Social Cognition, Vol. 37, No. 2, 2019, pp. 145–173

ACTIVATION IS NOT ALWAYS INFERENCE: WORD-BASED PRIMING IN SPONTANEOUS TRAIT INFERENCES

Diana Orghian
Massachusetts Institute of Technology, MIT Media Lab; CICPSI, Faculdade de Psicologia, Universidade de Lisboa

Tânia Ramos
Faculdade de Psicologia, Universidade de Lisboa; New York University

Leonel Garcia-Marques
CICPSI, Faculdade de Psicologia, Universidade de Lisboa

James S. Uleman
New York University

People infer, without any intention or awareness, personality traits about actors enacting diagnostic behaviors. This phenomenon is known as spontaneous trait inferences (STIs). The activation of a trait is considered to be a true inference when it results from processing the meaning of the whole behavioral description. However, a trait can also become activated due to intra-lexical associations with individual words in the description. Here, we suggest a method to distinguish the two sources of activation and explore the influence that word-based priming has on some of the most popular paradigms used to study STIs. Results show that in the probe recognition task, word-based priming plays a considerable role and can, in the absence of an appropriate control, mimic spontaneous trait inference occurrence. However, in the false recognition task and in the explicit trait judgment task, the role of this spurious activation is negligible and the real trait inference can be easily detected.

Keywords: Spontaneous trait inference, word-based priming, delayed and immediate measures, lexical decision, probe recognition, false recognition

This research was supported by the Portuguese National Foundation for Science and Technology under the Grant DFRH-SFRH/BD/84668/2012.

Disclosure of interest: The authors report no conflicts of interest.

Address correspondence to D. Orghian, Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal; E-mail: dorghian@mit.edu.

INTRODUCTION

When we perceive someone's behaviors, we go far beyond their concrete actions and infer traits about their personality. Based on such inferences, we can predict what other behaviors to expect from people around us and we can also adjust our own behavior as a function of those predictions. Moreover, the research has shown that people infer traits from behavioral descriptions even in the absence of intentions to form impressions (Winter & Uleman, 1984), a phenomenon called spontaneous trait inference (STI). For example, if you find out that someone won the chess tournament, you will, without any such intention, infer that the person is intelligent.

Evidence of STIs' occurrence has accumulated at an impressive rate (for a review, see Skowronski, Carlston, & Hartnett, 2008; Uleman, Newman, & Moskowitz, 1996; Uleman, Rim, Adil Saribay, & Kressel, 2012). A large part of the research about the topic has been dedicated (i) to show that trait inferences occur during the encoding of the behavior (as opposed to its retrieval) and (ii) to explore the link that is established in memory between the inferred trait and the actor of the behavior (e.g., Carlston & Skowronski, 1994, 2005; Orghian, Garcia-Marques, Uleman, & Heinke, 2015; Orghian, Ramos, & Garcia-Marques, 2018; Todorov & Uleman, 2002). To investigate these issues, researchers have been using written behavioral descriptions as a proxy for observed behaviors. Written descriptions, because simpler, are easier to create and to control than other type of stimuli (videos or static representations of actions such as silhouettes). However, an important methodological issue has been ignored in most of the STI research using written behavioral descriptions. While the general assumption is that the trait is a sentence-based inference (i.e., is activated based on the meaning of the behavioral description as a whole), an alternative possibility is that the trait is being activated by specific words in the description through word-based priming mechanisms. Going back to the previous example, in the sentence "She won the chess tournament," individual words like "chess" and/or "won" can also prime the trait "intelligent," since they are semantically related to the trait. Importantly, this can happen independently of any trait inference being derived from the meaning of the entire sentence.

Word-based priming might be a common manifestation in trait inference studies because strong associates emerge naturally in our discourse as a reflection of the way we represent behavioral information. As a consequence, these associates are present in the behavioral material that researchers end up creating for their experiments. This means that previous results in this literature might, at least in part, reflect priming effects due to the presence of strong associates, which may have contributed to the final activation of the target traits. Since this process has not been systematically investigated and controlled for, the degree to which previous findings are contaminated by word-based priming remains unknown. The objective of this article is to investigate to what extent word-based priming plays a role in STIs and how it affects the various measures typically used to study this phenomenon and the reliability of the conclusions reached by using them. Ultimately, this article will be crucial to STI researchers, by clarifying a methodological issue that has long been recognized in the literature but has not been consistently investigated. It will also contribute to our knowledge of the cognitive processes producing STIs.

SPONTANEOUS TRAIT INFERENCES

Initial evidence for spontaneous trait inferences was provided by Winter and Uleman (1984) with a cued-recall task. In their studies, participants memorized trait-implying behaviors for a later recall test performed under different cueing conditions. Results showed that providing the implied trait as a cue improved the recall of the sentence compared to a no-cue condition, and that the trait was equally or more effective than strong semantic associates of words presented in the sentence. These findings were interpreted as evidence that participants had spontaneously inferred the trait during the comprehension of the behavior, and that the behavior and the corresponding trait had been encoded together in memory. As a consequence, when the trait was presented as a cue, it facilitated the retrieval of the behavior. In these studies, Winter and Uleman revealed some concern about how the presence of strong word associates in the sentences could distort their results. For instance, through a pretest, they only chose actor nouns that did not lead to the generation of the target traits. This was done with the purpose of preventing the trait (e.g., "thrifty") from serving as a retrieval cue for the actor noun (e.g., "businessman") at test, and indirectly facilitating the retrieval of the sentence (e.g., "The businessman thoroughly studied his investment in the business venture") without any inferences being made. However, in these studies, the authors did not control for associations between the trait and other words in the sentence (e.g., "venture" and "investment").

Since Winter and Uleman's original findings, other ingenious tasks have been developed to detect the occurrence of STIs. These tasks can be divided into two categories: immediate measures and delayed measures (for a review of the different paradigms see Orghian, Smith, Garcia-Marques, & Heinke, 2017). In immediate measures (also called activation measures), the trait inference is measured immediately after the participant reads the description, giving the researcher access to the inference as it happens (i.e., on-line). The underlying logic is straightforward: if the trait is inferred during the comprehension of the behavior, then the activation level of the trait will be high at the moment that immediately follows the reading of the sentence. The probe recognition task (McKoon & Ratcliff, 1986) is an example of an immediate measure that has been widely applied to study spontaneous trait inferences (Ham & Vonk, 2003; Newman, 1991, 1993; Ramos, Garcia-Marques, Hamilton, Ferreira, & Van Acker, 2012; Uleman, Newman, & Moskowitz, 1996; Van Overwalle, Drenth, & Marsman, 1999; Wang, Xia, & Yang, 2015; Wigboldus, Dijksterhuis, & Van Knippenberg, 2003; Wigboldus, Sherman, Franzese, & van Knippenberg, 2004). In this procedure, each behavioral sentence is followed by a probe word. Participants have to indicate, as fast as possible, whether the word was part of the previous sentence or not. In the critical trials, trait-implying sentences are followed by the implied traits. It is assumed that if the trait was inferred during reading, it will be harder for the participant to correctly indicate that it was not present in the sentence. Other immediate measures used to study STIs include the lexical decision task (Zárate, Uleman, & Voils, 2001) and the word-stem completion task (Whitney & Williams-Whitney, 1990). Naming and modified Stroop tasks are also immediate measures. However, so far, they have been mainly applied to investigate non-social inferences (e.g., Dosher & Corbett, 1982; Forster, 1981; Potts, Keenan, & Golding, 1988).

