ones that can disconfirm the hypothesis (Klayman & Ha, 1987). Some would like to connect this to confirmation bias, but according to Klayman and Ha (1987) this is not suitable because making this connection suggests that the positive testing strategy is faulty. In fact, the strategy can be a useful heuristic in many situations. However, since it is a shortcut, it can lead to problems when used incorrectly. One evident problem that can arise from using the strategy is that some events receive too much attention while others receive too little, resulting in inaccurate responses (Klayman & Ha, 1987). In a contingency judgment task, using the strategy means that you present the cause often, resulting in a high P(cause) value, and focus on the cause-present cases. Then, if the possibility of the outcome is high, exposure to cases where the outcome appears without the cause is reduced (Barberia et al., 2013). For example, imagine that people often recover from a sickness without being given medicine. If you give medicine to these people frequently, you will fail to notice that people often recover without the medicine. That is, you will fail to notice that the outcome (the recovery) frequently is present when the cause (the medicine) is absent. In this case, the strategy is likely to result in an inaccurate response, since the medicine-absent events receive too little attention.

Lagnado and Sloman (2006) points out a third variable that can increase the illusion of causality, and it has to do with the closeness in time between events. The closeness in time between events is a necessary cue for us when we infer causality, but it can mislead us to infer causality where none exists. So, from events that by chance appear close in time we can erroneously draw the conclusion that one causes the other. Barberia et al. (2013) calls this variable “cause-outcome coincidences” (p.5), which reflects the idea that if the potential cause and the outcome coincides, we are likely to draw the conclusion that the potential cause in fact causes the outcome, even when the two only coincide by chance.

Consequently, in order to decrease the illusion of causality, either the frequency of the cause or the outcome can be decreased. However, in real life it is usually not possible for a person to decrease the frequency of the outcome. What remains is the possibility to decrease the cause, and this is often within a person’s control. For example, a person might not be able to decrease how often she becomes ill, the outcome, but she can decrease how often she takes the medicine, the cause, to find out if the medicine cures the illness (Barberia et al., 2013). To conclude, even though we might be unable to control the frequency of the outcome, being aware of the fact that a frequently occurring outcome can increase causal illusion means that we can prepare for these sorts of situations and try to make sure that we won’t develop the illusions (Matute et al., 2015).

Debiasing

Debiasing is defined as “a procedure for reducing or eliminating biases from the cognitive strategies of a decision-maker.” (Arnott, 2006, p. 62). By debiasing our thinking, we become better thinkers (Croskerry et al., 2013). 0

Teaching people scientific thinking might be the best, if not the only, way to debias people from illusion of causality (Matute et al., 2015). This has to do with the fact that people don't have the ability to assess causality naturally. Matute et al. (2015) mean that "..scientific methods should always be used when assessing causality." Unfortunately, people don't think scientifically by nature (Lilienfeld et al., 2012). Scientific thinking can reduce or eliminate illusion of causality by influencing people to: refrain from presenting the cause often, observe what happens when the cause isn't presented, make sure to look for complete information and be aware of the possibility of being biased. It is not only important to learn to use basic scientific thinking, it seems to be if not even more important to learn in what situations this kind of thinking is crucial (Matute et al., 2015).

Lilienfeld et al. (2009) point out several aspects that can negatively affect the success of debiasing techniques. Firstly, people don't believe that they are affected by bias. This is called the bias blind spot. People can acknowledge that biases exist and that they affect others, but people believe that when it comes to themselves, they perceive the world objectively (Pronin, Lin & Ross, 2002). Secondly, people fail to realize that debiasing is valuable for their everyday life. For some debiasing interventions this might mean that in order to be effective, participants need to be shown the consequences of biased decisions taken in the daily life (Lilienfeld et al., 2009). Barberia et al. (2013) included a deceptive part in their debias intervention to show the participants how easily they could be affected by bias, in order to avoid the bias blind spot aspect. In addition, the deceptive part had connections to daily life, so that the participants would more easily recognize the value of the debiasing (Lilienfeld et al., 2009). Thirdly, issues connected to scientific thinking can make debiasing efforts unsuccessful. Scientific thinking is both hard to teach and to do. For instance, telling people what they should do when facing a problem doesn't mean that they will be able to implement the advice. In order to think scientifically, people need practice as well as knowledge about the specific domain that the problem appears within. Furthermore, people tend to focus on the surface structure of problems, rather than on the deep structure. For example, people might not realize that two mathematical problems have the same deep structure, i.e. require the same mathematics, because they focus on the surface structure, the differing scenarios that the problems appear in. However, through experience people can learn to recognize the deep structure, and then the knowledge of how to solve a problem can transfer to other problems with the same deep structure (Willingham, 2008).

Lilienfeld et al. (2009) mean that there is a need for a lot more research on effective debiasing techniques, on what makes them effective and how the knowledge gained in them can be generalized for use in everyday life. Most techniques that have been developed so far don't have a proper connection to everyday life. In fact, so far researchers have spent more effort on discovering biases than on constructing techniques to debias people (Lilienfeld et al., 2009). Since debiasing faces plenty of difficulties, any good techniques that are invented to debias might only achieve a small change. However, according to Lilienfeld et al. (2009), the topic of debiasing should be given priority. Debiasing people will lead to a wiser, and hopefully safer, world.

Previous studies

The experiment of this study is a based on the experiments of two earlier studies made by Blanco et al. (2011) and Barberia et al. (2013).

In the study by Blanco et al. (2011), two experiments were conducted to assess the participants' illusion of control. University students from an introductory course in psychology were tested in a computerized contingency judgment task. The contingency judgment task was set up like a computer game in which the participant was to imagine that he or she was a medical doctor. The participants' task was to find out if the medicine was effective in healing a fictitious disease. In the first experiment, the participant was presented with 50 fictitious patients suffering from the fictitious disease. For each patient, the participant could choose to give or not to give a medicine, the potential cause, and then the participant received feedback regarding the patient's health. If the patient was healed, the desired outcome had occurred. It was then an outcome-present case. If the patient was not healed, it was an outcome-absent case. After the participant had been presented with all the 50 patients, he or she was asked to judge how effective the medicine was on a scale from 0 – 100. A second, similar experiment, in which the number of trials was doubled, was performed by Blanco et al. (2011). The idea was to examine if the results from the first experiment were due to insufficient training, and that the overestimation of contingency that was found in the first experiment therefore would not be seen in the second, longer experiment, as predicted by some models (e.g. Rescorla & Wagner, 1972). In the experiments, the outcome was fixed to be present in 75% and 76% of trials respectively, in a randomized order. The percentages differed between the experiments because of the difference in the total number of trials (Blanco et al., 2011).

Results from the two experiments showed that there was a significant difference between the participants' judgments of contingency and the preprogrammed contingency of 0. This means that the participants overestimated the contingency in both the experiments, that is, the participants had developed illusions of control (Blanco et al., 2011). The main finding in the first experiment replicated findings from Matute (1996), namely that higher frequency of presenting the cause, P(cause), resulted in higher judgments of contingency. In other words, the participants who presented the cause in most of the trials developed the strongest illusions of control. As mentioned previously, this is called the P(cause) effect. In addition, these results meant that judgments of contingency could be predicted by how often participants chose to give the medicine. This got further support in the second experiment (Blanco et al., 2011).

Results from the first experiment revealed that the actual contingencies that the participants experienced were close to zero, but there was a significant difference between the actual contingencies and the preprogrammed contingency of 0. There was also some variability in the actual contingencies that the participants experienced. Blanco et al. (2011) tested if actual contingency went up when frequency of presenting the cause, P(cause), went up. This turned out to be the case. However, a regression analysis showed that the actual contingency could not predict the participants' judgment of contingency (Blanco et al., 2011). In the second experiment, the actual contingencies did not significantly differ from the preprogrammed contingency of 0, and the variance of the actual contingencies was reduced. Blanco et al. (2011) meant that these results taken together indicated that the overestimation of contingency by the participants did not develop as a result of the actual contingencies that they experienced.

