Running Head: Repetition and Belief Formation

Gullible but functional:

Information repetition and the formation of beliefs

Christian Unkelbach & Alex Koch

Universität zu Köln

Word count: 5710

Corresponding Author:

Christian Unkelbach

Social Cognition Center Cologne

Universität zu Köln

Richard-Strauss-Str. 2

50931 Köln

Germany

Email: christian.unkelbach@uni-koeln.de

Gullible but functional: Information repetition and the formation of beliefs

When considering how people come to form beliefs about the world they live in, they seem to be rather gullible. People are convinced by weak arguments (Petty & Cacioppo, 1986, Petty, Wells, & Brock, 1975), they do not weight information properly (Dawes, 1979), and most critically, they fail to discount irrelevant information when forming beliefs about the world (see Wilson & Brekke, 1994, for an overview). One of the most notorious ways people are influenced by irrelevant information is mere repetition. On first sight, simply repeating information should not change its informational value, should not increase its validity, change its veracity, or alter its content. Encountering the same piece of information twice should not influence people's assessment of this information and its impact on people's beliefs.

However, simply repeating information *does* increases its subjective truth (Hasher, Goldstein, & Toppino, 1977; see Dechêne et al., 2010, for a review). This seemingly irrational tendency was already discussed and acknowledged by Wittgenstein in his "Philosophical Investigations" (1955). Wittgenstein famously stated that repeating informational input does not help to ascertain that information, and one cannot "… buy several copies of the morning paper to ensure that the content is true" (Wittgenstein, 1955, p. 147). Similarly, Begg and colleagues stated 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).

Repetition may thus seem "empty" from philosophical, computational, and logical perspectives; however, from psychological perspective, repetition is a key element in learning and memory (Ebbinghaus, 1885/1971; Hintzman & Block, 1971; McClelland, McNaughton, & O'Reilly, 1995). It increases the subjective value of stimuli (Betsch, Plessner, Schwieren, & Gütig, 2000; Unkelbach, Fiedler, & Freytag, 2007; Zajonc, 1968), and it establishes and strengthens perceived links between stimuli as in associative and evaluative learning (Rescorla & Wegner, 1972; De Houwer, Thomas, & Baeyens, 2001). Considering this substantial impact of repetition on many psychological processes, it is less surprising that repetition also influences how people judge information's truth and impacts personal beliefs.

In experimental psychology, the phenomenon that simply repeating information increases its subjective truth is labelled the repetition-induced truth effect (Hasher, Goldstein, & Toppino, 1977; see Dechêne, Stahl, Hansen, Wänke, 2010, for a review). This repetition-induced truth effect and its influence on the formation of beliefs is one of the empirically most robust psychological effects (see Dechêne et al., 2010, p. 239). It is also of great practical interest. For example, when information is ever more often shared, reposted online, or multiplied via social media, the increase in subjective truth due to mere repetition may explain the apparent increase in evidently false beliefs. Prominent examples are conspiracy theories (e.g., "9/11 was an inside job"; "Vaccinations cause autism"), urban legends (e.g., "The hoover dam is built with dead bodies"; "Children tattoos contained LSD in the Sixties"), but also single pseudo facts (e.g., "The Great Wall of China is visible from the moon") or "fake" news (e.g., "FBI agent suspected in Hillary [Clinton] email leaks found dead in apartment murder-suicide"). Due to repetition, these statements might become more believable.

In the following, we will first provide examples of the effect and then explain its theoretical backgrounds. Based on these backgrounds, we will argue that while the repetition-induced truth effect sometimes has detrimental effects on what people believe to be true (e.g., conspiracy theories, urban legends, single pseudo facts, or "fake" news), it is overall psychologically functional to believe repeated information more than novel information.

Truth by Repetition

The idea that repetition is a key variable in persuasion, subjective truth, and ultimately, the formation of beliefs about the world, is well-embedded not only in psychology. In the classic treaty "The crowd: A Study of the popular mind" by Gustave Le Bon (1895/1996), the author already stated: *"*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."* (Book II, Chapter III: 2. The Leaders of Crowds and their Means of Persuasion).

Similarly, in Aldous Huxley's (1932/2008) novel *Brave New World*, children are taught not only knowledge, but also moral lessons by repeating the same notions time and again while they

sleep: “Sixty-two thousand four hundred repetitions make one truth (p. 47).”[1] And in general, people use the notion that repetition indicates information’s truth with the simple rule of thumb that if they have learned something somewhere before, it is likely to be true. For sure, if someone remembers reading the information in the Encyclopedia Britannica, it is well justified to believe this information is more than completely novel and false. However, as we will see, repetition effects extend well beyond what can be rationalized so easily. Repetition also influences people’s beliefs if it comes from the identical source (“I told you, vaccinations cause autism.”), if it is labelled as false (“It is false that vaccinations cause autism”), or even when the initial presentation is incompatible with the second presentation (“Vaccinations do not cause autism”; see below). So how does this strong influence of repetition on subjective truth and the formation of beliefs occur? Below, we provide a historical overview of the explanations for the repetition-induced truth effect.