In delayed measures of STIs (also called memory measures) by contrast, there is an interval between the reading of the sentence and the measurement of the trait inference. The goal in these measures is to access a more stable representation of the inference. In order to perform a delayed memory task, participants need to access their long-term representation of the text or event. Two popular delayed measures are the savings in re-learning procedure (Carlston & Skowronski, 1994) and the false recognition task (Todorov & Uleman, 2002). In both of these tasks, there is an initial exposure phase during which the participant is exposed to a series of behavioral descriptions and photographs of people, which are said to be the actors of the behaviors. In the false recognition paradigm, during a later recognition test, it has been observed that participants falsely recognize the trait as having been presented in the sentence previously paired with the actor. In the saving in re-learning paradigm, there is usually a recall advantage for re-learning trait–actor pairs after exposure to the corresponding trait-implying sentence paired with the actor, when compared with new trait–actor pairs. These results are interpreted as showing that the trait was inferred after reading about the behavior and was linked to the mental representation of the actor. Several studies have provided convincing support for these assertions (Carlston, Skowronski, & Sparks, 1995; Crawford, Skowronski, Stiff, & Scherer, 2007; Todorov & Uleman, 2004).

Thus, the occurrence of STIs has been demonstrated across different laboratories and across a variety of tasks. However, very little is known about the text-processing mechanisms by which the trait is activated. Specifically, it is unclear whether the trait is being activated at the sentence level, at the word level, or both. The effect of word-based priming has been, however, widely acknowledged in the discourse comprehension field (e.g., Forster, 1981; Keenan & Jennings, 1995; Keenan, Potts, Golding, & Jennings, 1990; Sharkey & Sharkey, 1992), as discussed in the following section.

EFFECTS OF WORD-BASED PRIMING

Inference generation involves the integration of what is explicitly provided in the text (or observed) with prior general knowledge stored in the perceiver's memory. Since the majority of the studies that explore the occurrence of inferences use written descriptions of episodes, it is essential to understand how the information provided through text interacts with the prior knowledge of the reader. Keenan and colleagues (Keenan & Jennings, 1995; Keenan et al., 1990) proposed that, as a result of this interaction, a concept can be activated via two very different processes. One way is via sentence-based priming; that is, inferring the concept by considering the meaning of the text as a whole. An alternative way is through word-based priming mechanisms, that is, from reading individual words in the sentence. This type of activation is not generally considered a "true" inference, because it involves nothing more than passive spreading of intra-lexical activation, with no propositional construction or elaboration taking part. Inferring a personality trait from a behavioral description clearly involves applying a higher order type of processing to the provided information. However, although word-based and sentence-based primings are different processes, their effects can be indistinguishable. In fact, word-based priming effects may perfectly mimic sentence-based inferences, and ignoring the distinction between these two sources of activations may lead us to erroneous conclusions about the occurrence of trait inferences. Specifically, evidence of intra-lexical activations can be wrongly taken as evidence of spontaneous trait inferences.

Keenan and colleagues (1990) offered two alternatives to control for word-based priming effects. The first consists of eliminating any words in the sentence that are highly related to the target concept. However, as the authors acknowledged, this is an extremely difficult task. The meaning of the sentence as a whole must derive, to some extent, from the meaning of the individual words in the sentence and thus, we should always expect some amount of association between the meaning of individual words and the trait. An alternative suggested by the authors is to create appropriate controls. The control sentences should contain the same words as the implying sentences, but the words should be rearranged in such a way that the meaning of the resulting sentences changes and no longer implies the concepts of interest. Because the two sentences are equated in terms of individual words, if the activation of the target trait is higher following the encoding of the trait-implying sentence than following the encoding of the rearranged version, then the activation can safely be attributed to sentence-based inferences.

This solution has been implemented in the discourse comprehension field (Calvo, Castillo, & Schmalhofer, 2006; McKoon & Ratcliff, 1986; Otten & Van Berkum, 2009). A representative example is McKoon and Ratcliff's work (1986) on predictive inferences. In their studies, a target word (e.g., "dead") was presented immediately after a predictive sentence (e.g., "The director and the cameraman were ready to shoot close-ups when suddenly the actress fell from the 14th story") or after a rearranged sentence that contained roughly the same words as the predictive sentence (e.g., "Suddenly the director fell upon the cameraman, demanding close-ups of the actress on the 14th story"). When asked to indicate whether the target word was part of the sentence, participants had more difficulty in saying "no" after reading the predictive sentence than after reading the rearranged sentence. Since word-based priming effects are being controlled for, these discrepancies were attributed to sentence-based inferences.

We are not the only ones concerned with word-based priming in the context of spontaneous trait inferences. Newman (Newman, 1991, 1993) showed concern about word-based priming and used rearranged sentences in several of his studies.

However, the differences between trait-implying and rearranged sentences were not always significant in the expected direction, making it difficult to conclude that real inferences were taking place. Uleman, Hon, Roman, and Moskowitz (1996), also inspired by McKoon and Ratcliff's work, implemented the rearranged controls solution in their studies. In their work, participants were presented with trait probes (e.g., "lazy") after reading a trait-implying sentence (e.g., "He drove to the newsstand, only half a block away") or after reading a rearranged version containing almost the same words (e.g., "He drove to the only newsstand, 20 blocks away"). In their studies, expected differences in the reaction times for rejecting the probes after trait-implying and rearranged sentences were significant in the participants' early trials, but disappeared later with practice. To our knowledge, no other articles in the spontaneous trait inference literature have included rearranged sentences to control for word-based priming (an exception is our recent article, Orghian et al., 2017).

Here, we investigate the effect of word-based priming on four different paradigms commonly used in the literature. We tried to cover different categories of paradigms: an explicit trait rating task, two immediate tasks (the lexical decision and the probe recognition paradigm), and one delayed task (the false recognition paradigm).

OVERVIEW OF EXPERIMENTS

To conduct this investigation, we first created a collection of sentences that imply personality traits (referred to as "trait-implying sentences" from now on). Based on these sentences, we created control sentences containing approximately the same words as the trait-implying ones (e.g., "She won the chess tournament") but rearranged in such a way that they did not imply the same traits anymore (e.g., "She won a ticket to a chess tournament"). This method of rearranging the words guarantees that the effect of word-based priming is comparable in the trait-implying and the control sentences, since the words in the two sentences are similar. Any observed difference in the likelihood of inferring the target trait from the two sentences can only be attributed to sentence-based inferences. We will refer to these control materials as "rearranged sentences."

Here, we assume that activation needs to reach a threshold for the trait to become active. In other words, the trait is assumed to be either active or not. When a trait-implying sentence leads to a sentence-based inference and simultaneously contains words that are strongly related with the target trait, the trait is expected to be activated above the threshold in both trait-implying (either through sentence-based priming or word-based priming, we would not know) and rearranged sentences (through word-based priming). The only situation that leads to different states of activation in the trait-implying and rearranged sentences is when the sentences do not contain strong associates but do lead to sentence-based inferences. When that happens, the critical trait becomes activated in the trait-implying sentence (due to sentence-based inference) and does not in the rearranged one since there are no strong associates (or the existent activation coming from the words in the sentence does not reach the threshold). This is the only situation where sentence-based inferences can be detected with the guarantee that they are not being confounded with word-based inferences.