The theory of insufficient training was not supported by the results. According to the theory, the overestimations seen in the first experiment should have decreased in the second experiment. However, the results revealed the opposite. The mean value for judgments of contingency in the second experiment was significantly higher than in the first experiment. So, the overestimation of contingency became stronger when the amount of trials was increased. This was measured by P(cause), since it was an indirect measure of judgment of contingency (because of its ability to predict judgment). Consequently, the results showed that the participants did not decrease their responses during the increased amount of trials; they increased them. That is, the participants' illusions of causality became stronger (Blanco et al., 2011).

The other study, conducted by Barberia et al. (2013), used a between-groups design to test the effectiveness of a debias intervention aimed to prevent the formation of illusions of causality in secondary school students (the average ages were 14.26 in the control group and 14.84 in the experimental group). According to Barberia et al. (2013), they were the first to test an educational intervention aimed to reduce illusions of causality.

A computerized contingency judgment task was used to assess illusions of causality in the participants. The intervention consisted of two separate phases. The first phase involved staging a scenario in which a bogus product was said to enhance physical and intellectual abilities. This part of the intervention was aimed to show the participants how easily illusory perceptions of causality can develop. The participants were then told that the product was fake, and that they had been deceived. The second phase of the intervention involved educating the participants about contingency information as the correct way to infer causality. The phase also involved information about the importance of comparing the probability of an outcome in both the presence and the absence of the potential cause. Different examples were used to illustrate the concepts that were covered in the second phase. The control group did the contingency judgment task before taking part in the intervention, and the experimental group did the contingency judgment task after taking part in the intervention (Barberia et al. 2013). In the contingency judgment task the same cover story was adopted as in the experiment of Blanco et al. (2011). Participants were told that they would play a computer game, in which they were asked to imagine that they were medical doctors that would be presented to 40 fictitious patients suffering from a fictitious disease. The participants were not aware that this computer game was a part of the experiment. For each patient the participant could choose to give or not to give a medicine (cause), and directly afterwards they found out if the patient was healed (outcome) or not. After the participants had been presented with all the 40 patients they were asked to judge how effective they thought the medicine was on a scale from 0 – 100. The participants did the contingency judgment task twice. The first task was preprogrammed as a zero contingency condition in which the outcome (patient being healed) was present in 75% of cases, regardless if the cause (medicine) was presented. So, 30 out of 40 patients were healed, with or without medicine. However, the second task was a positive contingency in which 1 out of 8 patients who did not get the medicine were healed, and 6 out of 8 patients who got the medicine were healed. The positive contingency condition was added to make sure that the experimental group didn't simply rate the contingency low because they were suspicious after having been deceived in the intervention (Barberia et al., 2013).

As in the study of Blanco et al. (2011), the participants developed causal illusions. However, the intervention was successful. In the zero contingency condition, the participants in the experimental group made more accurate judgments of the contingency compared to the control group. In the positive contingency condition, both groups judged the contingency similarly and fairly accurately. Taken together, the results showed that the participants in the experimental group did not rate the contingency lower because they were suspicious after the intervention. Rather, they were able to detect when evidence indicated a contingency, and when it didn't. In addition, as the researchers had expected, the control group presented the cause more often than did the experimental group. To conclude, the debias intervention affected the participants in the experimental group in two ways. It had a significant direct effect on their contingency judgments, as well as a significant indirect effect on their judgments by affecting them to decrease their presentation of the cause resulting in a lower P(cause) value (Barberia et al., 2013).

Purpose and hypotheses

Researchers point to the importance of a future worldwide debiasing program (e.g. Barberia et al., 2013; Matute et al., 2015). However, to the best of our knowledge, studies have only been conducted in Western countries so far. If a future worldwide debiasing program is the objective, research needs to be performed in other contexts as well, in order to establish external validity of results from previous studies. Moreover, several researchers are pointing to the gap in research regarding debias interventions (e.g. Lilienfeld et al., 2009; Barberia et al., 2013; Matute et al., 2015). This study aims to examine if results from previous studies on illusion of causality (Blanco et al., 2011; Barberia et al., 2013) and a specific debias intervention (Barberia et al., 2013), can be generalized to another context. More specifically, the study will examine the effectiveness of a debiasing intervention aimed to reduce causal illusions, conducted in rural Kenya. To the best our knowledge, this is the first study that uses a repeated measures design when testing a debias intervention aimed to reduce causal illusions. Furthermore, to the extent of our knowledge, it is the first study on debiasing against causal illusions conducted outside Western countries. Based on previous research we have the following four hypotheses:

1. The participants will overestimate the contingency in the contingency judgment task.
2. The study's main hypothesis is that the participants will make more accurate estimates of contingency after the intervention, than before the intervention.
3. The actual contingency will not predict the participants' estimates of contingency.
4. The frequency of the participants' responses, P(cause), will predict the participants' estimates of contingency.

Method

Participants

The participants consisted of 30 people, 13 women and 17 men, aged 16-51 (*M* = 27, *SD* = 8.27) from rural Kenya. The mean number of years of education was 13.37 (*SD* = 3.03).

The sample was selected through convenience sampling with the help of a local non-governmental organization based in west Kenya. The organization has a large social network and social media was used to request participants. The organization had been informed that the participants needed to be over the age of 15, to be able to read and write and that preferably an equal amount of women and men would participate.

Material

To assess illusions of causality, a contingency judgment task with a zero contingency was used. In the pre-test, the outcome was preprogrammed to be present in 77,5% of the trials in a randomized order which was fixed and identical for all participants. In the post-test, the outcome was preprogrammed to be present in 75% of trials in a randomized order that was fixed and identical for all participants. So, there was no relationship between medicine and recovery. The high ratio of outcome-present trials was used as it has been shown to promote overestimations of contingency (Matute et al., 2015). The order in which patients were healed or not healed differed from pre-test to post-test. The reason for the difference in outcome percentages between the pre-test and the post-test was a mistake that occurred when the task was prepared. However, we believe that this error did not influence the results. If it did the, influence was minimal. This will be examined in more detail in the discussion.

In previous contingency judgment tasks, computers have been used. (Blanco, et al., 2011; Crump et al. 2007; Barberia et al. 2013). In this study, both papers and computers were used. Papers were used to simulate patients, to give feedback to the participants, to let the participants rate effectiveness and to let the participants rate their understanding of the intervention. The paper patients consisted of a paper with a sad smiley, and the only difference between them was the patient number on the top of the page. The feedback to the participants consisted of a paper with either a happy smiley with the text “The patient is healed!”, or a sad smiley with the text “The patient was not healed”. The paper used for effectiveness rating consisted of a scale from 0 to 100, “Ineffective” to “Entirely effective”, and a text saying ”To what extent do you think “Elovix” has been effective to heal the patients?” or ”To what extent do you think “Batatrim” has been effective to heal the patients?”. The paper for rating understanding of the intervention consisted of a scale from 1 to 5,” Not at all” to “Very well”, and the text “How well did you understand the lesson?”.

Computers were used to record general information about the participants, to record the participants’ responses, and to look up what feedback the participants should receive. Papers were partly used in the experiment because we were unable to find a preexisting computerized task that could be used in the test, and we didn’t have the skills to programme one ourselves. In the intervention, a flip board was used to illustrate examples. Statistical analysis was done using SPSS. The ΔP index was used to calculate the participants’ actual contingencies. Materials used in the task can be found in the appendix.

Design and procedure

The experiment had a repeated measures design and consisted of a pre-test, an intervention and a post-test. The tests consisted of the same contingency judgment task, set up as a game. The only alteration from pre-test to post-test was the names of the medicine and disease.

The participants arrived at different times. In small groups they were informed that they would participate in a test, in a lecture and then in a test similar to the first one. They were also told the instructions of the game (see appendix). Then, after the participants had been asked if they had any questions, they were asked to sign a letter of consent.