A history of explanations for the repetition-induced truth effect

The basic effect

Hasher, Goldstein, and Toppino (1977) presented the first empirical evidence of the repetition-induced truth effect in the psychological literature, and their basic design is still prevalent today. Participants heard statements from a large pool of topics and different subjects during a presentation phase (e.g., “The thigh bone is the longest bone in the human body”) and were told that some of these are false and some are true. Factually, half of the statements were false, half were true. Two, four, and six weeks later, participants rated lists of “new” (not heard before) and “old” (heard before) statements on a scale from 1 (“definitely false”) to 7 (“definitely true”). They found higher truth ratings for repeated compared to new statements even up to six weeks later. However, they provided no direct evidence for a psychological mechanism but concluded that people use mere frequency to attribute validity (i.e., “truth”) to statements.

The recognition explanation

Bacon (1979) tested and offered the explanation that people assign truth to repeated information simply because they remember the statements. Bacon showed that there is a correspondence between participants’ recognition judgments (i.e., “old” vs. “new”) and their rated

truth; that is, whether they recognized a statement influenced the truth judgments. The data also contradicted the frequency explanation by Hasher and colleagues (1979) because the factual status of a statement (i.e., repeated vs. novel) had less influence than the subjective repeated vs. novel (or "old" vs. "new") status of that statement. Thus, Bacon concluded that: "Consequently, the repetition effect is not really a repetition effect after all but a recognition effect." (p. 251). Please note that at this point, if recognition would be the sole correct explanation of the repetition-induced truth effect, then the effect would be highly irrational and dysfunctional. That is, it would be irrational for people to judge statements to be true just because they remember them from an experimental session two weeks earlier; because two weeks earlier, they also learned that they might be true or false(see above for the basic paradigm).

The familiarity explanation

Arkes, Boehm, and Xu (1991) offered and tested two explanations for the repetition-induced truth effect. The first mechanism underlying the effect the authors labelled as *referential validity*. If two independent sources provide the same information, that is, repeated information, then the information is more likely true, just because it is very unlikely that two independent sources provide the same false information. This follows because statements can be false in many different ways while there is usually only one true version of a statement. This is particularly true when it comes to statements about the physical world (see Alves, Koch, & Unkelbach, 2017, for a more general version of this argument).For example, Budapest may be a city in Romania or Argentina or Turkmenistan (all false), while there is only one correct statement: Budapest is a city in Hungary. This reasoning is also employed for judging the validity of eyewitness testimonies (i.e., two eyewitnesses independently reporting the same information makes it more likely true). It is also one of the most frequent strategies to assess validity when people search information online. For example, when two people independently praise the virtues of a given product in their reviews, or when two news sites independently report the same political events, they are more likely to be taken as true. For the repetition-induced truth effect, this implies that people recognize information but overlook that it

comes from the same source as before (i.e., another experimental session rather than an independent outside source).

However, across their experiments, Arkes and colleagues (1991) found no evidence that forgetting the information source is necessary. Rather, they found support for their second suggested mechanism, namely that subjective familiarity with the statements determines judged truth. As familiarity also influences recognition judgments, subjective familiarity was a candidate for replacing recognition (Bacon, 1979) as the underlying mechanism of the repetition-induced truth effect. Do people believe information that feels familiar?

To pit recognition against familiarity, Begg, Anas, and Farrinacci (1992) employed a so-called process dissociation procedure (see Jacoby & Kelley, 1992, for an easy introduction). Their participants heard statements from sources that were labelled as "true" or "false"; for example, given a male and a female speaker, participants learned that all statements by the male speaker would be false, and all statements by the female speaker would be true. Begg and colleagues found that repeated, and thus, more familiar statements from a "false" information source were more likely to be judged as true compared to new statements. Repeated, more familiar statements from a "true" information source, however, were most likely to be taken as true. Thus, the authors concluded independent contributions of both familiarity and recognition to judged truth.

The central role of familiarity was further supported by experiments showing that even statements labelled as blatant lies benefited from repetition (Brown & Nix, 1996), or that even information ("Crocodiles sleep with their eyes open") that directly contradicted the original information ("Crocodiles sleep with their eyes closed") became more believable due to the repetition of the semantic content (Garcia-Marques, Silva, Reber, & Unkelbach, 2015). These effects should not occur if people would factually remember the original encounter or recognize the statements.

The fluency explanation

The experiments by Begg and colleagues (1992) placed the repetition-induced truth effect into one larger category of effects caused by experiences (here: familiarity) elicited by the stimulus, such as the mere exposure effect (Mandler, Nakamura, & Van Zandt, 1987; Zajonc, 1968), the

revelation effect (Watkins & Peynircioglu, 1990; see also Topolinksi & Reber, 2010), or the false fame effect (Jacoby, Kelley, Brown, & Jasechko, 1989). Whittlesea (1993) proposed that information familiarity is not a direct output from memory, but results from the automatic attribution that fluent processing of the respective information is due to a previous encounter. Thereby, processing fluency, which is the experienced ease of ongoing mental processes (Unkelbach & Greifeneder, 2013), became a candidate as the central explanatory construct for the repetition-induced truth effect.

Reber and Schwarz (1999) directly tested whether processing fluency (i.e., the feeling of easy processing) influenced judged truth directly, without actually repeating information. Instead of repeating statements, they presented simple statements (e.g., “Osorne is a city in Chile”) in fluent, easy-to-read colors (e.g., dark red or dark blue) or disfluent, difficulty-to-read colors (e.g., green and yellow). Although their overall effect was small, participants rated statements in difficult-to-read colors as less true compared to statements in easy-to-read colors. Similarly, McGlone and Tofighbakhsh (2000) showed that people believe aphorisms that rhyme and are thus fluently processed (“Woes unite foes”) more than content-identical aphorisms that do not rhyme and are thus less fluently processed (“Woes unite enemies”).