With this framework in mind, in Experiment 1, we used a lexical decision task to measure the level of activation of target traits after reading trait-implying and rearranged control sentences. Based on the difference in the activation, we divided the stimuli sentences into two sets. In one set, the lexical decision for trait-words was faster after reading the trait-implying sentences than after reading the rearranged sentences. Given that the words are very similar in both sentences, the additional facilitation observed after reading the trait-implying behaviors was interpreted as evidence of sentence-based inferences. The pairs in this set were labeled as the strong pairs, since the evidence for sentence-based inferences in the trait-implying sentences in these pairs is strong. In contrast, for the second set of pairs of sentences, the trait-implying and the rearranged sentences led to a similar amount of facilitation in the lexical decision task. In these pairs, labeled weak pairs, the evidence for sentence-based priming is weak, because no activation is detected beyond that of the word-based activation, meaning that this activation masks the detection of eventual sentence-level inferences.

These two sets of materials were then used in three additional experiments. In Experiment 2, participants performed a trait rating task. The goal was to compare intentional inferences of target traits in strong and weak pairs of sentences when participants have the explicit goal of inferring traits. Note that our claim about the effect of word-based priming concerns only STIs. Because we don't believe that word-based passive activation can override explicit judgements about personality traits, we don't expect a difference between strong and weak trait-implying sentences in this task. This hypothesis is in agreement with the literature on implicit and explicit measures that shows that implicit measures are not necessarily reflected in explicit judgements (e.g., Gawronski & Bodenhausen, 2006; Gawronski & LeBel, 2008). Moreover, Experiment 2 should provide evidence that both weak and strong sentences communicate personality traits to a similar extent, so that any differences that we might find between the two in the following studies cannot be due to one set being better at communicating the traits than the other.

In the last two experiments, our hypothesis was that immediate and delayed measures would be differently affected by word-based priming. Immediately after the encoding of trait-implying and rearranged sentences, we expect a large overlap of activation between the two (mostly due to word-based priming). This overlap may be enough to cloak the extra activation coming from the sentence-based inference taking place in the trait-implying descriptions. Later on, after a delay in which a more stable representation of the event is created and after spurious activations are inhibited, the overlap between the two sentences should decrease, which should make the real trait inference easier to detect. According to the Construction-Integration model (Kintsch, 1988) popular in the text comprehension literature, inferences from words are indiscriminate, multiple, and context-free right after encoding. But during a later integration phase (where the multiple activations are constrained by the surrounding context), they stabilize and only relevant concepts are included in the final representation of the text. (For more on this see the General Discussion.)

In Experiment 3, we used the probe recognition task—an immediate measure of STIs. If, as we predict, word-based priming plays a role in this measure, then different patterns are expected for trait-implying sentences from weak and strong pairs. Difference between trait-implying and rearranged should be detected only for the strong pairs. Finally, in Experiment 4, we used the false recognition task—a delayed measure of STIs. If the effects of word-based priming are inhibited or edited out during the delay, then the trait inference effect should be observed for both strong and weak pairs.

EXPERIMENT 1

METHOD

Participants. One hundred and eight undergraduate students participated in this experiment (19 males). The average age of the sample was 19.95 years old. The sample size was arbitrarily determined based on another experiment that was being conducted in the same sessions, which had a requirement of 100 valid participants minimum. The data from all the experiments reported in this article were only analyzed once the reported samples were complete, no additional data being collected afterwards.

Material. Forty-eight pairs of trait-implying sentences and their rearranged versions were selected from a previous pretest conducted by Orghian and colleagues (Orghian, Ramos, Reis, & Garcia-Marques, 2018). All stimuli used in the current experiments can be acquired here: https://osf.io/cbefw. The rearranged sentences were created in such a way that they preserved as many words from the trait-implying sentences as possible, but did not imply the target traits. Note that the rearranged sentences were controlled for the activation of specific traits, which does not mean that other traits are not implied by them. Our concern was to create sentences that do not imply traits that would be relevant or related to the critical traits implied in the trait-implying sentences. This includes avoiding negation and synonymy. One might think that the negation of the trait-implying sentence would be a good control sentence. However, the negation of the trait-implying sentence (e.g., "She didn't win the chess tournament") might still lead to scenarios that imply the trait (in our example the negation still means the person took part in the tournament, which might still lead to the inference of the trait "intelligent"). Moreover, by negating something, the perceiver might get the impression that the standard is to perform the action, and as such imply the exact opposite trait, which is still not ideal because antonyms are strongly related in our mind (Gross, Fischer, & Miller, 1989).

We also had 96 neutral filler sentences that did not imply any trait (e.g., "Bananas and mangoes are important sources of potassium and magnesium"). None of the neutral filler sentences conveyed behavioral information. These sentences were included to disguise the goal of investigating trait inferences from behavioral descriptions. The 48 traits implied by the trait-implying behavioral sentences served as probes in the lexical decision task. Additionally, we also used 24 non-trait words (e.g., "sword") and 72 non-words taken from Domingos and Garcia-Marques's norms for Portuguese non-words (2013).

PROCEDURE

All the experiments in this article were built and conducted using OpenSesame, a free and open source graphic experiment builder (Mathôt, Schreij, & Theeuwes, 2012). With the exception of Experiment 3, participants were recruited with the ORSEE tool (Greiner, 2015).

Participants completed Experiment 1 in groups of up to 8 people. As a cover story, they were told that the study was investigating multitasking capabilities. Specifically, they were told that they had to memorize sentences and to perform an unrelated lexical decision task simultaneously. Thus, immediately after reading each sentence, a string of letters was presented and their task was to decide, as quickly and accurately as possible, whether it was a word or a non-word. Each trial started with a 500-millisecond fixation cross, followed by the presentation of the sentence, which remained on the screen until the participant finished reading the sentence and pressed the space bar. After pressing the space bar, a string of letters was presented for lexical decision, and it remained on the screen until the participant answered. Participants had to press the letter "b" on the keyboard to indicate that the target was a word and the letter "n" to indicate that the target wasn't a word. Participants started with 6 practice trials followed by 144 experimental trials.

In 24 of the 144 trials, participants memorized trait-implying sentences and in another 24 they memorized rearranged sentences. These 48 trials were followed by the presentation of the corresponding traits. There were also 24 trials where neutral sentences were followed by non-trait words and 72 trials where neutral sentences were followed by non-words. Note that for half of the trials the correct response was "yes" (for 48 traits plus 24 non-trait words), and for the other half of the trials the correct response was "no" because the letter string was not a word (72 non-words). Each sentence and each string of letters appeared only once for each participant. The order of the trials and the pairing between the neutral sentences and the strings were random for each participant.

In half of the 48 trials in which the probe was a trait, the sentences implied the probed trait and for the other half the sentences were rearranged versions of non-presented trait-implying sentences. However, the same participant never saw the two sentences (the trait-implying and the rearranged) corresponding to the same trait. In order to do this, two versions of the experiment were created, and each participant was randomly assigned to one of them. In the first version, half of the traits (set 1) was preceded by sentences implying those traits, while in the other half (set 2) they were preceded by rearranged sentences. In the second version, the traits from set 1 were preceded by the corresponding rearranged sentences, while the traits from set 2 were preceded by sentences implying those traits. The assignment of the sentences to set 1 or 2 was performed randomly. The total experiment took approximately 15 minutes.