All participants were given the same instructions. In the first game, the participant was to imagine that he or she was a medical doctor. As this medical doctor, the task was to find out if a new, imaginary medicine that was called “Elovix” was effective or not in healing a fictitious disease that we called “MacGregor’s syndrome”. The participant was going to be presented with forty fictitious patients, and for each patient be given the choice to give or not to give “Elovix” to the patient. At the end of the game, the experimenter was going to ask the participant to rate the effectiveness of the medicine.

Besides the authors of this study, a third experimenter was recruited and trained to assist with the contingency judgment task. In the pre-test, the participants were randomly paired with one of the three experimenters to do the game. In the following, we will explain in detail how the experiment was conducted.

First, the participant was seated opposite the experimenter. The experimenter asked the participant for general information, such as age, education level and gender. Since some people had to wait for a while before they were tested, the experimenter made sure to ask if the participant had any questions before the experiment started. At this point, few questions were asked. Then, the experimenter started the game and showed the paper with the first patient to the participant. The participant told the experimenter if he or she wanted to give or not to give the medicine to patient one. The experimenter recorded the response in the computer and looked up the preprogrammed order to find out what response to give to the participant, if the excel sheet showed “healed” or “not healed”. If it showed “healed”, the experimenter held up the paper with a happy smiley and the text “The patient is healed!”. If it showed “not healed”, the experimenter held up the paper with a sad smiley and the text “The patient was not healed”. Then the experimenter showed patient number two, and the same procedure was repeated. This was repeated for all 40 patients. The computer was facing away from the participants to make it look like it was computing an answer, so that the participants would not figure out that the outcome was preprogrammed. When the participant had given a response for all patients, the experimenter showed the scale (see appendix) ranging from 0 to 100, with the sentence ”To what extent do you think “Elovix” has been effective to heal the patient?”. The participant was asked how effective he or she thought the medicine was in healing the patient and to respond by pointing out a number on the scale. Then the participant was done with the first game. At this point, participants were asked not to talk about the test before completing the whole study to avoid spreading of information among participants that could prove a threat to the internal validity of the experiment.

When all the participants were done with the first game, they participated in the intervention together. The intervention aimed to debias the participants and consisted of a lesson in basic scientific thinking, based on the intervention by Barberia et al. (2013). The lesson covered the scientific method, bias, correlation and causality. The participants were shown the different steps of the scientific method and why the logic of the method can have useful applications in everyday life. The consequences of bias on thinking and decisions were presented shortly, and connections were made to scientific thinking as a tool to work against biases. The difference between correlation and causation was discussed, as well as necessary steps for figuring out the occurrence of causation, such as refraining from presenting the cause (see appendix for the complete intervention).

After the intervention, all participants did the contingency judgment task individually again. But before the game began, they were asked to rate their understanding of the lesson on the scale ranging from 1 to 5. The participants were then told that their task was to find out if another new, fictitious medicine called “Batatrim” was effective in healing the patients that were sick in a different disease than before, called “Haokoman syndrome”. When the participants had given their response for 40 new patients they were asked to rate this new medicine on the same scale as before, from 0 to 100, but this time the scale was accompanied with the sentence “To what extent do you think “Batatrim” has been effective to heal the patient?”. After the second contingency judgment task was done, participants were debriefed and financially compensated for participating.

Ethics

All participants in the study were above the age of 15. Before deciding to participate in the study, all participants read and signed a letter of consent. The letter included information about how long the study would take, that all data would be kept anonymous and confidential, that the study did not cause any physical or psychological harm, that they were free to stop participating at any time and that all participants would be financially compensated for participating. All participants were given a number that was used during the whole study, to avoid having to record names. After the study, the participants were debriefed regarding the nature of the study and they also had an opportunity to ask questions at this time.

Results

In line with the first hypothesis, the participants overestimated the contingency in both the pre-test and the post-test (see Table 2 for descriptive statistics). A paired-samples t-test showed that the participants' estimated contingency (i.e., rated effectiveness of the medicine) in the pre-test ($M = 64.53$, $SD = 25.35$), differed significantly from the preprogrammed contingency of 0, t (29) = -13.95, $p > .001$. A paired-samples t-test showed that the participants' estimated contingency in the post-test ($M = 70.6$, $SD = 17.46$), differed significantly from the preprogrammed contingency of 0, t (29) = -22.15, $p > .001$ as well.

In contrast to our second hypothesis, the participants did not decrease their estimates of contingency in the post-test compared to the pre-test. In fact, a paired-samples t-test showed that there was an increase in rated effectiveness ($M = 6.06$), but this increase between pre-test ($M = 64.53$, $SD = 25.35$) and post-test ($M = 70.60$, $SD = 17.46$) was not statistically significant, t (29) = -1.78, $p = .09$.

A paired samples t-test showed that there was no significant difference between the actual contingencies and the preprogrammed zero contingency in the pre-test, t (29) = 1.25, $p = 0.22$, or in the post-test, t (29) = 1.64, $p = .11$. This indicates that the overestimations of contingencies were not due to actual contingencies experienced by participants. In addition, there was no significant difference between the actual contingency in the pre-test ($M = 5.57$, $SD = 24.35$) and the actual contingency in the post-test ($M = 9.07$ $SD = 30.35$), t (29) = -0.54, $p = .59$, either. Paired-samples t-tests were conducted to evaluate the difference between the actual contingency and the estimated contingency. There was a statistically significant difference between the actual contingency ($M = 5.57$, $SD = 24.35$) and estimated contingency ($M = 64.53$, $SD = 25.35$), t (29) = 9.70, $p < .001$, in the pre-test. The mean difference between actual contingency and estimated contingency was 58.97. There was a statistically significant difference between the actual contingency ($M = 9.07$ $SD = 30.35$) and estimated contingency ($M = 70.6$, $SD = 17.46$), t (29) = -10.09, $p < .001$, in the post-test as well. The mean difference between actual contingency and estimated contingency in the post-test was 61.53. These results further indicate that the overestimation of contingency was not due to the actual contingency.

There was some interindividual variability. In the pre-test, the minimum actual contingency was – 26, and the maximum actual contingency was 82. In the post- test, the minimum actual contingency was – 29, and the maximum actual contingency was 75. In the post-test, the lowest estimate of contingency that was reported by a participant was 20. This participant had estimated the contingency in the pre-test to 40.

Two multiple regression analyses were conducted to test the third and fourth hypotheses. The analyses revealed that the actual contingency could not predict the estimated contingency in the pre-test, $\beta = .002$, $t\ (29) = .01$ $p = .99$, nor in the post-test, $\beta = -.13$, $t\ (29) = -.67$ $p = .51$. These results support the third hypothesis. Regarding the fourth hypothesis, the analyses revealed that there was no significant effect of P(cause) on rated effectiveness in the pre-test, $\beta = .20$, $t\ (29) = .88$ $p = .39$. However, there was a significant effect of P(cause) on rated effectiveness in the post-test, $\beta = .53$, $t\ (29) = 2.82$ $p <.01$. Based on these results, P(cause) can predict the rated effectiveness as hypothesized, but this effect is seen in the post-test, not in the pre-test.

A paired samples t-test was conducted to evaluate the difference between P(cause) in the pre-test and P(cause) in the post-test. There was a significant increase in P(cause) in the post-test as compared to the pre-test, $t\ (29) = 2.06$, $p < .05$. These results show that the participants were more active (presented the cause more often) in the post-test, than in the pre-test.

After the intervention and before the post-test, each participant was asked how well they understood the lesson (intervention) on a scale from 1 to 5, 1 meaning “not at all” and 5 meaning “very well”. The mean rated understanding of the intervention was 3.93 (*SD* = 1.11).

A visual inspection of normal probability plots and scatter plots failed to show any major deviations from normality, linearity and homoscedasticity.

In sum, results from the study found support for the first and third hypotheses. The second hypothesis was not supported and the fourth hypothesis was partially supported. See figure 1 for results from the pre-test and the post-test summarized in a bar graph.