Fluency as the central explanatory construct for repetition-induced truth was further supported by experiments by Unkelbach (2007), which addressed two critical points: First, showing that the repetition-induced truth effect it is not a mere exposure effect. Second, it is indeed processing fluency that mediates the effect. Concerning the first point, if the repetition-induced truth effect is indeed due to subjective experiences elicited by a stimulus (e.g., a statement “feels” familiar), it might not be a fluency effect, but rather a mere exposure effect (Zajonc, 1968, 2001). The mere exposure effect is the acquisition of preferences due to the repeated exposure to stimuli; in other words, people like repeated things. The repetition-induced truth effect may then follow simply because people like repeated information more than novel information and express this preference with a positive truth rating. Alternatively, people might employ a “positive, therefore true” heuristic (Unkelbach, Bayer, Alves, Koch, & Stahl, 2011; but see Hilbig, 2012).

Unkelbach (2007) argued that fluency effects depend on the interpretation of the fluency experience (see also Unkelbach & Greifeneder, 2013), and a truth effect follows because people interpret fluent processing as a cue for a statement's truth. If people learn a different interpretation of processing fluency when judging truth (e.g., fluent processing as a cue for a statements falseness), a fluency explanation predicts that people should take repeated and thus fluently processed information as false instead of true. In contrast, mere exposure should unconditionally lead to higher rated truth of repeated information. Thus, participants encountered statements in a training phase for which truth correlated with their processing fluency. In a standard condition, truth and processing fluency were positively correlated; for example, true statements such as "Dolphins are mammals" were presented in dark blue or dark red and thus easy to read, while false statements such as "Lead is lighter than aluminum" were presented in light green or light yellow and thus difficult to read. In a reversed condition, *false* statements were easy to read and *true* statements were difficult to read. This latter condition reversed the typical color-based truth effect found by Reber and Schwarz (1999) in the following test phase when participants judged the truth of easy or difficult to read statements (see Olds & Westerman, 2012, for similar fluency reversals). The training with colors also reversed the repetition-induced truth effect based on repeated and novel statements that were both printed in black against a white background; that is, the fluency training transferred from one fluency source (i.e., color contrast) to another fluency source (i.e., repetition)

This finding clearly showed that it is people's interpretation of processing fluency that underlies the repetition-induced truth effect; otherwise standard versus reversed training with color should not influence the effect of repetition on judged truth. Further, mere exposure would have predicted a main effect of repetition independent of standard versus reversed training with color. Thus, processing fluency was established as the construct that explains both the repetition-based and non-repetition-based (e.g., color, rhyming) truth effect.

Yet, empirically, non-repetition-based fluency manipulations usually yield smaller truth effects than repetition-induced truth effects. For example, Hasher and colleagues (1977) reported a truth effect of $d = 0.84$ for repeated compared to new statements (estimated from Table 1 in Hasher

et al., 1977). The color-based truth effect by Reber and Schwarz (1999) was substantially smaller, namely *d* = 0.13. With some exceptions (e.g., Unkelbach, 2007, Exp. 2), stronger repetition-based truth effects are apparent in most data sets (e.g., compare Hansen, Dêchene, and Wänke, 2008, with Hansen, Dêchene, and Wänke 2009). Obviously, this could be due to fluency effects from repetition being stronger than fluency effects from, for example, color contrast. However, the pattern may also suggest that processing fluency and repetition influence truth via different processes. Addressing this issue, Silva, Garcia-Marques, and Mello (2015) directly compared perceptual-based and repetition-based fluency effects on truth and concluded that repetition and perceptual fluency influence truth judgments in different ways: "It seems that repetition has a stronger connection to truth, which is also less malleable than in the case of perceptual fluency […] truth effects due to perceptual fluency are likely to have another origin…" (p. 13).

A referential theory

Besides the empirical problem of smaller truth effects based on perceptual compared to repetition-induced fluency, the fluency explanation necessitated additional assumptions. In a nutshell: Why do people use fluency as a cue for truth rather than for falseness? One answer was provided by assuming that people learn to interpret processing fluency as "truth" (Unkelbach, 2006; Greifeneder & Unkelbach, 2013). Another answer was that people have lay theories for the meaning of processing fluency (Schwarz, 2004; Greifeneder & Schwarz, 2014). That is, people either need to learn that fluent processing is indicative of truth, or they need a communicated lay theory that makes the connection between truth and fluency (e.g., "If it feels fluent, it must be true"). This is non-trivial because one must assume a highly benevolent learning environment that allows to establish the link between truth and fluency, or an explicit source. While possible, both additional assumption.

