Cognition 160 (2017) 110–126

Contents lists available at ScienceDirect

Cognition

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

Original Articles

A referential theory of the repetition-induced truth effect

Christian Unkelbach *, Sarah C. Rom

University of Cologne, Germany

ARTICLE INFO

Article history:
Received 10 August 2016
Revised 27 December 2016
Accepted 30 December 2016
Available online 12 January 2017

Keywords:
Illusory truth effect
Referential theory
Fluency
Familiarity

ABSTRACT

People are more likely to judge repeated statements as true compared to new statements, a phenomenon known as the illusory truth effect. The currently dominant explanation is an increase in processing fluency caused by prior presentation. We present a new theory to explain this effect. We assume that people judge truth based on coherent references for statements in memory. Due to prior presentation, repeated statements have more coherently linked references; thus, a repetition-induced truth effect follows. Five experiments test this theory. Experiment 1–3 show that both the amount and the coherence of references for a repeated statement influence judged truth. Experiment 4 shows that people also judge new statements more likely "true" when they share references with previously presented statements. Experiment 5 realizes theoretically predicted conditions under which repetition should not influence judged truth. Based on these data, we discuss how the theory relates to other explanations of repetition-induced truth and how it may integrate other truth-related phenomena and belief biases.

1. Introduction

> "*It was Napoleon, I believe, who said that there is only one figure in rhetoric of serious importance, namely, repetition. The thing affirmed comes by repetition to fix itself in the mind in such a way that it is accepted in the end as a demonstrated truth.*"
> [Gustave Le Bon, The Crowd: A Study of the Popular Mind, 1895]

Judging whether information is true or not is one of the most important tasks people perform. And as stated by Gustave Le Bon in 1895, there is one robust effect that influences such truth judgments: repetition. People believe repeated statements more than new statements. Since the seminal demonstration of this "illusory truth effect" by Hasher, Goldstein, and Toppino (1977), the effect has replicated across many experiments (see Dechêne, Stahl, Hansen, & Wänke, 2010, for a review), with information ranging from trivia questions to statements about societal issues to opinions about consumer products (Arkes, Hackett, & Boehm, 1989; Bacon, 1979; Johar & Roggeveen, 2007, respectively); the effect is present with repetition intervals from minutes to weeks to months (Brown & Nix, 1996; Garcia-Marques, Silva, Reber, & Unkelbach, 2015; Schwartz, 1982, respectively); and the effect bridges different areas of psychology, from cognitive psychology (Begg, Anas, & Farinacci, 1992), to social psychology (Fragale & Heath, 2004), to consumer research (Hawkins & Hoch, 1992), to the psychology of aging (Parks & Toth, 2006), and to cognitive neuroscience (Wang, Brashier, Wing, Marsh, & Cabeza, 2016).

The effect that people judge repeated information as relatively more true may seem trivial. Yet, it becomes a puzzling problem upon longer consideration: Why should someone believe repeated information more than new information? Wittgenstein already addressed this puzzle in his "Philosophical Investigations" (1955, p. 147), and compared the tendency to believe information simply because it is repeated to "... buying several copies of the morning paper to ensure that the content is true" (see also Unkelbach, Fiedler, & Freytag, 2007). Begg and colleagues (1992) similarly asserted that "...there is no logical reason for repetition to affect rated truth or for earlier information to be trusted more than later information" (p. 447). So indeed, why do people believe repeated information more?

Previous researchers suggested explanations such as an accrual in a statement's corresponding frequency attribute (Hasher et al., 1977), convergent validity (Arkes, Boehm, & Xu, 1991), repetition-induced familiarity (Begg et al., 1992), or processing fluency (Reber & Schwarz, 1999; Unkelbach, 2007). Building on these accounts and the available data, we assume that people judge truth based on corresponding references for statements in memory and the coherence of these references. We first discuss the currently dominant processing fluency explanation; then, we delineate the present theory; and finally, we present five experiments testing the theory's main propositions that judged truth is a function of available corresponding references and their coherence.

* Corresponding author at: Social Cognition Center Cologne, Universität zu Köln, Richard-Strauss-Str. 2, 50931 Köln, Germany.
E-mail address: christian.unkelbach@uni-koln.de (C. Unkelbach).

http://dx.doi.org/10.1016/j.cognition.2016.12.016
0010-0277/

1.1. The processing fluency explanation

Processing fluency is a meta-cognitive experience that people use in many judgments (for overviews see Alter & Oppenheimer, 2009; Unkelbach & Greifeneder, 2013). As repetition increases the fluency of information processing (Feustel, Shiffrin, & Salasoo, 1983), this increase might be responsible for the truth effect. Processing fluency also explains truth effects that are not based on repetition; for example, when statements are presented in high contrast rather than low contrast (Reber & Schwarz, 1999), or when statements rhyme (McGlone & Tofighbakhsh, 2000). Processing fluency thereby provides a potential unified explanation for truth effects. In addition, fluency explains why statements initially labeled as "false" are regarded as true over time (Brown & Nix, 1996; Garcia-Marques et al., 2015): the "false" label is lost over time, but the increased fluency due to prior processing remains intact (Unkelbach & Stahl, 2009).