Table 2. *Descriptive statistics from pre-test and post-test. The actual contingency values have been re-scaled from a 0-1 scale to a 0-100 scale.*

	Pre-test		Post-test	
	M	SD	M	SD
Preprogrammed contingency	0	0	0	0
P(cause) in %	68.27	14.21	74.50	19.96
Actual contingency	5.57	24.35	9.07	30.35
Estimated contingency	64.53	25.35	70.6	17.46

Note:
Preprogrammed contingency = Preprogrammed contingency between cause and outcome.
P(cause) = Probability of the potential cause (giving medicine) in percentage.
Actual contingency = Contingency experienced by participants calculated by ΔP.
Estimated contingency = Rated effectiveness of the medicine provided by participants.

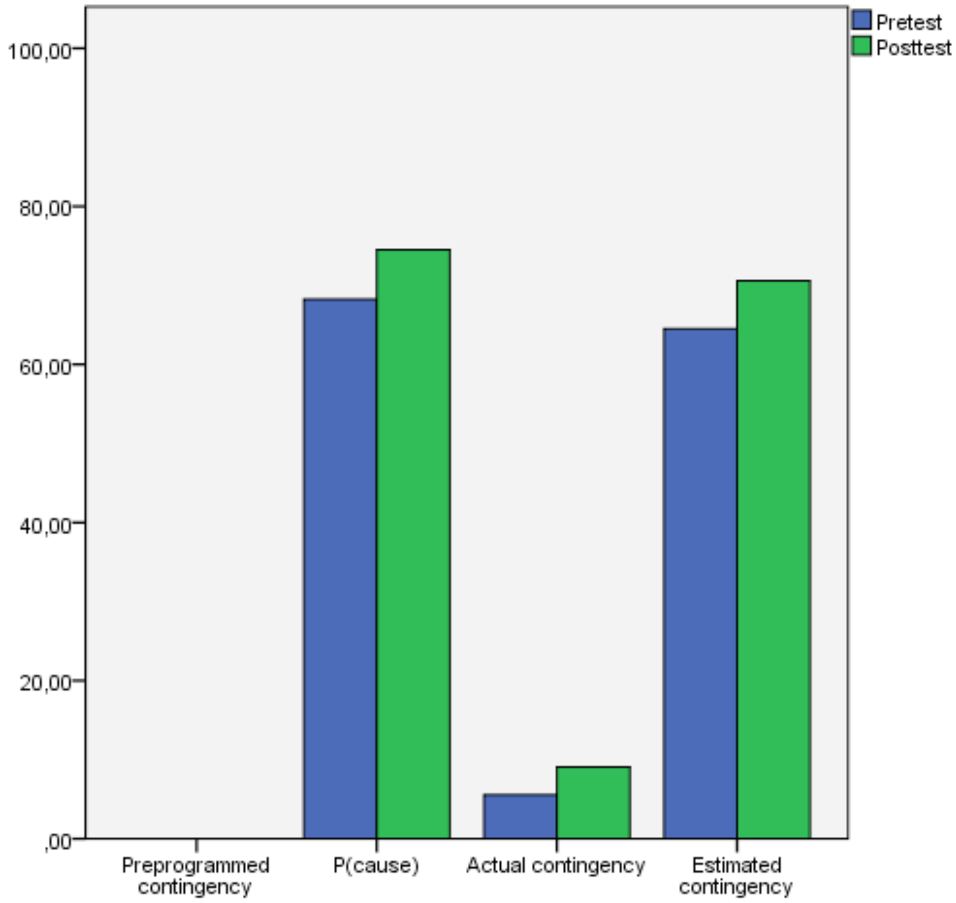


Figure 1. *Graph showing results from pre-test and post-test. The actual contingency values have been re-scaled from a 0-1 scale to a 0-100 scale.*

Discussion

The aim of this study was to examine if results from two previous studies on causal illusions could be generalized to rural Kenya. The study did replicate some results from the previous studies. However, the main hypothesis was not supported. The debias intervention that was based on the intervention by Barberia et al. (2013) did not succeed in debiasing the participants. Consequently, the study could not replicate the main finding of Barberia et al. (2013).

As hypothesized, the participants were biased in their estimates of contingency. They rated the medicine as effective in both the pre-test and the post-test, when in fact it was not effective at all. As previously mentioned, some studies show that people are sensitive to actual contingencies (e.g. Shanks & Dickinson, 1987), but our results replicate the findings of both Blanco et al. (2011) and Barberia et al. (2013), namely that people generally overestimate contingencies. Only one participant gave a correct estimate (0) of the contingency, and another participant gave an estimate of 10 in the pre-test. The lowest estimate of contingency in the post-test was 20. The fact that the first hypothesis was supported, and that the participants developed such strong causal illusions, supports the importance of more research on debiasing interventions.

It was ruled out that the actual contingency affected the participants to develop causal illusions. Firstly, the actual contingency did not differ significantly from the preprogrammed zero contingency. Secondly, there was a significant difference between the participants' actual contingencies and estimated contingencies in both pre-test and post-test. Thirdly, actual contingencies could not predict the participants' overestimation of contingencies. Consequently, the actual contingency cannot explain why the participants overestimated the contingency. These results replicate the findings of both Blanco et al. (2011) and Barberia et al. (2013) and further deviates from studies that report that participants are sensitive to actual contingencies (e.g. Shanks & Dickinson, 1987). There is one main difference between Shanks and Dickinson's (1987) study and our study that might explain the contrasting results. In Shanks and Dickinson's (1987) study the potential cause was pressing the spacebar on a computer and the outcome was a flashing triangle on the computer screen. In our study the potential cause was giving medicine to a patient and the outcome was the recovery of the patient. It is possible that it is easier for participants to infer causality from recognizable events, such as medicine and recovery from a disease, than from more abstract events like pressing a key and a flashing triangle. That is, people might expect a medicine to cause recovery in a patient to a greater extent than they expect the pressing of a key to cause a triangle to flash on a screen. Hamilton and Rose (1980) argue that people's previous experiences are likely to bias their expectations about the relation between two variables.

The fourth hypothesis, that the frequency of the participants' responses P(cause) would predict the participants' estimates of contingency, was based on results from Blanco et al. (2011). This hypothesis was only partially supported. The frequency of the participants' responses could predict their estimates of contingency in the post-test, but not in the pre-test. In the post-test, active participants, those that presented the cause more frequently, were more likely to develop a stronger overestimation of the contingency than those who presented the cause less often. That is, the more active participants developed stronger illusions of causality.

The main hypothesis, the second one, was not supported. Based on Barberia et al. (2013), we hypothesized that the participants would report more accurate estimates in the post-test than in the pre-test. However, these findings were not replicated. The participants maintained their illusions of causality. The results support the fact that our tendency to detect causal relationships even when they do not exist is very strong (Barberia et al., 2013). The results also indicate that the participants used a positive testing strategy, meaning that they presented the cause often and focused on the cases in which the medicine was presented (Klayman & Ha, 1987). Since the tests were high-outcome conditions, 31 respectively 30 patients were healed, it was likely that the cause and outcome would coincide often by chance. This means that the three variables most known to produce illusions of causality appeared in the task. The high outcome was predetermined, but the P(cause) effect appeared because the participants were active. The combination of the two increased the likelihood of cause-outcome coincidences.