To address the empirical difference between "pure" fluency effects and repetition-based fluency effects, and the source of the interpretation, Unkelbach and Rom (2017) proposed a *referential theory* of the repetition-induced truth effect. The theory starts from a philosophical point asking how people may in general judge the validity, veracity, or "truth" of a given piece of

information or statement in a truth judgment paradigm. Most philosophical theories of truth incorporate two major elements for such judgments, namely correspondence and coherence (Kirkham, 1992). Psychologically, one may see correspondence as the references in memory that provide meaning for the elements of a given statement. Coherence is then the relational consistency of these corresponding references. For example, the statement “The world’s highest tree is a Sequoia tree in California.” should have corresponding references in memory that provide meaning for the statement’s elements. Upon hearing the statement, most people will have memory references for the elements “world”, “Sequoia”, “tree”, and “California”. For most people, these corresponding references have a high degree of consistency. For example, California is a U.S. American state within the world, Sequoias are trees that grow in California, and Sequoias are also typically tall trees. Thus, the statement has corresponding references that are coherent and is thus likely to be judged as “true”.

How correspondence and coherence may inform truth becomes apparent if one changes one element in the statement, for example, “The world’s highest tree is a Sequoia in Antarctica.” Assuming that California and Antarctica have the same number of corresponding references, the two statements do not differ in their number of references, but the California statement is highly coherent, while the Antarctica statements is incoherent. Sequoia trees do not grow in the Antarctic. As a result, people on average should believe the California statement, but not the Antarctica one. Figure 1 illustrates this process (adapted from Unkelbach & Rom, 2017, Figure 1a).

Figure 2 (adapted from Unkelbach & Rom, 2017, Figure 1b) then illustrates the implications of truth judgments based on corresponding memory references and their coherence for the repetition-induced truth effect. When people hear or read a novel piece of information or a novel statement (illustrated by the light grey lines), as in the typical exposure phase of a truth experimentin the tradition of Hasher and colleagues (1977), the statement activates corresponding references within memory (e.g., “California” etc.) and their respective links. For novel information in the statement (e.g., “Sequoia” for a person who does not know a Sequoia is a type of tree), a corresponding reference will be formed and – if no inconsistency is apparent – coherently linked to

the activated corresponding references (“Sequoias are trees that grow in California”). In the subsequent test phase, repeated statements will thus have more corresponding references that are coherently linked (see Figure 2’s right part) than new statements; this is why new statements will appear relatively less true than repeated statements.

A truth effect based on fluent processing without repetition then occurs because many corresponding references that are coherently linked increase a statement’s fluency of processing; thus, fluency is also an output from memory rather than the ultimate explanation why repetition increases judged truth. Conceptualizing processing fluency this way solves the theoretical problem of how people learn to associate fluent processing with truth. The referential theory assumes that people believe statements and judge them to be true when they activate many coherently linked corresponding references. As many activated, coherently linked corresponding references increase both subjective truth and subjective processing fluency, people experience and learn that truth and fluency are correlated. This learned interpretation of processing fluency as a truth signal then leads to fluency-based truth illusions when fluency is manipulated independent of repetition (e.g., by color contrast or rhyming). I

It is also important to emphasize that the referential theory predicts that new statements do not appear false, but simply as relatively less true, a boundary condition that is established for typical truth paradigms (Hansen, Dechêne, & Wänke, 2008; see also Wänke & Hansen, 2015). The theory thereby shares Gilbert’s (1991) view that people by default believe incoming information the same way they believe the existence of physical objects upon seeing them. However, new statements will appear less true than repeated statements because they have typically less corresponding references that are coherently linked.

The theory is consistent with philosophical considerations of how people judge truth, and it fully explains all the data on the repetition-induced truth effect; statements that have more coherently linked corresponding references will also be higher in recognition rates, higher in familiarity, and higher in fluent processing. And if one assumes that the links between items in memory constitute “knowledge” (e.g., a Sequoia is a tree), the theory also explains the interactions

of knowledge and repetition when judging truth (Fazio, Brashier, Payne, & Marsh, 2015; Unkelbach & Stahl, 2009).

Finally, the referential theory also explains political and other kinds of partisanship or, in other words, why people with agendas and tastes maintain beliefs that contradict others' beliefs. For example, conservatives may repeatedly fantasize about, speculate with fellow partisans about, or one-sidedly read or hear about positive but not necessarily true aspects of conservative, past society. This should increase the amount of coherently linked corresponding references activated by statements conveying the same or related positive but not necessarily true aspects of conservative, past society. As a result, statements in favor of conservative, past society should become more and more believable, whereas due to an increased number of incoherent memory links and decreased processing fluency statements in favor of progressive, future society should become less believable (vice versa for progressive partisans, of course).

In sum, the referential theory may provide insights into the cognitive, not necessarily intended and / or controllable mechanisms by which opposing agendas and interests evolve into ever more polarized and conflicting beliefs (see also below): At some point statements by those with opposed agendas and interests may become so incoherent with activated corresponding references and thus so difficult to process that the likelihood of reasonable synthesis from thesis and antithesis approaches zero.

Why truth by repetition may be functional

The repetition-induced truth effect seems to be an extreme form of human gullibility. Again, simply telling people the same thing twice should not make the message more believable. But apparently it does. However, we want to argue that the repetition-induced truth effect might be functional after all, both from a fluency perspective as well as from a referential perspective.

To argue that such a truth effect might be functional necessitates a definition of what is functional, rational, or adaptive. This is not as trivial as it seems (see Reber & Unkelbach, 2010, for a more detailed treaty). For example, William James (1909/1975) argued that a belief is justified if the belief increases the utility of the believer. However, such a utilitarian notion of belief justification

does not necessarily align with the factual state of affairs or what is commonly seen as "truth". For example, people with a family history of cardiovascular diseases might worry about their blood pressure, thereby increasing their blood pressure and incidence chance of cardiovascular malfunctions. Believing the factually false information that they have no such family history may actually have beneficial effects for their blood pressure (see Kirkham, 1992, for other justifications of belief).

