Consciousness and Cognition 63 (2018) 11–28

Contents lists available at ScienceDirect

Consciousness and Cognition

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

Full Length Article

Illusory correlations despite equated category frequencies: A test of the information loss account

Michael Weigl[a,*], Axel Mecklinger[a], Timm Rosburg[a,b]

[a] Department of Psychology, Experimental Neuropsychology Unit, Saarland University, D-66041 Saarbrücken, Germany
[b] Department of Forensic Psychiatry, University Psychiatric Clinics Basel, Wilhelm Klein-Str. 27, CH-4002 Basel, Switzerland

ARTICLE INFO

Keywords:
Illusory correlation
Distinctiveness
Memory
Stereotyping
Availability

ABSTRACT

Illusory correlations (IC) are the perception of covariation, where none exists. For example, people associate majorities with frequent behavior and minorities with infrequent behavior even in the absence of such an association. According to the information loss account, ICs result from greater fading of infrequent group-behavior combinations in memory. We conducted computer simulations based on this account which showed that ICs are expected under standard conditions with skewed category frequencies (i.e. 2:1 ratio for positive and negative descriptions), but not under conditions with equated category frequencies (i.e. 1:1 ratio for positive and negative descriptions). Contrary to these simulations, our behavioral experiments revealed an IC under both conditions, which did not decrease over time. Thus, information loss alone is not sufficient as an explanation for the formation of ICs. These results imply that negative items contribute to ICs not only due to their infrequency, but also due to their emotional salience.

1. Introduction

The ability to extract regularities from a limited number of observations is one of the most fundamental tools an organism needs for adaptive behavior. Humans are even able to implicitly learn complex artificial grammars (e.g. Reber, 1967). In fact, humans have such a strong propensity to detect patterns that they perceive contingencies in environments, even when there are no contingencies – a phenomenon called illusory correlation (Chapman, 1967; Hamilton & Gifford, 1976).

More technically, an illusory correlation (IC) is a subjectively perceived correlation between two events, which differs systematically from the actual covariation between those events (e.g. Chapman, 1967; Fiedler, 2000). The two events might actually not correlate at all or correlate in the opposite direction as reported. ICs have been investigated in basic research (e.g. Chapman, 1967; Tversky & Kahneman, 1973) as well as in applied research, like psychodiagnostics (Chapman & Chapman, 1967), clinical psychology (Alloy & Abramson, 1979), or organizational psychology (Feldman, Camburn, & Gatti, 1986). Very fruitful investigations on the IC were conducted in stereotyping research (e.g. see Hamilton, 1981 or Stroessner & Plaks, 2001 for reviews).

There are (at least) two different types of ICs discussed in the stereotyping literature – each associated with a specific pattern of results (Hamilton, 1981; Tversky & Kahneman, 1973; see Fiedler, 2000, for a more fine-grained classification): the expectancy-based and the distinctiveness-based illusory correlations. In expectancy-based ICs, participants already have an expectation about the relationship of two variables, based on their experiences and personal beliefs. When study participants have to judge the covariation between well-known groups (e.g. accountants and salesmen) and certain traits (e.g. timid and talkative) in a new set of stimuli, their

* Corresponding author at: Department of Psychology, Experimental Neuropsychology Unit, Saarland University, Campus, Building A2.4, D-66123 Saarbrücken, Germany.

E-mail address: m.weigl@mx.uni-saarland.de (M. Weigl).

https://doi.org/10.1016/j.concog.2018.06.002
Received 26 September 2017; Received in revised form 28 April 2018; Accepted 3 June 2018
1053-8100/

judgment on the new set is usually consistent with their pre-experimental expectations (Hamilton & Rose, 1980).

In contrast, participants in experiments on the distinctiveness-based IC have to infer a correlation about material for which they do not possess preexisting expectations about the relationship. In the framework of the current study, we primarily refer to the distinctiveness-based IC.

In the seminal study of Hamilton and Gifford (1976), participants read short descriptions about members of two fictional groups – group A and group B, with group A having twice as many members as group B. For both groups, two-thirds of the description referred to desirable behavior and one-third to undesirable behavior. In other words, group membership and behavior were uncorrelated. Nevertheless, the participants showed a tendency to associate the majority with the frequent, desirable behavior and the minority with the infrequent, undesirable behavior. This pattern was observed consistently across a range of dependent measures (evaluative trait ratings, frequency judgments, and cued recall performance; Hamilton & Gifford, 1976; see Mullen & Johnson, 1990 for a review). The popularity of the concept of ICs stems from the fact that it offers a cognitive explanation for the formation of stereotypes. Moreover, the experimental set-up resembles the situation we encounter in our modern societies: minorities are by definition smaller than majorities and most people behave in a norm-consistent, desirable way (e.g. Alves, Koch, & Unkelbach, 2017a; Fiske, 1980; Kanouse, 1984).