RESULTS AND DISCUSSION

Only the trials where the targets for the lexical decision task were traits were analyzed. To assure than only the motivated people were included in the analysis, and because the correct implementation of the following experiments relies on the result of this experiment, we trimmed the least accurate 5% of the participants (5 participants). The average accuracy of these 5 participants was 77.5%, whereas the accuracy of the remaining sample was 96.17%. Only the reaction times of the correct responses were considered in the following calculations. Moreover, to include only participants' most immediate responses, reaction times larger than 2500 ms were eliminated (1.44% of the correct responses). Because this study was conducted with the purpose of distinguishing strong pairs of sentences (those that show evidence of sentence-based inferences) from weak ones (those that show only word-based trait priming), all the analyses reported in the article were conducted per items and not per participant.

The mean reaction time for each trait, in each of the two conditions (when preceded by a trait-implying and when preceded by rearranged sentences) was computed. Next, the difference between the two versions was calculated for each trait (trait-implying minus the rearranged; see Appendix A for the statistics of each pair). By using the median split of these differences, two groups were obtained: a group with the most positive differences (ranging from 44 to -15 ms), that is, a group in which the trait-implying sentences did not lead to an RT reduction when compared with the rearranged versions, and a second group with the most negative differences (ranging from -16 to -169 ms), that is, a group in which an RT reduction for the trait-implying sentences was detected when compared with their rearranged versions. The first group is referred to as weak and the second as strong. In the strong pairs group, the reaction times to the trait-implying sentences (M = 807.40, SD = 67.91) are significantly shorter than to the rearranged sentences (M = 866.59, SD = 77.18), $t(23) = 7.26$, $p < .001$. The opposite effect was found for the weak group, with larger reaction times in the trait-implying (M = 835.71, SD = 62.73) than in the rearranged condition (M = 825.98, SD = 57.07), $t(23) = -3.18$, $p = .004$. As discussed in the introduction section, we think that the weak pairs do not work as expected (i.e., greater facilitation for the trait-implying than for the rearranged version) because word-based priming effects are overriding the trait inference effect. For strong pairs, the trait inference is detected regardless of intralexical priming effects.

Moreover, one could argue that the difference between groups is driven by the difference in the inferential power (sentence-based trait inferences) between trait-implying strong and trait-implying weak sentences. However, the reaction time in the strong group (trait-implying condition: M = 807.40, SD = 67.91) does not differ significantly from the reaction time in the weak group (trait-implying condition: M = 835.71, SD = 62.73), $t(23) = 1.50$, $p = .140$, which makes this explanation less plausible.

To verify the robustness of our conclusions, besides the median split, we performed a different split based on terciles. The analysis and the graphs for the three groups obtained are presented in Appendix B1.

Next, we explored whether weak and strong sentences differ in other aspects. To compare the sentence length, we counted the number of words in the weak and the strong pairs of sentences. We found no significant difference between the length of the rearranged (M = 13.08, SD = 2.42) and the length of the trait-implying sentences (M = 13.10, SD = 2.78), $t(47) = -.07$, $p = .947$. The length of rearranged sentences in the weak set (M = 13.67, SD = 2.32) was marginally larger than the length of rearranged sentences in the strong set (M = 12.50, SD = 2.43), $t(47) = 1.70$, $p = .095$. For trait-implying sentences, the difference in length between the weak set (M = 13.33, SD = 3.19) and the strong set (M = 12.88, SD = 2.36) was not significant $t(47) = .57$, $p = .574$.

We also pretested these sentences on how easy they were to comprehend. Four judges, unfamiliar with the goals of our research, indicated how easy it was to comprehend each sentence on a 9-point scale (1–not easy at all, 9–very easy). The weak (M = 8.70, SD = .82) and strong trait-implying sentences (M = 8.69, SD = .55) were similar in terms of ease of comprehension, $t(23) = 0.005$, $p = .996$. A similar result was observed for rearranged weak (M = 7.13, SD = 2.11) and for rearranged strong sentences (M = 7.11, SD = 1.78), $t(23) = 0.031$, $p = .975$. Not surprisingly, the comprehension ratings were much lower for rearranged (M = 7.12, SD = 1.95) than trait-implying sentences (M = 8.70, SD = 0.71), $t(41) = -4.565$, $p < .001$. Finally, two different judges were asked to write down the first word that came to their mind for each rearranged sentence. For these studies, we only chose the sentences for which the judges did not generate any words related to the critical traits. Thus the weak and the strong rearranged sentences did not vary in these generated responses.[1] For more about these pretests see Orghian, Ramos, Reis, and Garcia-Marques (2018).

These two groups of sentences were used in the following experiments in order to examine to what extent word-based priming effects influence the results in the paradigms typically used in the spontaneous trait inference literature.

EXPERIMENT 2

In our second experiment, we used an explicit trait judgement task where the participants were asked to judge how much the target trait described the actor of the behavior. This experiment tests whether the trait is equally implied by the trait-implying sentences in the weak and the strong groups, when the participants are asked to explicitly infer the target traits. Critically, we do not expect the two

1. Only 43 of the 48 pairs of sentences were pretested.

groups of sentences to perform differently, because word-based activation is not expected to impact explicit intentional judgements. Our concern with word-based priming effects is only related to implicit measures. In the case of intentional inferences, the influence of word-based priming is expected to be minimal, since people are consciously analyzing the sentence as a whole with an explicit goal of inferring a personality trait.

METHOD

Participants. Fifty-eight undergraduate students (12 male) took part in this study. The sample's average age was 22 years old. The sample size in this experiment was defined based on an arbitrary criterion, the total number of show-ups in a week.

Materials. The same 48 critical pairs of trait-implying and rearranged sentences (24 weak pairs and 24 strong pairs) tested in Experiment 1 and their corresponding traits were used in this experiment. No neutral sentences were used because there was no need to disguise the purpose of the study to participants.

PROCEDURE

The experiment was introduced to the participants as a pilot study to test material for future experiments. There were 48 trials and each trial started with a fixation cross presented for 500 ms, followed by the presentation of a sentence. On the same screen, the trait was presented with a rating scale. The participants' task was to indicate the extent to which the actor described in the sentences possessed the trait, on the scale ranging from 1 (not at all) to 9 (totally). Again, we had two different versions of the experiment, such that only one of the sentences in each pair was presented to each participant (either the trait-implying or the rearranged version).

RESULTS AND DISCUSSION

A main effect of type of sentence was found, $F(1, 46) = 194.63$, $p < .001$, $\eta_p^2 = .81$, with higher ratings in the trait-implying condition ($M = 7.73$, $SD = .72$) than in the rearranged ($M = 4.28$, $SD = 1.44$). No other effects were observed, and critically as it can be seen in Figure 1, there is no difference between the weak and strong material, $F < 1$. These results suggest that word-based priming has no role in this intentional trait-rating task (see Appendix B2 and Figure B1 for the results based on tercile split).