Here, we do not follow such a utilitarian approach to belief justification, but follow a naïve empirical or rational approach. First, one needs to assume that there *exists* a true empirical state of affairs. Second, beliefs that correspond more with these true states are to be preferred over beliefs that correspond less with these true states. That is, the belief "The earth is round." is preferable over the belief that "The earth is flat." Although the former is not a perfectly true description of the oblate spheroid form of the earth, it corresponds more with the empirically accessible facts about the earth's shape. Third, an effect, such as the repetition-induced truth effect, is functional if the existence of the effect leads on average to more beliefs that correspond with the assumed true state of affairs than when the effect would be non-existent. Put more simply, does the repetition-induced truth effect lead to more "true" beliefs or to more "false" beliefs? In the following, we address such a functionality both from a fluency perspective and the referential perspective.

Functionality from a fluency perspective

Unkelbach (2007) explicitly assumed that people use processing fluency in truth judgements because there is a correlation between factual truth and processing fluency. In other words, there should be a positive ecological correlation between people's fluency experiences and truth (see also Hertwig, Herzog, Schooler, & Reimer, 2008; Herzog & Hertwig, 2013). This positive correlation may exist for a normative as well as a practical reason. First, normatively, Grice (1975) proposed the maxims of quality and manner in interpersonal communications. That is, people following theses maxims should communicate truthfully (quality) and in a fluent, comprehensible way (manner). As people follow these two maxims differentialy (e.g., people in a negative mood follow Grice's maxims

more than people in a positive mood; Koch, Forgas, & Matovic, 2013; Matovic, Koch, & Forgas, 2014), message receivers should observe a positive correlation between truth and processing fluency.

Practically, most people should communicate truthfully most of the time, as it is hard to imagine a functional society in which false information is more frequent than true information. Thus, true information should be more frequent compared to false information in the world. Higher frequency entails statistically a higher chance of redundancy and repetition (see also Alves, Koch, & Unkelbach, 2017; Koch, Alves, Krüger, & Unkelbach, 2016), also contributing to a correlation between truth and processing fluency. In addition, physical reality constrains truth such that there is only one way for information to be true, but many ways to be false. For example, one may state that the earth's shape is a plane, a cube, or sphere, but only one can be true. Similarly, the world's highest tree might be a spruce, a sequoia, or a Eucalyptus tree, but only one can be true. Thus, due to the high variety of potentially false information, true information is more likely to be repeated (leaving aside strategic miscommunication or false facts / fake news that are often repeated, see below).

One might reformulate this assumed positive correlation into the assumption that the conditional probability of truth given a fluently processed statement (or any information) is larger than 50% (see Reber & Unkelbach, 2010, for a full treaty). Given our definition above (i.e., beliefs are the more justified the more they correspond to truth), the repetition-induced truth effect is thus functional as long as $p(\text{true}|\text{fluent}) > .50$. Given the practical as well as the normative considerations outlined above, this relation is very likely to hold.

Functionality from a referential perspective

From a fluency perspective, one might simply state that the effect is functional as long as fluency is an ecologically valid cue (see also Unkelbach & Greifeneder, 2013). The $p(\text{true}|\text{fluent}) > .50$ assumption, although highly plausible, may remain untestable empirically, however. A similar empirical problem arises from the referential perspective, although one can also make a logical argument from this side with a simple thought experiment.

The sole necessary assumption to argue for the functionality of repetition-induced truth from a referential perspective is that experiences with regards to the physical world are consistent.

Imagine an environment in which information is only available from concrete input (i.e., direct experience), but not from symbolic input (i.e., language or pictures). Now imagine that someone in this environment observes that Person A gets sick after consuming Plant X. From a referential perspective, this should establish the according links between the corresponding references "Person A", "Plant X", and "sick". Now the observer sees that Person B, C, and D independently also get sick after eating Plant X. Now the network of the established links focuses around "Plant X" and "sick", as these are the corresponding reference in memory that are common to all the observations. Thus, there should be a strong association between "Plant X" and "sick", which might lead to the proposition that "Plant B is poisonous".

However, if the observer already has seen hundreds of other people consuming Plant B and not getting sick, the proposition that Plant B is poisonous will be judged as false. Or, if the observer has consumed Plant B himself/herself and does not get sick, this would provide a strong incoherent corresponding reference. As long as the informational input is based on such first-hand observations, the amount of corresponding references and their (in)coherence will approximate the "true" state of the world in his/her beliefs. Given that there *is* a true state of the world, the present assumptions will lead to beliefs about the world that are approximately correct. In other words, the apparently illusory truth effect might be rooted in a first-hand experience-based learning system that effectively approximates the truth about the world. And the same way the visual system is calibrated by haptic experiences (i.e., learning that the world is not upside down), people might learn that beliefs are true if they have a high and higher number of coherent compared to incoherent corresponding references in memory and can thus be processed fluently, with ease. Thus, the effect is functional according to our definition (i.e., beliefs should approximate truth) as long as it is based on first-hand experiences.