That the participants became more active in the post-test seems to contradict the participants' high ratings of understanding. Problems connected to teaching scientific thinking give a potential explanation to why the participants did not decrease their responses. Barberia et al. (2013) called the contingency judgment task a transfer task since the participants needed to be able to transfer the new knowledge from the intervention to the task, and implement it. Barberia et al. (2013) did not discuss the issues regarding teaching scientific thinking, but their results showed that they successfully managed to teach the participants scientific thinking that was relevant to increase the accuracy in contingency judgments. Since our participants rated their understanding of the intervention as high, perhaps the problem was to transfer the knowledge into practical application in the task. In their intervention, Barberia et al. (2013) presented examples that had not only the same deep structure as in the contingency task, but also the same surface structure. For example, they used an example with an herb and recovery. Based on Willingham's (2008) points regarding the difficulties with teaching scientific thinking, this probably increased the likeliness that the knowledge of how to solve the problem transferred to the task. As mentioned, the participants in the study of Barberia et al. (2013) did not know that the game that they played was part of the study. Since our participants were aware that the game was part of the study, we didn't believe that our examples should be as similar to the medicine and recovery as they were in the study of Barberia et al. (2013). It was not desired that we simply told the participants how to do the task. In conclusion, it is possible that the participants were unable to transfer their knowledge to the post-test, and as a result causal illusions persisted in the post-test. However, this doesn't explain why the participants became more active in the post-test; the P(cause) value increased significantly from the pre-test to the post-test. According to Blanco et al. (2011), illusion of causality increases when the amount of trials increases. They examined the development of P(cause) along the trials and found that contrary to some predictions, P(cause) increased successively. Since P(cause) could predict the estimated contingency, these results indicated that causal illusions increased along with the trials. We were unable to assess if this occurred, since we only replicated the effect of P(cause) on estimates of contingency in the post-test. However, the non-significant increase in estimates of contingency between the tests approximated significance, and the P(cause) increased in the post-test. In addition, both participants that gave low estimates of contingency in the pre-test gave higher estimates of contingency in the post-test. These results taken together are in direct contrast to what the intervention aimed to achieve, and more in line with what you would expect if illusions increased along with the trials. We were not able to examine it, but if causal illusions do increase along with the trials in a repeated measures design, effects of the intervention might not be discovered because the participants are likely to rate the contingency higher in the post-test than in the pre-test. This might then be a confounder, and prevent the discovery of an effective intervention. However, since we were unable to examine this, and since it has not been examined in a repeated measures design or with an intervention aimed to decrease illusions, it is not possible to draw any conclusions at this point.

Method discussion

To the best of our knowledge, this is the first study to test a debias intervention aimed to reduce causal illusion in an experiment with a repeated measures design. Barberia et al. (2013) argued in their study that future studies similar to theirs should examine the effectiveness of the intervention in a repeated measures design. Ideally, the repeated measures design should also include a control group with no intervention.

Barberia et al. (2013) included deception in their intervention. The participants were deceived into believing that a bogus product was effective, and they did the contingency judgment task without knowing that it was a part of the experiment. Barberia et al. (2013) chose to include this to clearly show the participants the usefulness of accurate judgment of causality, as well as to avoid that the participants due to the bias blind spot (i.e. believing that bias doesn't affect oneself) failed to realize that they were in need of debiasing. Although those reasons seem valid, deception was not included in the intervention of this study. We believe that for several of the participants, this was their first experience of participating in research. Including deception could potentially make the participants feel like they were tricked, or that our expectations on them were low. After all, the majority of the participants in the study of Barberia et al. (2013) reported that they felt cheated after the intervention. The cost versus benefit of using deception in psychological science is a complex issue. Some researchers mean that the use of deception can decrease faith in science, and that it increases suspicion towards science (e.g. Baumrind, 1985). This risk, in combination with the specifics of the context, resulted in an intervention free from deception. Barberia et al. (2013) believed that the deception was an important part of the intervention. At the same time, they did discuss that they were unsure of which part of the intervention that was vital for the success. Perhaps including deception is an important part of debiasing, if not only to make participants aware of the fact that they are prone to biases.

Due to human error, the outcome-present trials were 31 (77.5%) in the pre-test. This number should have been 30 (75%), as it was in the post-test. However, we believe that this had minimal, if any, effect on the results. Firstly, the difference in percentages had no impact on the preprogrammed zero contingency since the outcome was independent of the participants responses, regardless of the percentage of outcome-present trials. Secondly, values of P(cause) and actual contingencies are based on the participants' responses. It is hard to speculate about how and if participants' responses would be affected by an extra outcome-present trial in the pre-test. Blanco et al. (2011) compares blocks of 10 trials each to the whole sequence of 40 trials on P(cause) values. In each of these blocks the percentage of outcome-present trials is 80% as compared to 75% in the whole sequence of 40 trials. The authors argue that these percentages are very similar and warrant a fair comparison. Based on this we believe that the difference is too small to have any impact on our results.

One possibility as to why our main hypothesis was not supported could be that the participants did not understand the information that we tried to communicate in the intervention. This could have been due to language barriers. The experimenters and the participants spoke different native languages, and spoke English with different accents. Participants' ratings of understanding suggest that the participants rated their understanding of the intervention as high. There is, however, a possibility that the participants' ratings of understanding did not reflect their true understanding of the intervention. This could for example be due to a desire to please the experimenters and therefore not wanting to give the intervention a low rating as it could be interpreted as a judgment of the experimenters' skills. The experimenters asked each participant to rate their understanding of the intervention before starting the post-test. A suggestion for future research is to let participants rate their understanding of the intervention without the experimenters being present. This could avoid the risk that the ratings do not reflect participants' true understanding due to a desire to please experimenters. The potential problem of language barriers could also have resulted in difficulties in understanding the contingency judgment tasks. If this was the case, then estimates of contingency could have been arbitrary and not really based on information about contingency deducted from the trials.

Aside from potential difficulties in communication, there is a potential risk that the design of the intervention was flawed. If there was a design flaw in the intervention, then it could not have debiased the participants even when controlling for all other factors. Barberia et al.

(2013) did not report their intervention in detail. So, our intervention was based on the concept and the general themes of their intervention. The intervention used in this study can be found in detail in the appendix.

The contingency judgment task in this study differed from previous studies because the task was conducted with papers and computers, while the previous ones were conducted with only computers. A problem that could have resulted from this is that the experimenters affected the participants' responses through cues, for example by unintentionally rewarding the participants when they used a good strategy. We were aware of this risk, and did our best to make sure that this didn't affect the test. In addition, for the test to be valid it was important that the participants believed that the computer calculated the response that we gave them (healed/not healed). We believe that this crucial part was upheld during the tests, since no participant seemed to realize that the order in which the outcome appeared was preprogrammed. The results, the persisting overestimations, supports that this was not a problem for the study.

Regarding the reliability of the study, the contingency judgment task is standard in judgment and decision-making research according to Barberia et al. (2013). Based on this, we found the reliability of the contingency judgment task to be high.

Conclusion and future studies

The results of this study further strengthen what several studies have shown, that people are prone to develop causal illusions. The results also support what is discussed in current research, that debiasing is a difficult task that faces many obstacles. That the participants became more active was a surprising finding, and it is uncertain what caused this.

The practical implications of the current study are in line with those reported by Blanco et al. (2011). The example that was used in both studies, consisting of medicine as cause and recovery as outcome, is similar to real life situations. Based on how people act in the contingency judgment task, it seems likely that they easily could develop the belief that a pseudomedicine is effective. If it has no or few side effects people will administer the medicine often. Even if a disease tends to disappear on its own, people will not expose themselves to cases that indicate that the medicine doesn't cause recovery. Consequently, people might spend money on useless pseudomedicines, and some illnesses might become severe if they are not properly treated.

Further research is needed to establish external validity of the debias intervention by Barberia et al. (2013). In other words, research to examine if the educational intervention is a globally effective strategy to debias people from illusions of causality. For future studies we believe that it is important to consider that illusions may increase when trials are increased, especially for studies with a repeated measures design.

References

Allan, L. G. (1993). Human contingency judgments: Rule based or associative?. *Psychological Bulletin*, *114*(3), 435-448. doi:10.1037/0033-2909.114.3.435

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin*, *110*(3), 486-498. doi:10.1037/0033-2909.110.3.486

Arnott, D. (2006). Cognitive biases and decision support systems development: A design science approach. *Information Systems Journal*, *16*(1), 55-78. doi:10.1111/j.1365-2575.2006.00208.x

Barberia, I., Blanco, F., Cubillas, C. P., & Matute, H. (2013). Implementation and assessment of an intervention to debias adolescents against causal illusions. *Plos ONE*, *8*(8), 1-9. doi:10.1371/journal.pone.0071303

Baumrind, D. (1985). Research Using Intentional Deception: Ethical Issues Revisited. *American Psychologist, 40*(2), 165-174.