EXPERIMENT 3

In Experiment 3, we investigated how the strong and the weak groups of sentences perform in an immediate measure frequently used in the STI literature—the probe

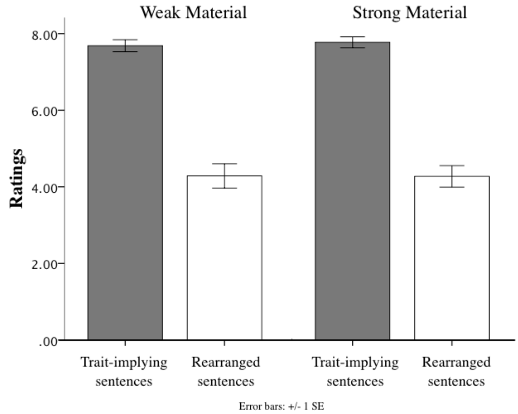

FIGURE 1. Average ratings as a function of type of material (weak vs. strong pairs of sentences) and type of sentence (trait-implying vs. rearranged) in Experiment 2.

recognition task. If our hypothesis that word-based activation influences immediate measures is true, then we should be able to observe a pattern similar to the one that was found in the lexical decision task (Experiment 1), that is, a greater activation of the trait after reading the trait-implying sentences than after the rearranged sentences, but only for the strong group of sentences.

In the recognition probe task (Ham & Vonk, 2003; Newman, 1991; Uleman, Hon et al., 1996; Van Overwalle et al., 1999; Wigboldus et al., 2003, 2004), participants read a sentence and immediately afterwards they are presented with a probe word. Their task is to indicate, as quickly and accurately as possible, whether the probe word was presented in the sentence previously seen. This is an immediate measure of trait inference because the trait inference is tested immediately after the reading of the sentence. Note that in this paradigm, larger reaction times are associated with a greater activation of the trait probe (contrary to what happens in the lexical decision task where shorter reaction times are associated with more activated traits). This means that if the probe is being activated by the sentence, then it should be more difficult to indicate that the probe word was not part of the sentence and thus, it should take longer for the participants to provide a correct response. As such, we predict that participants will respond slower to trait probes after reading a sentence that implies the trait than after reading its control version, which would indicate that a text-based trait inference was generated. However, for our hypothesis to be supported, this pattern should be observed only for the strong pairs of sentences and not for the weak pairs.

METHOD

Participants. Ninety-four undergraduate students (32 males) took part in the study. Their average age was 23 years old. Using the G*Power 3.1 tool (Faul, Erdfelder, Lang, & Buchner, 2007), for a medium partial eta-squared of .06 and a power of .80, a minimum of 24 participants is required to detect the interaction between the type of material (strong versus weak) and the type of sentence (trait-implying vs. rearranged). For a small effect size of .01 (.80 power), 138 participants are required. Since we don't know the size of the effect, we collected a sample size somewhere in between the two suggested samples, taking into account the number of available participants in the lab at the time the experiment was conducted.

Materials. The same 48 target traits and the corresponding pairs of sentences from previous experiments were used in this experiment. Twenty-four of these pairs correspond to the strong group and 24 correspond to the weak group. Additionally, 56 filler trials consisting of neutral sentences and 56 non-trait words were used. The total number of trials was 104 (48 critical plus 56 fillers).

PROCEDURE

Participants were told that this study had the objective of investigating working memory abilities. Initially, participants completed 8 practice trials that were followed by 104 experimental trials. Each trial started with a centered fixation dot presented for 700 ms, followed by the presentation of a sentence. Participants were instructed to press the space bar when they had finished reading the sentence. After they press the space bar, a 100-ms blank screen followed. Next, the probe was presented and remained on the screen until a keyboard response was detected. Participants were instructed to press the letter "q" on the keyboard if the answer was "no" (i.e., to indicate that the word was not part of the previous sentence) and to press "w" if the answer was "yes" (i.e., to indicate that the word was in the previous sentence). In 32 of the filler trials, the probe was actually presented in the sentence in order to prevent the correct response from always being "no." In the remaining trials (24 filler trials plus the 48 trait trials), the probe was not part of the sentence. As in the previous experiments, there were two versions of the experiment and each participant was randomly assigned to one of these versions. If a trait was preceded by its trait-implying sentence in one version, it was preceded by its rearranged pair in the other version, and vice versa. In each version, half of the critical sentences were from the strong group and the other half from the weak group. As before, the type of sentence (trait-implying versus rearranged) was manipulated within-item, while the type of material (strong versus weak) between-item.

RESULTS AND DISCUSSION

The overall accuracy rate was 91%. Since there were no significant effects for accuracy, this dependent variable will not be mentioned further. Only the reaction times of correct responses are considered. Responses with reaction times lower that 250 ms or larger than 2500 ms were also excluded (less than 1% of the data) from the analyses. A mixed effects ANOVA was conducted, with the average reac-

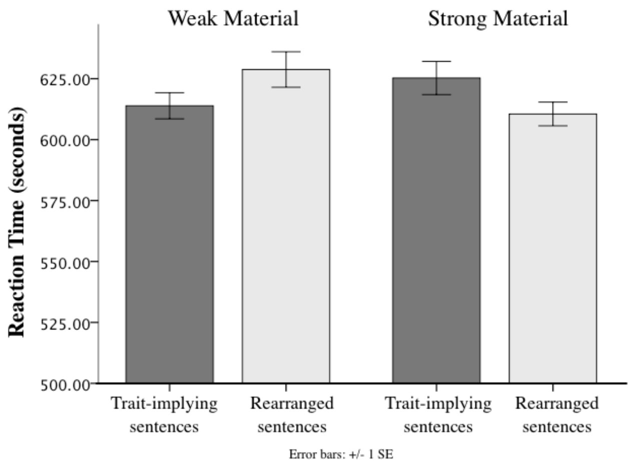

FIGURE 2. Mean reaction time as a function of type of material (weak vs. strong pairs of sentences) and type of sentence (trait-implying vs. rearranged) in Experiment 3.

tion time per item being the dependent variable and the type of sentence (trait-implying sentence vs. rearranged sentence) and the type of material (strong pairs vs. weak pairs) being the independent variables. A significant interaction was observed between the two factors, $F(1, 46) = 11.90$, $p = .001$, $\eta_p^2 = .21$.

For the strong pairs, longer reaction times were observed for the trait-implying sentences ($M = 625$, $SD = 34$) than for the rearranged sentences ($M = 611$, $SD = 24$), $F(1, 23) = 8.11$, $p = .009$, $\eta_p^2 = .25$. This result indicates that it was more difficult for the participants to reject the trait probe when it was preceded by a sentence that implied that trait than when it was preceded by a rearranged version. Thus, we found evidence for the occurrence of spontaneous trait inferences in the strong pairs condition. For the weak pairs, however, there was no evidence of spontaneous trait inferences. In fact, an opposite pattern was observed, with longer reaction times for the rearranged versions ($M = 629$, $SD = 35$) than for the trait-implying sentences ($M = 614$, $SD = 26$), $F(1, 23) = 4.57$, $p = .044$, $\eta_p^2 = .17$ (see Figure 2).