Detrimental effects of repetition-induced truth

The two caveats for both approaches are immediately apparent. From a fluency perspective, one might argue that in these times of effortless automated online communication, false information is as likely to occur as true information. If Grice's (1975) quality but not manner maxim is strategically violated, and if the same false information is strategically repeated, fluency might signal falsehood

and the repetition-induced truth effect would no longer be functional. Such cases occur typically under what is labelled “propaganda”, but might be generalized to any strategic communication attempt. Interestingly, already Le Bon (1896) listed repetition in his chapter on how leaders might control the masses.

Second, from a referential perspective, truth by repetition may no longer functional if memory is no longer based on first-hand experiences, but on symbolic experiences mediated by language because some information is more likely to be communicated than other information. For example, our observer of persons eating Plant X might not see himself/herself cases Person A-D … but read about them in the newspaper, hear about them in the radio, or clearly, find out about them on the internet. Why such symbolic experiences may reduce the functionality of inferring truth from coherence and fluency by repetition is immediately apparent: in our first-hand environment, all persons eating the plant have an approximately equal chance to be observed (perhaps not if the observer is a doctor). As a result, if the factual probability of getting sick after eating Plant X is low, one will observe few cases of sickness after consuming the plant but many cases of Plant B consumptions followed well-being. This will prevent the erroneous belief that Plant B is poisonous. In the world of modern media, however, all the consumption cases without sickness will be most likely not reported, preventing the corresponding references to be established and incoherently linked to “Plant B” and “sick”. Rather, by all likelihood, the few sickness cases will get reported, and most likely repeatedly, increasing the probability that a proposition such as “Plant B causes sickness” will have many coherent corresponding references and thus will become an established belief.

The example parallels some of the most unfortunate false beliefs (here, we use “false” as improbable by any scientific standard; e.g., “Vaccinations cause autism”). All vaccinated children that did not become autistic do not get reported in the newspapers, the radio, on TV, and the internet; if anything, they appear as a summary statistic. In the referential view, however, the single case of autism after vaccination has almost the same impact and value as the summary statistic of all the cases where no evidence for autism after vaccinations was found (i.e., it may be represented as a single corresponding reference). Thus, although the referential perspective implies that the

repetition-induced truth effect is functional, it also provides a boundary model for how symbolic communication can lead to the formation of false beliefs.

Conclusion

The repetition-induced truth effect in the laboratory is a prime example of human gullibility. Whether it turns lies into truths, fiction to fact, or advertisements into successful persuasion, it is seemingly an easy-to-exploit effect. Based on the two explanations for the effect, the fluency explanation and the referential explanation, however, we argue that the effect is in essence highly functional. Easily processed information is more likely true than false, and nature does not contradict itself. Thus, inferring truth from repetition may be an easy and useful shortcut to adequate truth judgment. However, in cases of strategically sent and repeated false communications, inferring truth from repetition comes at the cost of sometimes false beliefs. To the best of our knowledge, however, the belief that the highest tree in the world is a Sequoia in California, is correct.

Footnotes

1. Please note that the character in whose train of thoughts this statement occurs, Bernard Marx, factually does not endorse this practice and reflects on the sentiment with the internal exclamation: “Idiots!”

References

Alves, H., Koch, A., & Unkelbach, C. (2017). The "common good" phenomenon: Why similarities are positive and differences are negative. *Journal of Experimental Psychology: General*, *146*, 512–528.

Arkes, H. R., Boehm, L. E., & Xu, G. (1991). Determinants of judged validity. *Journal of Experimental Social Psychology*, *27*, 576-605.

Bacon, F. T. (1979). Credibility of repeated statements: Memory for trivia. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 241-252.

Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General, 121*, 446-458.

Betsch, T., Plessner, H., Schwieren, C., & Gütig, R. (2001). I like it but I don't know why: A value-account approach to implicit attitude formation. *Personality and Social Psychology Bulletin, 27*, 242-253.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. American Psychologist, 34, 571-582.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). Mix me a list: Context moderates the truth effect and the mere-exposure effect. *Journal of Experimental Social Psychology, 45*, 1117-1122.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review, 14*, 238-257.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Association learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853-869.

Ebbinghaus, H. (1885/1971). *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie [Concerning memory: Investigations in experimental psychology]*. (1st Ed., Leipzig 1885). Darmstadt, Germany: Wissenschaftliche Buchgesellschaft.

Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, *144*, 993-1002.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46,* 107119.

Greifeneder, R., & Schwarz, N. (2014). Metacognitive processes and subjective experience. In Sherman, J. W., Gawronski, B., & Trope, Y. (Eds.). *Dual-process theories of the social mind* (pp. 314-327). New York, NY: Guilford Press.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.). Syntax and semantics: Vol. 3. Speech acts (pp. 41–58). New York: Academic Press.

Garcia-Marques, T., Silva, R. R., Reber, R., & Unkelbach, C. (2015). Hearing a statement now and believing the opposite later. *Journal of Experimental Social Psychology*, *56*, 126–129.

Hansen, J., Dêchene, A., & Wänke, M. (2008). Discrepant fluency increases subjective truth. *Journal of Experimental Social Psychology, 44*, 687–691.

Hansen, J., & Wänke, M. (2010). Truth from language and truth from fit: The impact of linguistic concreteness and level of construal on subjective truth. *Personality and Social Psychology Bulletin*, *36*, 1576-1588.

Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior, 16*, 107-112.

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 1191-1206.