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. doi:10.1080/17470218.2011.552727

Buehner, M. J. (2005). Contiguity and covariation in human causal inference. *Learning & Behavior*, *33*(2), 230-238. doi:10.3758/BF03196065

Croskerry, P., Singhal, G., & Mamede, S. (2013). Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Quality & Safety*, *22 Suppl. 2*ii58-ii64. doi:10.1136/bmjqs-2012-001712

Crump, M. C., Hannah, S. D., Allan, L. G., & Hord, L. K. (2007). Contingency judgements on the fly. *Quarterly Journal Of Experimental Psychology*, *60*(6), 753-761. doi:10.1080/17470210701257685

De Houwer, J., & Beckers, T. (2002). A review of recent developments in research and theories on human contingency learning. *Quarterly Journal of Experimental Psychology: Section B*, *55*(4), 289-310. doi:10.1080/02724990244000034

Evans, J. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59255-278. doi:10.1146/annurev.psych.59.103006.093629

Greville, W., & Buehner, M. (2010). Temporal predictability facilitates causal learning. *Journal of Experimental Psychology-General*, *139*(4), 756-771.

Hamilton, D. L., & Rose, T. L. (1980). Illusory correlation and the maintenance of stereotypic beliefs. *Journal of Personality & Social Psychology*, *39*(5), 832-845

Haselton, M. G., Nettle, D., & Andrews, P. W. (2015). The evolution of cognitive bias. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (p. 724-746). John Wiley & Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/9780470939376.ch25

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs*, *79*1:1-17.

Kahneman, D. (2012). Tänka, snabbt och långsamt. Stockholm: Volante, 2012 (Tyskland).

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*(2), 211-228.

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(3), 451-460. doi:10.1037/0278-7393.32.3.45

Lilienfeld, S. O., Ammirati, R., & David, M. (2012). Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology*, *50*(1), 7-36.

Lilienfeld, S. O., Ammirati, R., & Landfield, K. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare?. *Perspectives on Psychological Science*, (4). 390.

Matute, H. (1996). Illusion of control: Detecting response-outcome independence in analytic but not in naturalistic conditions. *Psychological Science*, (5). 289.

Matute, H., Yarritu, I., & Vadillo, M. A. (2011). Illusions of causality at the heart of pseudoscience. *British Journal of Psychology*, *102*(3), 392-405. doi:10.1348/000712610X532210

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, (2015)*, doi:10.3389/fpsyg.2015.00888/full

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality & Social Psychology Bulletin*, *28*(3), 369.

Rescorla, R. A. (1968). Probability of shock in the presence and absence of cs in fear conditioning. *Journal of Comparative And Physiological Psychology*, *66*(1), 1-5. doi:10.1037/h0025984

Stanovich, K. E. (2010). *Rationality and the reflective mind*. Lanham, MD : Oxford University Press.

Shanks, D. R., & Dickinson, A. (1987). Associative accounts of causality judgment. In GH Bower (Ed.), *The psychology of learning and motivation* (pp. 229–261). San Diego, CA.: Academic Press.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, (4157). 1124.

Trimmer, P. C. (2016). Optimistic and realistic perspectives on cognitive biases. *Current Opinion in Behavioral Sciences*, *12*(Behavioral ecology), 37-43. doi:10.1016/j.cobeha.2016.09.004

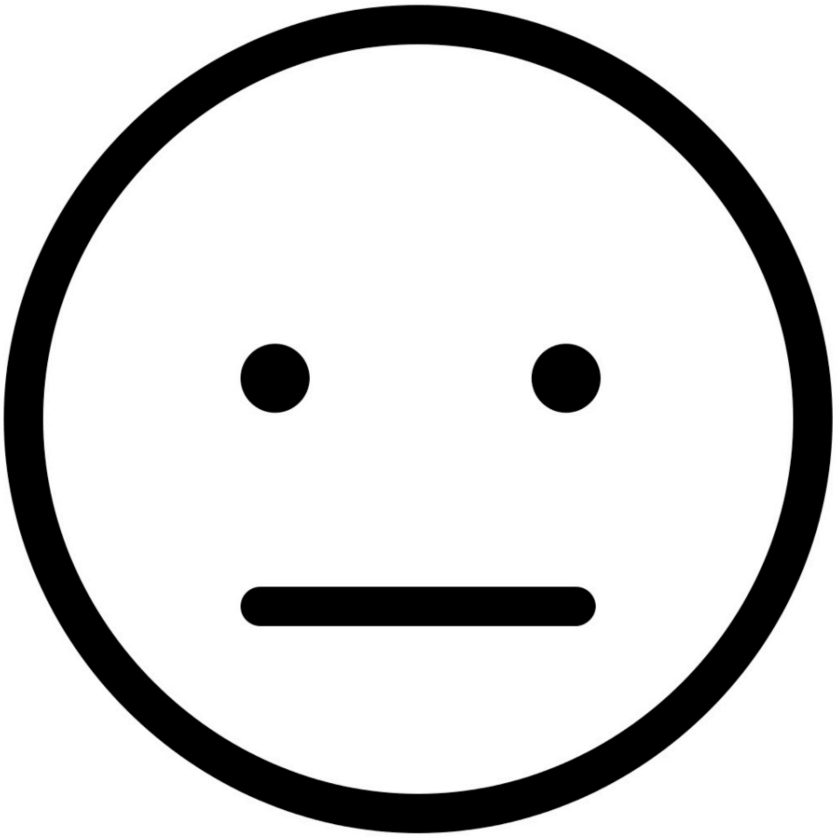
Willingham, D. T. (2008). Critical thinking: Why is it so hard to teach?. *Arts Education Policy Review*, *109*(4), 21-32.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin*, *116*(1), 117.

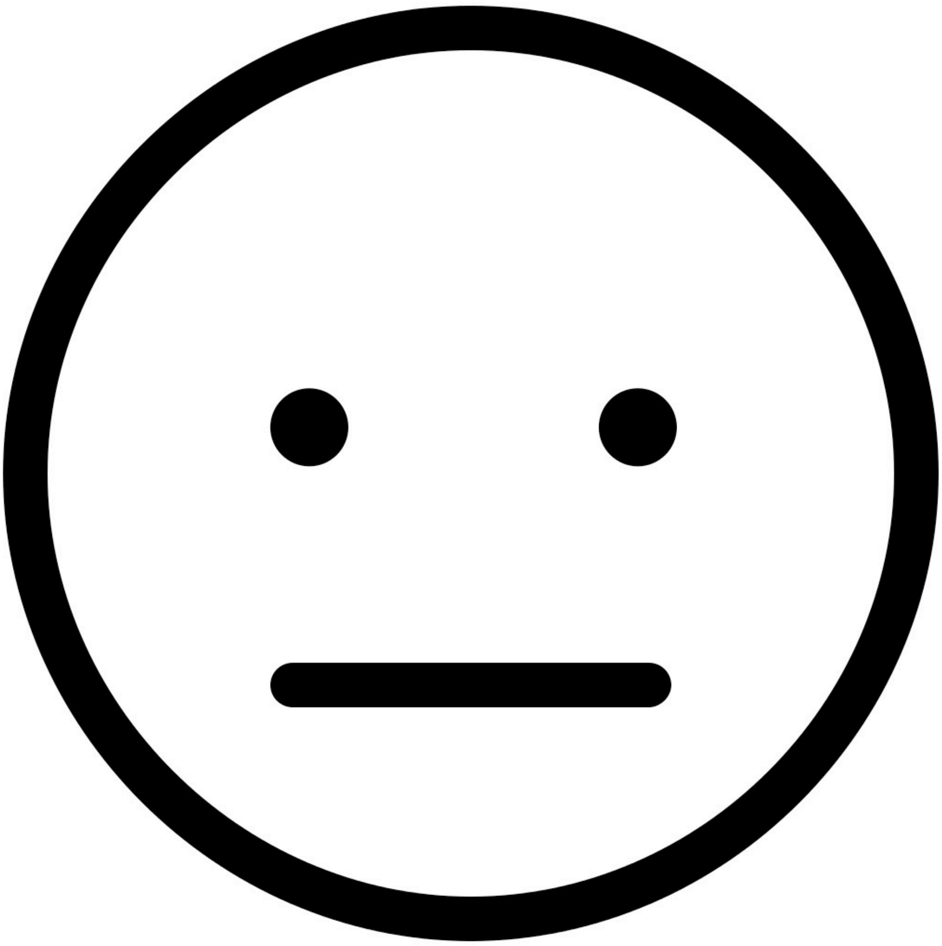
Yarritu, I., Matute, H., & Vadillo, M. (2014). Illusion of control: The role of personal involvement. *Experimental Psychology*, *61*(1), 38-47.