We do not have a clear explanation for why the pattern reversed for the weak pairs but it might be related to the extra analytical thinking required by the trait-implying sentences; this may have interfered with the intra-lexical activations. A different but related explanation has to do with how traits are being activated in the trait-implying sentence. It is very likely that we don't activate only one trait per sentence but activate a semantic area (or multiple closely related traits), while the word-based priming only activates a unique trait. Thus, there might be more competition among the multiple traits activated by the trait-implying sentence than in the word-based priming. Finally, there is the fact that the rearranged sentences are more difficult to understand or less natural than the trait-implying ones (for a similar argument see Newman, 1993). However, this possibility does not explain why the same is not happening in the strong material.

In sum, as predicted, we found a strong interaction between the type of material and the type of sentence, with more trait interference in the trait-implying condition than in the rearranged sentences for the strong pairs, but not for the weak pairs.

In Appendix B3 and Figure B2 we present additional analyses where the material is divided in groups based on terciles.

EXPERIMENT 4

In Experiment 4, we used the same material as in the previous experiments, but participants were asked to perform a delayed memory task—the false recognition task (Goren & Todorov, 2009; Rim, Min, Uleman, Chartrand, & Carlston, 2013; Todorov & Uleman, 2002, 2003, 2004). Typically in this task, participants see pairs of sentences and photographs of faces of people under memory instruction. Each sentence is said to describe a behavior performed by the person in the photograph. In a delayed test phase, participants are presented with the same photographs, and each photograph is paired with a trait word. The task is to indicate whether the word was present or not in the sentence that was paired with that photograph in the study phase. In some trials, the trait presented is the one that was implied by the sentence paired with that photograph (match trials), whereas in others, the trait presented was implied by a behavior previously paired with some other person (mismatch trials). It is assumed that if the trait is inferred from the behavioral sentence and linked to the actor of that behavior, then participants will exhibit a larger rate of false recognitions in the match condition than in the mismatch condition. In the mismatch condition a lower rate of false recognitions should be verified since no link was created between the presented trait and the person in the photograph during the first part of the experiment.

As in the previous experiments, the type of sentences was manipulated (trait-implying sentences vs. rearranged sentences) and the type of material (strong pairs vs. weak pairs). If the influence of word-based priming is only a problem for immediate measures, then the type of material should not make a difference in this paradigm. That is, we should observe evidence of STIs for both strong and weak material, contrary to the results found in Experiment 3. However, if the word-based priming lasts long enough to affect the delayed test phase, the pattern of results should be similar to the one in the previous experiment, that is, STIs will only be detected for the strong pairs.

METHOD

Participants. Twenty-nine students (9 males) took part in the experiment. The average age was 17 years. The participants were high school students visiting the psychology department, and thus, the sample size was defined by the number of visitors. Consent from the participants' parents was obtained. One may argue that the sample size of this experiment is small, but, the effects in the false recognition paradigm are usually strong. For example, in the studies reported by Todorov and Uleman (2003), r varied from .72 to .87 and the sample size varied from 27 to 38 participants.

Materials. The material consists of the same 48 pairs of sentences and their corresponding 48 traits used in the previous experiments. Additionally, 24 trait-implying sentences were used as fillers. In the filler sentences, the trait implied by the behavior was presented in the sentence (e.g., "She is so superstitious that she knocked on wood after mentioning that something bad could happen"). These sentences were only included to ensure that the correct response to the question made at test (i.e., whether the word was part of the sentence) wasn't always "no." Seventy-two photographs of faces of people with neutral expressions were paired with these sentences.

PROCEDURE

The experiment was conducted in the laboratory and each session had a maximum of 8 participants. The experiment was introduced as a memory study. Participants were told that the experiment consisted of two phases, a first phase where they would have to memorize pairs of photographs of people and sentences with information about those people. The second phase was said to be a memory test of the material learned during the first phase (the type of test was not specified). Each trial started with a 500-ms fixation cross, followed by a photograph presented in the middle of the screen and a sentence displayed below the photograph. Each pair was presented on the screen for 8 seconds. The study phase started with 4 practice trials. For each participant, our experimental software randomly chose 24 rearranged and 24 trait-implying sentences from the 48 critical pairs, with the only criteria being that two versions of the same trait could not be presented to the same participant. There were also 24 filler trials, in which the trait was actually presented in the sentence. Overall the first phase consisted of 72 experimental trials. The order of the trials and the pairing between the sentence and the photograph were randomized for each participant. After this phase, participants completed a 3-minute distracting task, followed by the test phase. In the test, they were presented with the previous photographs, each one paired with a trait. Their task was to indicate whether the word was part of the sentence presented in the study phase with that person. They had to press letter "s" to indicate "yes" and letter "l" to indicate "no." Each trial in the test phase started with a 500-ms fixation cross, followed by the presentation of the trait and the photograph, which remained on the screen until a response was given. After a response was given, a 500-ms blank screen followed. This phase started with 4 practice trials, followed by 72 experimental trials. There were two critical types of trials in the test phase: the match trials, where the trait presented at test was implied in the sentence presented with that photograph during study phase, and the mismatched trials, where the trait presented was implied in a sentence but the sentence was previously paired with a different person. From the 48 target trials, half (24) were paired, in the study phase, with rearranged sentences and half with trait-implying sentences. Within each of these two sets of trials, 12 were match and 12 were mismatch trials. The remaining 24 trials corresponded to the fillers. In the fillers, the trait was actually presented in the sentence, so the correct response for these was "yes."

RESULTS AND DISCUSSION

Only the target trials were included in the following analyses. Again the material was analyses in accordance with the two groups created in Experiment 1. (See Appendix B4 and Figure B3 for the results based on terciles split.) A mixed effects ANOVA was conducted with the false recognition rate as the dependent variable. The type of sentence (implying sentence vs. rearranged sentences) and the pairing (match vs. mismatch) were the within-item independent variables, and the type of material (strong group vs. weak group) was the only between-item independent variable. A main effect of type of sentence was found, $F(1, 46) = 9.81$, $p = .003$, $\eta_p^2 = .18$, with a higher false recognition rate for the trait-implying sentences ($M = .37$, $SD = .19$) that for the rearranged versions ($M = .27$, $SD = .15$), indicating that the traits were more activated by the trait-implying sentences than by their control versions. A main effect of pairing was also observed, $F(1, 46) = 9.03$, $p = .004$, $\eta_p^2 = .16$, with more false recognitions in the match condition ($M = .36$, $SD = .18$) than in the mismatch condition ($M = .28$, $SD = .12$), meaning the trait was inferred and linked to the person presented in the photograph. A marginal interaction between pairing and type of sentence was also found, $F(1, 46) = 3.70$, $p = .061$, $\eta_p^2 = .08$, with more false recognitions observed in the match condition ($M = .43$, $SD = .26$) than in the mismatch condition ($M = .30$, $SD = .21$) for trait-implying sentences, $F(1, 47) = 9.58$, $p = .003$, $\eta_p^2 = .17$. The same was not true for the rearranged sentences, since the rate of false recognitions in the match condition ($M = .28$, $SD = .21$) did not differ significantly from the rate of false recognitions in the mismatch condition ($M = .26$, $SD = .16$), as it can be seen in Figure 3, $F < 1$. Moreover, there is no significant three-way interaction, $F < 1$, meaning that this pattern was similar for the weak and the strong sentences. This result suggests that, unlike the probe recognition task, the false recognition task is not affected by word-based priming. In other words, independently of the type of material, we did find evidence of spontaneous trait inference, indicating that the trait was inferred from the sentence and linked to the actor for the trait-implying sentences and not for the rearranged ones.