Skewed frequency distributions are assumed to be essential for ICs to arise (e.g. Fiedler, 1991, 1996; Hamilton & Gifford, 1976). In cases in which undesirable behavior is more frequent than desirable behavior, the IC is reversed, i.e. the majority is evaluated less favorable than the minority (Hamilton & Gifford, 1976, Exp. 2; Mullen & Johnson, 1990). However, there is a still ongoing debate about the mechanisms by which skewed frequency distributions influence our judgment and a variety of models have been put forward to explain illusionary correlation. As a consequence, ICs have been investigated from various theoretical perspectives (e.g. accentuation approach: McGarty, Haslam, Turner, & Oakes, 1993; availability account: Rothbart (1981); memory trace model: Smith, 1991; pseudocontingencies: Fiedler, Freytag, & Meiser, 2009; recurrent connectionist model: Van Rooy, Van Overwalle, Vanhoomissen, Labiouse, & French, 2003; Rescorla-Wagner model: Murphy, Schmeer, Vallée-Tourangeau, Mondragón, & Hilton, 2011; attention theory: Sherman et al., 2009). Two prominent accounts are the shared distinctiveness account (SDA) and the information loss account (ILA). Our study was designed to test predictions made on the basis of these two accounts. For the sake of clarity, we will describe only the SDA and ILA at this point; details of the other accounts can be found in the General Discussion.

The SDA states that infrequent combinations are more distinctive and, therefore, better encoded than more common ones (Chapman, 1967; Hamilton & Gifford, 1976; Tversky & Kahneman, 1973). Infrequent combinations are therefore more easily available in memory than others. As individuals estimate the frequency of the combinations on the basis of their availability, infrequent combinations are overestimated (Tversky & Kahneman, 1973).

The evaluation of the empirical evidence for the SDA is unfortunately hampered by the fact that researchers conceptualized distinctiveness quite differently. For example, Hamilton and Gifford (1976) defined distinctiveness as infrequency, whereas Feldman et al. (1986) also considered negativity as distinctive.

Schmidt (2012) identified four different types of distinctiveness: (1) primary distinctiveness (distinctiveness in the immediate context, usually due to infrequency), (2) secondary distinctiveness (distinctiveness over life-time, usually due to bizarreness), (3) emotional significance (emotional engaging stimuli) and (4) high priority stimuli (relevant, but non-arousing stimuli). For the purpose of the current study, we define a stimulus as being distinctive, if it fulfills one of these four criteria.

Evidence for better memory for shared distinctive items stems from studies using free recall (Hamilton, Dugan, & Trolier, 1985) and one-shot ICs (Risen, Gilovich, & Dunning, 2007). Further support can be found in the memory literature: Distinctive items are in general better remembered than non-distinctive items (e.g. Alves et al., 2015; Fabiani & Donchin, 1995; Hunt, 1995, 2009; von Restorff, 1933; see also Schmidt, 1991, 2012, for an integrative account). Furthermore, memory is even better for items that are distinctive on several stimulus dimensions (e.g. Hunt & Mitchell, 1982; Kuhbandner & Pekrun, 2013).

The information loss account (ILA) offers an alternative explanation for ICs without assuming any differential processing of information (Fiedler, 1991, 1996, 2000; Smith, 1991). Since perception and memory are far from perfect, noise can distort parts of information during encoding, storage, and retrieval. For example, due to such noise, we might misremember a rude person from the majority as a member of the minority. If participants are asked to make a judgment about their attitude towards the groups, unbiased aggregation of these distorted data alone is sufficient to lead to the erroneous conclusion that a correlation between groups and behavior is present. In the typical IC experiments the distribution of the valence of the behavior is skewed, i.e. positive stimuli are objectively more frequent than negative stimuli (or vice versa; Fig. 1). For the majority, the preponderance of positive (or negative) behaviors becomes evident to the participant during encoding, because they can aggregate over a large number of instances. Therefore, the subjective frequency estimates for the majority are not so much affected by noise and should roughly correspond to the actual frequencies. The minority, however, is more strongly affected by this noise, because single outliers have more influence on the estimates of smaller samples and the estimates regress to the mean (see 33% and 67% condition in Fig. 1). Thus, the main appeal of the ILA is its parsimony. ICs can be explained without assuming biased processing (e.g. different processing of distinctive and non-distinctive information).

Evidence for the ILA stems from computer simulations that reproduce the IC effect without the assumption of biased processing (Fiedler, 1996, 2000; Smith, 1991). But there is also experimental evidence that the overestimation of frequencies increases when categories are split into sub-categories (Fiedler, 1991; Fiedler & Armbruster, 1994) or that ICs can be observed even in the absence of distinctive or infrequent information (Fiedler, 1991; Shavitt, Sanbonmatsu, Smittipatana, & Posavac, 1999; Van Rooy, Vanhoomissen, & Van Overwalle, 2013).

The SDA and the ILA are not mutually exclusive. However, it is necessary to test boundary conditions to judge the relative merit of both accounts: One highly important boundary condition arises naturally from a closer look at the regression to the mean argument of