Learning and Motivation 44 (2013) 303–311

Contents lists available at ScienceDirect

Learning and Motivation

journal homepage: www.elsevier.com/locate/l&m

# Evidence for an illusion of causality when using the Implicit Association Test to measure learning

Miguel A. Vadillo[a,*], Jan De Houwer[b], Maarten De Schryver[b], Nerea Ortega-Castro[c], Helena Matute[c]

[a] University College London, London, UK
[b] Universiteit Gent, Ghent, Belgium
[c] Universidad de Deusto, Bilbao, Spain

ARTICLE INFO

*Article history:*
Received 4 January 2013
Received in revised form 23 March 2013
Available online 28 April 2013

*Keywords:*
Outcome-density effect
Contingency learning
Causal learning
Implicit Association Test

ABSTRACT

Our ability to detect causal relations and patterns of covariation is easily biased by a number of well-known factors. For example, people tend to overestimate the strength of the relation between a cue and an outcome if the outcome tends to occur very frequently. During the last years, several accounts have attempted to explain the outcome-density bias. On the one hand, dual-process performance accounts propose that biases are not due to the way associations are encoded, but to the higher-order cognitive processes involved in the retrieval and use of this information. In other words, the outcome-density bias is seen as a performance effect, not a learning effect. From this point of view, it is predicted that the outcome-density bias should be absent in any testing procedure that reduces the motivation or opportunity to engage in higher-order cognitive processes. Contrary to this prediction, but consistent with the most common single-process learning accounts, our results show that the outcome-density effect can be detected when the Implicit Association Test is used to measure the strength of cue–outcome associations.


One of the most remarkable features of human beings and other animals is our outstanding ability to adapt to the regularities in our environment. Quite surprisingly, however, our accurate sensitivity to statistical relations does not make us immune to blatant cognitive illusions, some of them with far-reaching consequences. For instance, many citizens in developed societies still recur to homeopathy and other complementary or alternative medicines (Barnes, Bloom, & Nahin, 2008), despite the fact that they are known to be ineffective (Shang et al., 2005) and also despite their huge economic costs (Nahin, Barnes, Stussman, & Bloom, 2009). The impact of pseudoscience and erroneous causal beliefs in the educative system is equally astonishing (Lilienfeld, Ammirati, & David, 2012).

During the last decades, cognitive psychologists have explored how erroneous beliefs arise as the result of confirmation biases, illusory correlations, overreliance on heuristics, and, most importantly, illusory perceptions of causality (Gilovich, 1991; Vyse, 1997). Current research on associative learning has contributed to our understanding of causal illusions by identifying factors that bias our ability to detect the covariation between a candidate cause and an effect. One of these factors is the probability with which the to-be-explained effect occurs. Imagine that you suffer very frequently from headaches and that you are looking for a remedy to ameliorate your condition. Even without any treatment, headaches tend to disappear very frequently in the interval of a few hours. This warranties that, whatever remedy you decide to take when you feel a headache, its consumption is very likely to be followed by a recovery, even if the remedy itself is absolutely non-effective. However, if

* Corresponding author at: Division of Psychology and Language Sciences, University College London, 26 Bedford Way, London WC1H 0AH, UK.
*E-mail address:* m.vadillo@ucl.ac.uk (M.A. Vadillo).

0023-9690/$ – see front matter 
http://dx.doi.org/10.1016/j.lmot.2013.04.001

this experience happens regularly, it is very tempting, almost unavoidable, to conclude that the remedy must be effective, because it has been followed by the recovery so many times in the past. Taking this into account, it is hardly surprising that until the development of placebo-controlled, double blind tests, the history of medicine has been "the history of the placebo effect" (Shapiro & Shapiro, 1997). This example illustrates why the overall frequency of an effect can bias the perception of causality. By mere chance, frequent effects will usually happen after other factors, which will be seen as potential causes.