Herzog, S. M., & Hertwig, R. (2013). The ecological validity of fluency. In C. Unkelbach, R. Greifender (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behaviour* (pp. 190-219). New York: Psychology Press.

Hilbig, B. E. (2012). How framing statistical statements affects subjective veracity: Validation and application of a multinomial model for judgments of truth. *Cognition, 125*, 37-48.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology, 88*, 297-306.

Huxley, A. (2008). *Brave new world*. Ernst Klett: Sprachen.

James, W. (1909/1975). *The meaning of truth* (Vol. 2). Harvard University Press.

Jacoby, L. L., & Kelley, C. M. (1992). A process-dissociation framework for investigating unconscious influences: Freudian slips, projective tests, subliminal perception, and signal detection theory. *Current Directions in Psychological Science, 1*, 174-179.

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology, 56*, 326-338.

Kirkham, R. L. (1992). *Theories of truth: A critical introduction*. Cambridge: Massachussets.

Koch, A. S., Alves, H., Krüger, T., & Unkelbach, C. (2016). A general valence asymmetry in similarity: Good is more alike than bad. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1171–1192.

Koch, A. S., Forgas, J. P., & Matovic, D. (2013). Can negative mood improve your conversation? Affective influences on conforming to Grice's communication norms. *European Journal of Social Psychology*, *43*, 326–334.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereotypes, traits, and behaviors: A parallel-constraint-satisfaction theory. *Psychological Review, 103*, 284-308.

Le Bon, Gustave (1895/1996). The Crowd: A Study of the Popular Mind/ Psychologie des foules. The Project Gutenberg Etext of The Crowd, http://www.gutenberg.org/cache/epub/445/pg445.html.

Mandler, G., Nakamura, Y., & Van Zandt, B. J. (1987). Nonspecific effects of exposure on stimuli that cannot be recognized. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 646-648.

Matovic, D., Koch, A. S., & Forgas, J. P. (2014). Can negative mood improve language understanding? Affective influences on the ability to detect ambiguous communication. *Journal of Experimental Social Psychology*, *52*, 44–49.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419-457.

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, *11*, 424-428.

Olds, J. M., & Westerman, D. L. (2012). Can fluency be interpreted as novelty? Retraining the interpretation of fluency in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 653-664.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology, 19*, 123-205.

Petty, R. E., Wells, G. L., & Brock, T. C. (1976). Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology, 34.* 874-884.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*, 338-342.

Reber, R., & Unkelbach, C. (2010). The epistemic status of processing fluency as source for judgments of truth. *Review of philosophy and psychology, 1,* 563-581.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory, 2*, 64-99.

Schwarz, N. (2004). Meta-cognitive experiences in consumer judgment and decision making. *Journal of Consumer Research, 14*, 332-348.

Silva, R. R., Garcia-Marques, T., & Mello, J. (2016). The differential effects of fluency due to repetition and fluency due to color contrast on judgments of truth. *Psychological Research, 80*, 821–837.

Topolinski, S., & Reber, R. (2010). Gaining insight into the “Aha” experience. *Current Directions in Psychological Science, 19*, 402-405.

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science, 17*, 339-345.

Unkelbach, C. (2007). Reversing the truth effect: learning the interpretation of processing fluency in judgments of truth. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 219-230.

Unkelbach, C., Bayer, M., Alves, H., Koch, A., & Stahl, C. (2011). Fluency and positivity as possible causes of the truth effect. *Consciousness and Cognition, 20*, 594–602.

Unkelbach, C., Fiedler, K., & Freytag, P. (2007). Information repetition in evaluative judgments: Easy to monitor, hard to control. *Organizational Behavior and Human Decision Processes, 103*, 37–52.

Unkelbach, C., & Greifeneder, R. (2013). A general model of fluency effects in judgment and decision making. In C. Unkelbach & R. Greifeneder (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behaviour* (pp. 11–32). New York, NY: Psychology Press.

Unkelbach, C., & Rom, S. C. (2017). A referential theory of the repetition-induced truth effect. *Cognition, 160*, 110-126.

Unkelbach, C., & Stahl, C. (2009). A multinomial modeling approach to dissociate different components of the truth effect. *Consciousness and Cognition, 18*, 22–38.

Watkins, M. J., & Peynircioglu, Z. F. (1990). The revelation effect: when disguising test items induces recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 1012-1020.

Wänke, M., & Hansen, J. (2015). Relative processing fluency. *Current Directions in Psychological Science, 24*, 195-199.

Whittlesea, B. W. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1235-1253.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: unwanted influences on judgments and evaluations. Psychological Bulletin, 116, 117-142.

Wittgenstein, L. (1955/1977). *Philosohpische Untersuchungen* [Philosphical Investigations]. Frankfurt, Germany: Suhrkamp.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*, 1-27.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science, 10*, 224-228.