Due to the substantial empirical support, the fluency explanation is widely accepted; however, it necessitates non-trivial additional assumptions. Processing fluency per se does not explain *why* people respond to fluently processed information with higher rated truth (see Unkelbach, 2006); in contrast, fluency is per se an explanation for increases in liking, because the experience itself seems to be hedonically marked (Winkielman, Schwarz, Fazendeiro, & Reber, 2003). Processing fluency's inherent positivity indeed leads to an additional assumption that might explain fluency effects on truth; because people regard "true" responses most likely more positive than "false" responses, participants might be biased by the positive fluency experience to respond positively ("true"), or they might even employ a direct "positive, therefore true" heuristic (Unkelbach, Bayer, Alves, Koch, & Stahl, 2011; but see Hilbig, 2012). A second possible assumption is that fluency is interpreted as familiarity, which may lead to an experience of convergent validity (Arkes et al., 1991; Hawkins & Hoch, 1992), and thus higher rated truth. A third assumption is that people have implicit or explicit theories about what fluent processing indicates (e.g., truth, frequency, or recency; Schwarz, 2004; Song & Schwarz, 2009). Similarly, Unkelbach (2006), Unkelbach (2007) as well as Herzog and Hertwig (2013) suggested that people learn to interpret fluency based on ecological correlations between fluency and external criteria. Both lay theories and learned interpretations have in common that the influence of fluency is context-dependent, and Unkelbach and Greifeneder (2013) summarized the idea of lay theories and learned interpretations into a general model of fluency effects on judgments. However, both accounts either need explanations where the lay theories come from or must rely on benign learning environments that provide veridical feedback (see Hogarth, 2001).

The following referential theory is compatible with the fluency explanation, and indeed, it assumes processing fluency as the cause for non-repetition-based truth effects (McGlone & Tofighbakhsh, 2000; Reber & Schwarz, 1999; Unkelbach, 2007). However, it solves the problem of non-trivial additional assumptions, it provides a process model for the repetition-induced truth effect, and it makes novel predictions regarding repetition-induced truth.

1.2. A referential theory

In linguistics, the term "referential theory" denotes the theory that the meaning of a word lies in the reference to the object it describes. The same way the real world gives meaning to a word, we propose that the judged truth of a statement is informed by (a) corresponding references in memory that give meaning to the elements in the statement and (b) the coherence of these references in memory. Thereby, we base judged truth on the two major constituents of truth from philosophy, namely correspondence and coherence (see Kirkham, 1992). A full treaty on these concepts from a philosophical side is beyond the present scope, but empirically, we will treat corresponding references as memory traces that provide meaning for the elements in a given statement, and coherence as the internal relational consistency of these corresponding references (see Dunwoody, 2009; Kirkham, 1992, p. 104; Kunda & Thagard, 1996).

Fig. 1a illustrates the hypothesized process of a truth judgment. Consider for example the statement: "The world's most poisonous snake is the Australian Inland Taipan." Fig. 1a's left part assumes that one might have corresponding memory references that provide coherent meaning for the words: Australia has many poisonous animals and that a Taipan is indeed a snake. As these are coherent corresponding references for the statement's elements,

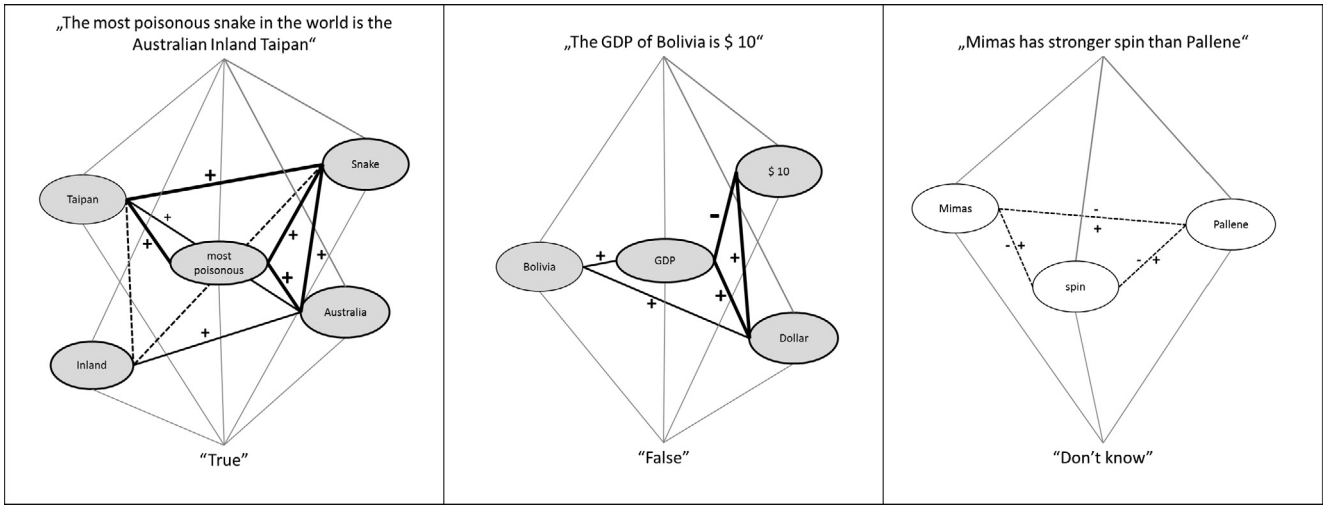

Fig. 1a. Illustration of three localized information networks. The light grey lines indicate incoming (here: the statement) and outgoing (here: the judgment) information. Corresponding references in memory give meaning to the elements in the incoming information and the links between the references determine whether the resulting information networks are coherent or incoherent, resulting in an observable judgment. Grey circles indicate existing references in memory, white circles indicate references that are instigated by the presented information. Solid black lines indicate references that are linked and dotted lines indicate references that are not a priori linked, but instigated by the presented information. Finally, plus signs indicate an excitatory link, minus signs an inhibitory link, with line and sign size indicating link strength. Following Kunda and Thagard (1996), the process is depicted as a propositional, symbolic network, but it is easily implemented within a parallel distributed, sub-symbolic network.