This effect has been studied extensively in the area of human contingency learning. In these experiments, participants are exposed to a series of trials in which a cue might be present or absent and an outcome might follow or not. Their task is to learn to predict the outcome based on the presence or absence of the cue in every trial. For example, a typical cover story used in these experiments invites participants to imagine that they are medical doctors who have to discover whether a patient is allergic to a given food. In each trial, they see whether the patient has taken that food and next, whether he or she suffered an allergic reaction. At the end of the experiment, participants are usually asked to rate the strength of the statistical or causal relationship between the candidate cause (the patient eating the food) and the outcome (her suffering an allergic reaction). Many experiments conducted with this or related tasks have found that even in situations in which there is no statistical relationship between the cue and the outcome, the participants still report that such a relationship exists if the outcome occurred in many trials (Allan & Jenkins, 1983; Allan, Siegel, & Tangen, 2005; Blanco, Matute, & Vadillo, in press; Musca, Vadillo, Blanco, & Matute, 2010). This outcome-density effect also takes place in situations in which participants are not judging the relationship between a neutral cue and an outcome, but the relationship between their own behavior and a consequence (Alloy & Abramson, 1979; Matute, 1995; Msetfi, Murphy, Simpson, & Kornbrot, 2005; Shanks, 1985, 1987).

Early explanations for the outcome-density bias were framed in associative terms (López, Cobos, Caño, & Shanks, 1998; Matute, 1996; Shanks, 1995). For a situation in which there is just one cue and one outcome, the standard associative explanation assumes that a node representing the cue and a node representing the context compete to become associated with the representation of the outcome. The strengths of the cue–outcome and the context-outcome associations are updated on a trial-by-trial basis by means of a simple error correction rule, such as the one proposed by Rescorla and Wagner (1972) in the area of classical conditioning. According to this learning rule, the cue and the context compete to become associated with the outcome, so that by the end of training their respective associations with the outcome will be proportional to their relative predictive validity. When adopting a competitive learning rule, in situations in which there is no statistical relation between the cue and the outcome, the context ends up accruing all the associative strength. However, it is often assumed that the salience of the context is lower than the salience of the cue. This implies that learning is slower for the context than for the cue or, in other words, that the context is a poor competitor during the first stages of learning. Therefore, early in training, the accidental pairings of the cue and the outcome can result in a spurious cue–outcome association that will only disappear afterwards, once the context has accrued enough associative strength. In sum, within association formation models of learning, the outcome-density effect is understood as a transient, preasymptotic bias arising from accidental cue–outcome pairings before the context becomes an effective competitor.

More recently, however, a dual-process model has been offered to account for the outcome-density effect. Allan et al. (2005) conducted a contingency learning experiment in which they manipulated the cue–outcome contingency and, orthogonally, the overall probability of the outcome. As in most contingency learning experiments, in each training trial participants first saw whether the cue was present and had to predict whether the outcome would follow using a yes/no discrete response. After their response, they were told whether the outcome really appeared afterwards and proceeded to the next trial. At the end of training participants rated the strength of the cue–outcome relationship. The judgments collected at the end of training showed the expected outcome-density effect: for any given level of cue–outcome contingency, participants' judgments varied as a function of the probability of the outcome. Most interestingly, however, this effect was absent in a dependent measure computed from the yes/no responses that participants gave during training. Specifically, Allan et al. compared the proportion of "yes" responses in cue-present trials and the proportion of "yes" responses in cue-absent trials. If participants think that there is a positive relationship between the cue and the outcome, they should predict the outcome more often when the cue is present than when the cue is absent. However, Allan et al. found that this measure was not sensitive to the outcome-density effect. From their point of view, the fact that the predictive responses were unaffected shows that the outcome-density effect does not influence how people encode the relationship between the cue and the outcome. However, participants' judgments are not based solely on the encoded cue–outcome relation. Additional processes involved in the production of the judgment would be responsible for the outcome-density effect (see also Allan, Siegel, & Hannah, 2007). In other words, the outcome-density effect is assumed to be a performance phenomenon, not a learning phenomenon. Quite interestingly, Perales, Catena, Shanks, and González (2005) have proposed a similar explanation for a related bias that happens when it is the cue, instead of the outcome, that occurs very frequently, namely the cue-density effect (see also Matute, Yarritu, & Vadillo, 2011; Vadillo, Musca, Blanco, & Matute, 2011).

Although the theoretical framework proposed by Allan et al. (2005) and Perales et al. (2005) is certainly inspiring and sheds new light on these effects, it is based on limited evidence. Dissociations between dependent variables can sometimes reflect the action of independent cognitive systems. But they can also be the product of a methodological artifact like, for instance, the limited reliability of one of the dependent measures (Shanks & St. John, 1994). In the case of the experiments conducted by Allan et al. and Perales et al., doubts can be raised about whether the dependent measures collected from trial-by-trial predictions are as reliable as the judgments provided at the end of training. For one thing, the outcome predictions were collected during training, that is, before the cue–outcome association had been properly encoded. Moreover, even if participants believe that there is a probabilistic relationship between the cue and the outcome, their predictive responses