One might say that our conclusion is based on a null effect (the lack of three-way interaction). However, while it is not surprising that people make inferences for the strong trait-implying sentences, it is very informative that they also make inferences for the weak trait-implying sentences. This is a non-null effect that strongly supports our hypothesis. More precisely, for the trait-implying condition, in the strong group, a significant inference effect was found with more false recognitions in the match condition ($M = .41$, $SD = .28$) than in the mismatch condition ($M =$

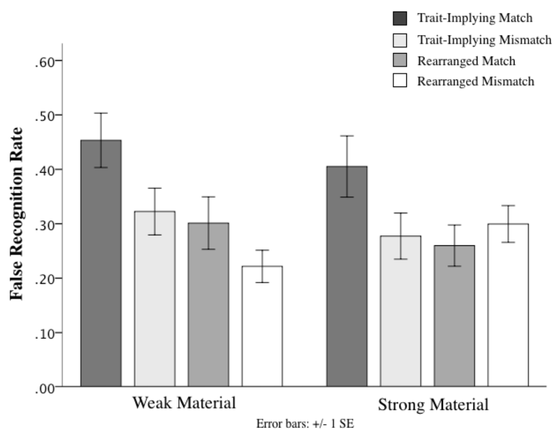

FIGURE 3. False recognition rate as a function of type of material (weak vs. strong pairs of sentences), type of sentences (trait-implying vs. rearranged), and pairing (match and mismatch) in Experiment 4.

.27, *SD* = .21), $F(1, 23) = 4.40$, $p = .047$, $\eta_p^2 = .16$. The same pattern was found for the trait-implying condition in the weak group, with more false recognition in the match condition (*M* = .45, *SD* = .25) than in the mismatch condition (*M* = .32, *SD* = .21), $F(1, 23) = 5.01$, $p = .035$, $\eta p^2 = .18$. When the same comparisons are conducted for the rearranged conditions, no inference is found in the strong group, $F < 1$, with similar amount of false recognitions in the match condition (*M* = .26, *SD* = .19) and in the mismatch condition (*M* = .30, *SD* = .17). This same difference did not reach significance in the weak group, $F(1, 23) = 3.02$, $p = .095$, $\eta_p^2 = .12$ (match condition: *M* = .30, *SD* = .24, mismatch condition: *M* = .22, *SD* = .15).

GENERAL DISCUSSION

The main goal of the present research was to examine the impact of word-based activation on various measures of STIs. In order to examine STIs occurrence, researchers usually create behavioral descriptions in the form of sentences or paragraphs. In the current article, we investigated the two ways through which a trait can be activated during the reading of these materials. One is based on pure associations between specific words in the sentence and the target trait to be inferred, and the second is by processing the sentence as a whole. Only the second type can be considered a real inference, since the implied trait results from a combination of the perceiver's knowledge and the global meaning of the sentence (i.e., more than the sum of the meanings of the individual words). Thus, in a sentence, when both word-based priming and sentence-based inference lead to the activation of the same trait, if appropriate controls are not used, it is impossible to tear these processes apart. A study that does not take into account the existence of these two sources of activation has no guarantee that the research is actually about trait inferences. It might be the case that what is being detected is no more than inter-lexical activations. To overcome this limitation, we strongly suggest the use of appropriate control materials. The controls that we propose and implemented in the current studies were created by rearranging the words from the trait-implying sentences in such a way that the newly created sentences do not imply the target traits anymore. By contrasting the activation of the target trait in the trait-implying and in the rearranged sentences, it is possible to determine whether the effect is driven by a real inference or by word-based priming.

We don't see the word-based priming as a serious threat to the interpretation of all the previous results reported in the spontaneous trait inferences literature, especially because it is highly unlikely that it would account for the large amount of converging evidence. We do think, however, that this is an important aspect to control for in future research. Moreover, we showed with the current studies that word-based priming is only a drawback for some paradigms.

In Experiment 1, we used a lexical decision task to divide our stimuli sentences into two groups: a group in which spontaneous trait inferences were detected (the strong group) and a group in which they were not (the weak group). In the strong group, the lexical decision about trait-word was faster after reading trait-implying than after reading rearranged sentences. The difference between these two types of sentences gives us access to the activation of the traits resulting from sentence-based inferences. In the weak group, the trait-implying sentences did not lead to a facilitation in the lexical decision task when compared to the rearranged sentences, which suggests that word-based priming and true spontaneous inference process cannot be distinguished in these pairs.

These two groups of materials were used in three other paradigms: a trait rating task (an explicit measure), the probe recognition task (an implicit immediate measure), and the false recognition task (an implicit delayed measure). We found that the performance in the strong and in the weak groups of materials varied as a function of the paradigm. In the probe recognition task (Experiment 3), consistently with the results in Experiment 1, we only found STI (i.e., a significant difference between trait-implying and rearranged sentences) in the strong group and no trait inferences were detected in the weak group. The pattern was different for the other two tasks (Experiments 2 and 4), where STIs were detected for both strong and weak groups.

The present article is the first to formally address how different STI paradigms are affected by word-based priming. However, future research should investigate the reasons underlying these differences. One possible explanation is the duration of the word-based priming effect. There are some theoretical and empirical findings that support the plausibility of this hypothesis. According to Kintsch's Construction-Integration model (1988), when we read a sentence each word activates, in parallel, concepts in our general knowledge network in a rather promiscuous way. As a consequence, the initial representation of the text includes several plausible concepts but not all of them are relevant. Some concepts are activated due to spreading activation mechanisms and are, in that sense, context-free. This is seen as the price to pay for the tremendous flexibility of our comprehension system. The model further states that this initial noisy and incoherent text-representation goes through cycles of activation until it stabilizes on the representation that is more coherent within the context of the sentence. During this stabilization process, which according to the model corresponds to an integration phase, meaning-inappropriate concepts are deactivated and excluded from the text representation. Thus, on the basis of Kintsch's model, it is possible to predict that priming effects at the lexical level will impact performance at initial processing stages, which would explain its impact on immediate tests, but not necessarily in delayed measures of STIs that access more stable representations. There are also empirical findings that converge to the same prediction. Several studies using ambiguous words have found that, as a word is encoded, all the possible meanings of the word become immediately activated. However, after some time, only the contextually appropriate meanings remain activated (Conrad, 1974; Kintsch & Mross, 1985; Lucas, 1987; Oden & Spira, 1983; Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979). These results suggest that the initial access to multiple meanings is automatic and context-independent. Onifer and Swinney (1981), for example, used a cross-modal priming technique to explore the effect of contexts on meaning activation of lexically ambiguous words. Participants attended to sentences that contain an ambiguous word (e.g., "bank"), either in a context that favored the primary meaning of the word (e.g., "All the cash that was kept in the safe at the bank was stolen last week when two masked men broke in") or in a context that favored the secondary meaning of the word (e.g., "A large piece of driftwood that had been washed up onto the bank by the last storm stood as a reminder of how high the water had actually risen"). Immediately after reading each sentence, a word was visually presented for lexical decision. The word was either related to the primary meaning (e.g., "money"), to the secondary meaning (e.g., "river"), or a control word. Results showed that lexical decision was facilitated for words related to both the primary and the secondary meanings, regardless of the semantic context. However, when the presentation of the probe word was delayed for 1.5 seconds, facilitation was observed only for the word related to the context-appropriate meaning. Applying this same logic to trait inference, if a behavioral description contains words highly associated with the target trait, we expect that the overall meanings of the sentence (via sentence-based inference) and the individual meanings of its words (via word-based priming) to be activated at an early stage of encoding. At this stage, a large overlap of activation is expected between the trait-implying and the rearranged sentences (mostly due to word-based priming), and this overlap may be enough to camouflage the extra activation coming from sentence-based inferences in the trait-implying sentences. Later on, after the different elements in the sentence are integrated and after context-irrelevant and spurious meanings are deactivated, only the relevant activations will remain, such is the case for the true inference. At this stage, the overlap between the rearranged and the trait-implying sentences should be null or much lower than initially. When the overlap is low, detecting the sentence-based inference should to be easier.