Figure 1

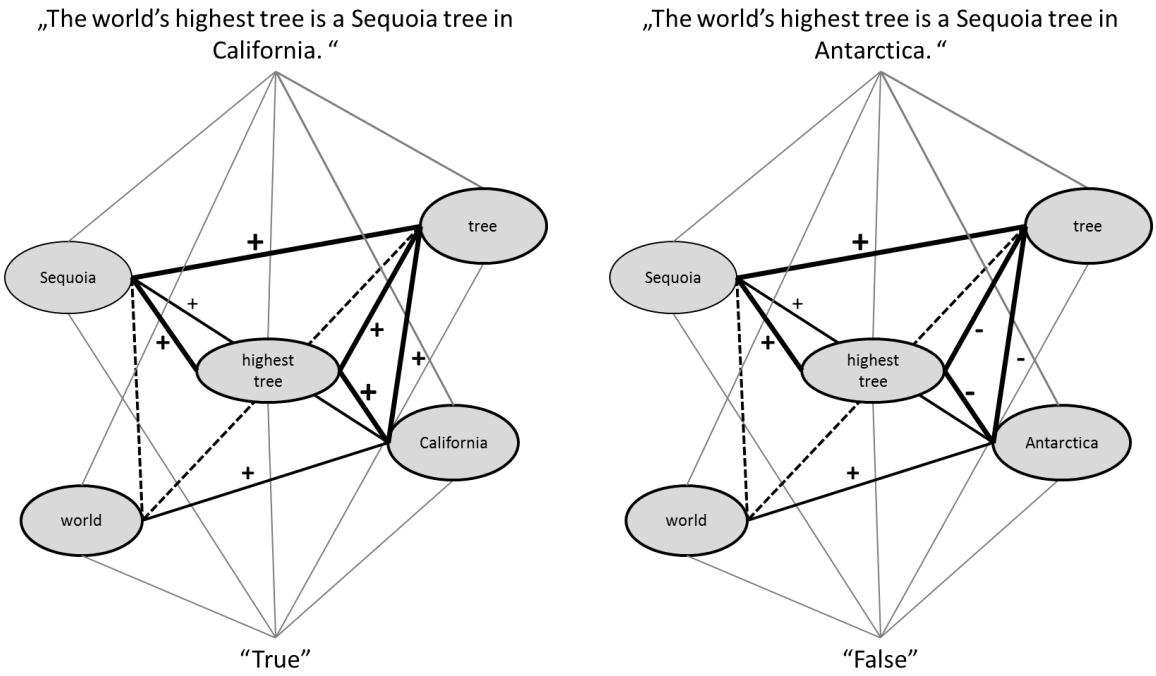

Figure 1. Illustration of how correspondence and coherence determine subjective truth. The solid grey lines indicate incoming information; here, a statement about trees. The grey circles ("tree") represent references in memory that provide meaning to the elements in the statement. Solid black lines indicate links between these references and the strength of the line indicates link strength. Dotted lines indicate links that are instigated by the incoming information. Finally, "plus" signs indicate an excitatory link and "minus" signs indicate inhibitory links. Coherence, defined as a parallel-constraint satisfaction solution (Kunda & Thagard, 1996), then defines the resulting subjective experience. If the statement's corresponding references form a coherent network, a "true" response follows, while an incoherent network of references results in a "false" response.

Figure 2

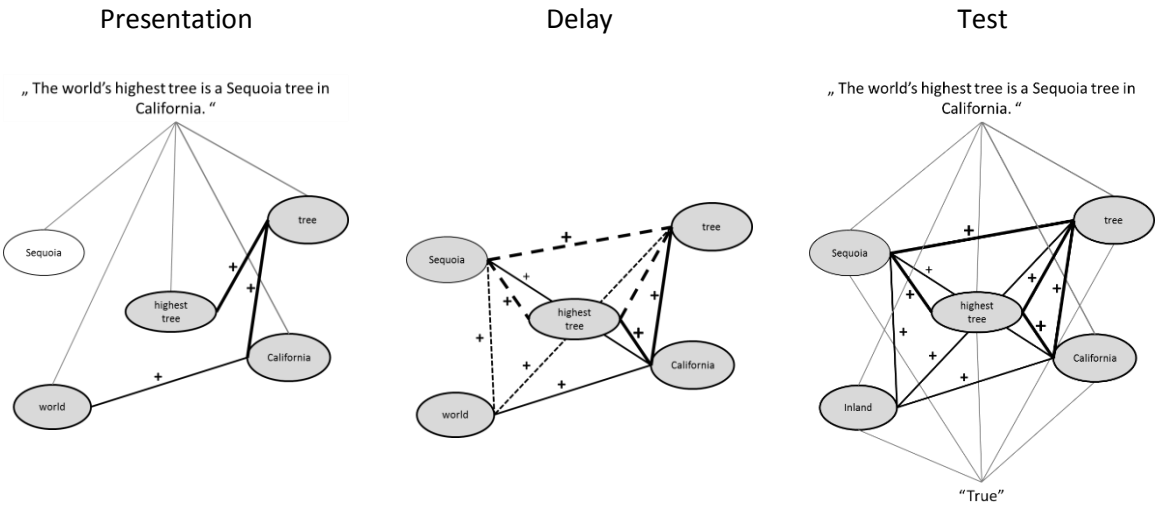

Figure 2. An illustration of the repetition-induced truth effect according to a referential explanation. The left panel shows the presentation phase when participants encounter a statement for the first time. If it is not incoherent with existing references (see Australia vs. Antarctica), the statement's corresponding references are linked within memory, as shown in the middle panel. At test, all of the repeated statements' elements are coherently linked, leading to a "true" judgment, as shown in the right panel. Novel statements at test are equivalent to the left panel. As the left panel has fewer coherently linked references compared to the right panel, a repetition-induced truth effect follows.