Appendix

Patient 1

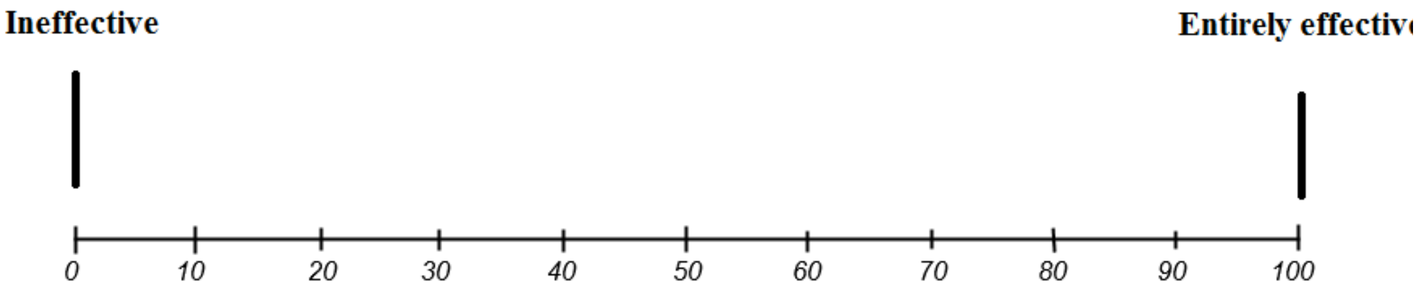


The patient has not been healed.

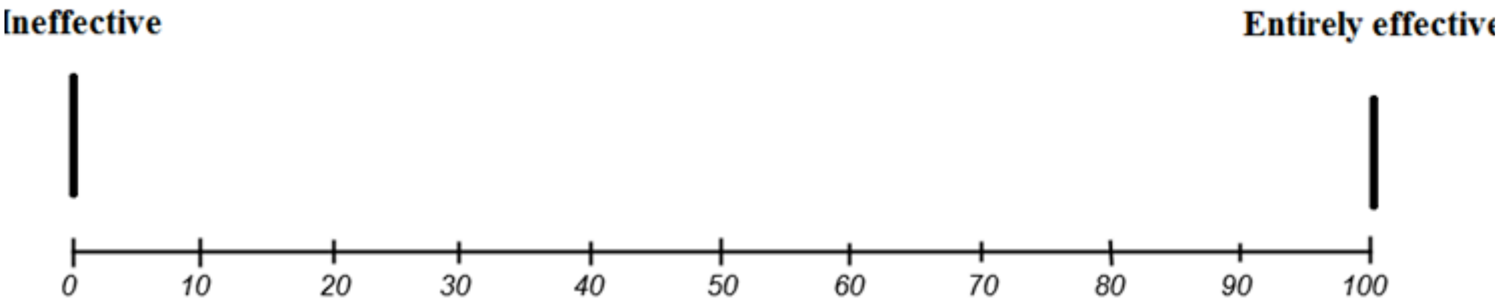


The patient is healed!

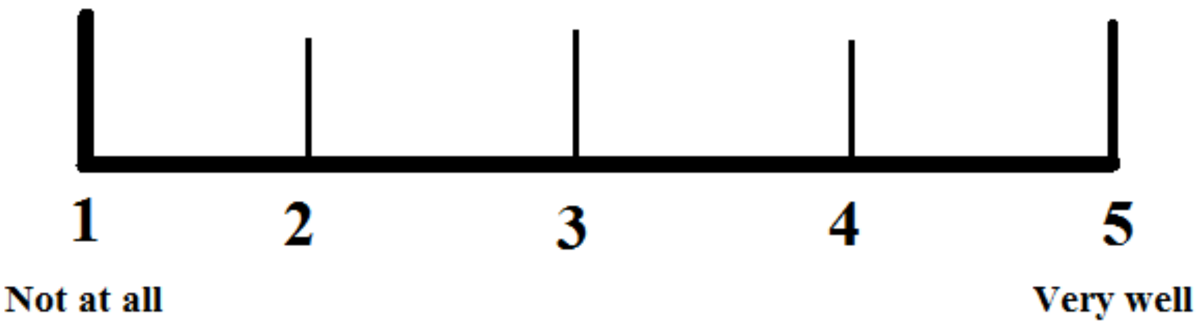
To what extent do you think Elovix has been effective to heal the patients?



To what extent do you think Batatrim has been effective to heal the patients?



How well did you understand the lesson?



Letter of consent

Hello, we are performing a study as part of our bachelor thesis in psychology for Lund University, Sweden. We are studying how people think about events. This study consists of a simple test, a lecture and then another simple test. The tests will take approximately 10 minutes each and the lecture will take an hour. The participants will receive a 400 KES compensation. We would like to ask you to participate in our study.

If you agree to take part in this study, you should know that:

- All data that we obtain will be kept confidential and anonymous.
- You may stop participating in this study at any time.
- If you wish you shall receive information about the results of the study after we have obtained results.
- All data that we obtain will be analyzed as grouped data, not as individual data.
- The study will not cause any physical or psychological harm.

I, ______________________________________, understand the nature of this study and I agree to participate voluntarily. I give the researchers permission to use my data as part of their study.

Signature: ____

Date: _________________

If you have any questions or wish to receive the results of the study, please contact any of the following:

Supervisor - Anna Kemdal Pho - anna.kemdal_pho@psy.lu.se

Linnea Cederlund - nea.cederlund@gmail.com
Henrik Josander - henrik.josander@gmail.com

Instructions

In a little bit you will play a game individually, with one of us. We will now explain how the game works.

When we begin the game you will imagine that you are a medical doctor. You should try to find out if a new, imaginary medicine, that we call Elovix, can heal people that are sick in a made-up disease that we call “MacGregor’s Syndrome”.

Forty people are sick, and you can choose to give or not to give medicine to each one of them. We will show you one patient at a time, until we have gone through all 40 patients.

For each patient you will tell me if you want to give or if you don’t want to give the medicine.
After you have done this, medicine or not medicine, we will show you with these papers if the patient was healed or if the patient is still sick.

Remember, you should try to find out if the medicine works or not, if it heals people or if it doesn’t.
When we have gone through all the patients we will ask you how good you thought the medicine was at healing patients.
During the test we will write down the answers in a computer.

Do you have any questions?

We will give you a participant number that you need to keep with you until you go home. We use this instead of recording your name, so that all your information will be kept anonymous, and we will need you to remind us of your number in the second game later today.

Let’s start the game.

Intervention

Does everyone know what a scientist does?

Basically, a scientist's job is to find answers to questions. Scientists study pretty much everything in the world, and everyday they improve our knowledge about how things work and how things look in the world.

What we really are interested in today is something called the scientific method. It is the method that scientists all over the world use to get answers to questions, to get new knowledge. Scientists get the answers through observations and experiments, so called research, and to make sure that they get correct answers when they do this they use the scientific method. Today we will show you how you can think like a scientist, and why this can be very useful.

Let's look at the scientific method in more detail by using an example.

The scientific method begins with a question. This question must be something that we can measure in some way. So, as an example, let's say our question is: Can tea make you perform better in school?

Then we make a guess about what the answer is going to be, the guess that we make is called a hypothesis. Our hypothesis in this example is: We believe that drinking tea will make you get better grades.

Then we try to answer the question. This we will do by making an experiment, which is the best way to get a precise answer to a question. This has to do with the most important thing in an experiment, something that is called control. Control in an experiment means that you make sure to set up the experiment in a way that makes you certain that you are testing what you want to test, that nothing else will have an influence on the answer that you get. For example, a very common thing to do in an experiment to have control is to have two groups, one group called the experimental group and the other called the control group. In our example, the experimental group is the group that drinks tea before they take an exam. The control takes the same exam, but they don't get to drink tea before taking the exam. If our guess, our hypothesis, is right, then the participants in the experimental group will get better grades than the participants in the control group.

When we have done our experiment we will have a result, an answer to our question. Does it look like we guessed it would look? Don't worry if it doesn't. It is common in research that the results don't look like you guessed that they would. And sometimes we learn even more from finding out that something doesn't look like we thought it would do.
If we did the experiment with the tea, I am quite sure that we wouldn't see any difference. As far as I know, tea doesn't make people perform better. So, if we would have done this experiment for real, our guess would have been wrong. But, we would have learnt something new. If we want to improve our grades, we shouldn't rely on tea!

We have to make sure that we document everything we do carefully, both when we plan and when we do the experiment. The reason to this is that other scientist are going to want to make sure that they can make the experiment just like the one we made, and that they get the same result. Because science is tough, the results that we get must stand several tests. So scientists make sure to double-check each other's results. If the same experiment is made several times, and the results differ from each other between the experiments, we need to think of how to make

a new, different experiment that might give a better answer to our question. And a better answer is an answer that we get in all the tests that we make. All this has to do with something called skepticism.
Scientists are skeptical; to believe something they need to see it several times, the same result must be repeated several times. They are careful about what they believe.

Finally, we share our results with everyone. Because sharing what you come up with is an important part of science. You share the things that went well in your experiment, as well as the things that didn't go as planned and the errors that you might have made. Honesty is a basic part of the scientific method, as well as one of the most important parts.

Scientists are always aware that the result they get is the best answer to the question so far, but it doesn't mean that we in the future won't come up with another experiment that gives a better, clearer answer to the question. This is why a scientist will say:
This is how it seems to work, not this is how it works.

Also, scientists are supposed to be as objective as possible. This means that when they are looking for answers to questions, it shouldn't matter what they personally might wish the answer to the question to look like. They should always look for the truth.

Why is this method so important?
To answer that, we need to look a little bit at how we usually think and make decisions. Most of the time our thought processes are fast and intuitive. Basically, we are unaware of how we think, and we act on the basis of "what feels right". This works quite well in many cases, but often this kind of thinking results in errors. We call them errors because they make us believe things that can be far off from what is actually going on in reality.

For example, one error that all of us often do is that we never really question what we believe. As a matter of fact, we tend to look for things that support what we already believe, and ignore information that shows us that we might be wrong. And, like I said, this is something we do without being aware of it. So, because of this error often we believe that we are right about something, when in fact we are wrong.

This is why the scientific method is so important. Compared to everyday thinking that is based on "what feels right", it is based on information that we get from observations and experiments. It tells us to look for information that may prove that we are wrong, instead of just looking for information that proves that we are right.
The method helps us to learn about things and act in a way that we can't if we just think like we do everyday. It is one of the best methods that we have to really figure things out, and it is used every day and all over the world.

Also, us humans tend to believe things, instead of disbelieving them.
This is why the scientific skepticism is so important, it reminds us that we need to question things before we believe them, we need to test things properly before we say that they are right or efficient.

So, what can we learn from the scientific method?
The scientific method can help us get answers that are useful for us, because they will more often be correct than the ones that we get from our everyday way of thinking.

And we make so many important decisions in our life, and who would want these decisions to be based on errors? You don't have to do experiments to make better decisions, knowing what the method is based on can help you make more well-informed decisions in your life.
So, it teaches us to look for all the information, look for information that can show that you are wrong, to question the information that we do get not accepting without thinking it through and to be objective, that is, open to the possibility that the answer that we get might not be the one that we wanted or expected.

- Correlation and Causation

Now we are going to talk about the relationship between different events. I am going to explain two concepts regarding to this. Namely, correlation and causation.

- What is correlation? - The word Correlation is made of Co- (meaning "together"), and Relation. So, correlation is when two things have a relation together, but that does not have to mean that one thing causes the other thing to happen. It just means that they have a relation.

There can be both negative and positive correlations. I will explain this with some examples. An example could be that your stomach hurts every time your friends from another village come to visit. When they visit, your stomach hurts, when they don't visit, your stomach doesn't hurt. So there is a relationship or a correlation between your stomachache and visits from your friends. In this case there is a positive correlation because the more visits from your friends, the more you have stomachaches.
Does this mean that your stomach hurts because of your friends?

Probably not. Maybe your friends always gives you some different food when they come to visit and the real reason that your stomach hurts is not because your friends visited but because you ate the food that you are not used to. In this example there is a correlation between visits from your friends and stomachache, but no causation. The real causation is between the food and stomachache. So correlation means that two things have a relationship but one thing doesn't necessarily cause the other. Causation means that one thing happens because of the other thing. Do you see the difference between correlation and causality?

Another example could be amount of time you sleep and how tired you are the next day. If you sleep many hours you will be less tired the next day. If you sleep few hours you will be more tired. In this case there is a negative correlation between hours slept and tiredness. The more you sleep, the less tired you are. And the less sleep you get, the more tired you will be. So you can see that the two things (sleep and tiredness) have a relationship.

Do you think there is causality or only correlation between hours slept and how tired you are the next day? It sounds probable that you are tired because you didn’t sleep enough, but you can’t be sure about it. Maybe there is something else that affects both how tired you are and how much you sleep. So what can you do to be more sure about it?

How to infer causality

A scientist wouldn’t say that less sleep causes you to be tired just because it sounds likely to be true. So how would a scientist do to actually say that there is causality between two things? They use the scientific method, like we talked about earlier. A scientist would follow the steps in the scientific method to get an answer that is more reliable than just assuming based on your own beliefs and experiences.

If a scientist wants to see if there actually is causality between two events, it is important to see if the two events can happen independent of each other. The main point of this is to see if you can reject and rule out that one event causes another. If you have observed that one event usually happens after the next, you might think that one event causes the other to happen. Like sleep and tiredness. In other words, you might think that too little sleep is the cause of tiredness. But if you can choose to sleep a lot and you are tired anyway, and especially if this happens a lot of times, then you might have to change your mind about the causal relationship between sleep and tiredness. Then you might find out that there is only correlation, and no actual causation between amount of sleep and tiredness. In the example with visits from friends and stomach ache that we talked about earlier, you might want to try to eat something else while your friends are visiting, or you might want to try to eat their food while your friends are away to see if it is actually your friends or the food that causes your stomach to hurt. This would be more in line with how a scientist would think to try to find out what is causing your stomach to hurt.

Let’s look at one more example. Imagine that you are a teacher in a school. You want to see if a new special book will help your students get better grades. So you let students sign their name on a list to get to read the new book. Then you compare the grades of the students who read the book with students who did not read the book.

It turns out that the students who read the new book got much better grades than the students who did not read the book.
Is this proof that the students got better grades because they read the special book? Or in other words, is there causation between reading the new book and getting better grades? Remember that in this example, we let the students sign up themselves to read the book. So it might be that the students who signed up already were more interested in the subject and therefore would have gotten better grades anyway than students who did not sign up to read the book.

This is again an example of a correlation between better grades and reading the new book, but it does not have to mean that reading the book causes students to get better grades. Correlation does not have to mean causation.

Why is this important?
In the example with the teacher and the new schoolbook it is very important that the teacher can see the difference between correlation and causation. If he or she thinks that the new book actually causes students to get better graders he/she might spend a lot of money on new books for all the students when in reality the books didn't have anything to do with the better grades.

Now we have talked about a few examples of when there is correlation between two things but not causality. When people mistakenly think that one thing causes another but there in fact is only a correlation, we call it "illusion of causality" and it happens all the time. Even highly educated scientists make this mistake sometimes but it is very important to know about it and to consider it before making any big decisions.