To further investigate the role of word-based priming and to test the hypothesis that its impact decays with time, it would be necessary to have exactly the same paradigm applied in an immediate and in a delayed fashion. Not all the paradigms are, however, suitable for this purpose. Neither using the lexical decision task in a delayed mode, nor the false recognition task in an immediate mode can be done without drastic modifications of the two paradigms. A suitable candidate may be the probe recognition paradigm that was already applied in the past in both immediate and delayed fashion (McKoon & Ratcliff, 1986; Van Overwalle et al., 1999).

In this article, we looked at a methodological limitation of written behavioral descriptions commonly used to study STIs. Clarifying this methodological aspect will give us a better understanding of the STI phenomenon, it occurrence, and the cognitive processes and mechanisms underlying it. We hope that this research can contribute to an improvement and refinement of our methodological and theoretical approaches to spontaneous trait inferences, so that our knowledge about how we perceive others gets more and more precise.

APPENDIX A

TABLE A1. The 48 traits and the statistics corresponding to the differences in reaction times between the trait-implying and the rearranged sentences in the lexical decision task: reaction times (rearranged sentence) minus reaction times (trait-implying sentence)

Trait	*t*-test	*df*	*p*-value	M dif.	*SE*	95% CI
open-minded (aberto)	0.545	99	0.587	23	43	[-62, 108]
joyful (alegre)	-1.178	100	0.242	-47	40	[-127, 33]
arrogant (arrogante)	0.019	97	0.985	1	47	[-92, 93]
adventurous (aventureiro)	-0.17	95	0.865	-9	51	[-110, 93]
kind (bondoso)	-1.066	98	0.289	-51	48	[-145, 44]
calculating (calculista)	0.237	101	0.813	16	69	[-120, 152]
calm (calmo)	-0.331	96	0.741	-12	37	[-86, 62]
affectionate (carinhoso)	-0.597	99	0.552	-31	52	[-114, 72]
boring (chato)	-0.057	101	0.955	-3	57	[-117, 111]
jealous (ciumento)	-0.52	98	0.604	-29	55	[-139, 81]
coherent (coerente)	-0.478	99	0.634	-26	54	[-132, 81]
confident (confiante)	-1.253	99	0.213	-55	44	[-143, 32]
cooperative (cooperante)	0.209	94	0.835	12	67	[-119, 148]
religious (crente)	0.714	99	0.477	44	62	[-79, 167]
knowledgeable (culto)	-0.848	101	0.398	-39	46	[-130, 52]

messy (desarrumado)	0.642	99	0.522	35	54	[-73, 142]
low-profile (discreto)	-0.373	101	0.71	-16	42	[-100, 68]
effective (eficaz)	0.032	99	0.974	1	37	[-71, 74]
enthusiastic (entusiasmado)	-0.617	101	0.538	-37	59	[-154, 81]
demanding (exigente)	0.791	100	0.431	39	50	[-59, 138]
flexible (flexivel)	0.332	100	0.741	13	39	[-64, 90]
cold (frio)	-1.935	101	0.056	-85	44	[-172, 2]
greedy (gananciosa)	-0.404	99	0.687	-24	59	[-140, 93]
waster (gastador)	-0.736	93	0.464	-49	66	[-179, 82]
impatient (impaciente)	0.434	100	0.666	24	55	[-86, 134]
incapable (incapaz)	-0.619	100	0.537	-28	45	[-116, 61]
inconsistent (inconsistente)	0.325	92	0.746	24	74	[-123, 172]
ungrateful (ingrato)	-0.648	96	0.519	-43	66	[-173, 88]
interesting (interessante)	-0.212	99	0.832	-11	54	[-118, 95]
sarcastic (ironico)	-0.083	100	0.934	-5	60	[-124, 114]
unrealistic (irrealista)	-989	100	0.325	-78	79	[-233, 78]
sexist (machista)	0.062	98	0.95	4	59	[-114, 121]
sly (manhoso)	-1.697	96	0.093	-108	64	[-235, 18]
liar (mentiroso)	0.083	99	0.934	-5	63	[-129, 119]
niggardly (mesquinho)	-1.892	97	0.061	-94	50	[-192, 5]
mysterious (misterioso)	-0.519	101	0.605	-30	58	[-146, 85]
optimistic (optimista)	0.101	101	0.92	5	47	[-89, 98]
passive (passivo)	-2.631	99	0.01	-159	61	[-280, -39]
patriotic (patriota)	-0.167	100	0.868	-7	44	[-95, 80]
punctual (pontual)	-0.984	99	0.328	-46	46	[-138, 46]
economical (poupado)	-1.236	97	0.22	-67	54	[-174, 41]
racist (racista)	-2.253	101	0.026	-169	75	[-318, -20]
fast (rapido)	-0.342	99	0.733	-15	45	[-104, 74]
respectful (respeitador)	0.397	99	0.692	26	65	[-104, 155]
nice (simpatico)	-0.033	101	0.974	-1	38	[-77, 75]
social (sociavel)	-1.003	101	0.318	-45	45	[-134, 44]
stubborn (teimoso)	0.032	99	0.974	1	56	[-109, 113]
shy (timido)	-1.507	101	0.135	-70	46	[-162, 22]

Note. **Bold** = traits that are part of the strong group whereas the remaining are part of the weak group.

APPENDIX B. ADDITIONAL ANALYSES

B1. EXPERIMENT 1

Additionally to the median-based split of the material (presented in the body of the article), to show that our conclusions are robust and do not depend on this specific method of obtaining the groups, we partitioned the data (after ordering it in function of the difference in reaction times between trait-implying and rearranged conditions) in three groups, each containing one-third of the data. Thus, in Group 1 (weak) the difference in reaction times (trait-implied minus the rearranged) varied from 44 to -3 ms, in Group 2 (intermediate) from -3 to -37 ms, and in Group 3 (strong) from -39 to -169 ms.

B2. EXPERIMENT 2

The groups obtained by splitting the material based on terciles were used to conduct a mixed ANOVA. A main effect of type of sentence was found, $F(1, 45) = 191.99$, $p < .001$, $\eta_p^2 = .42$, with higher ratings for the trait-implying sentences ($M = 7.73$, $SD = .73$) than for the rearranged ones ($M = 4.28$, $SD = 1.45$). No other effects were found, $F < 1$, meaning the patterns of inference is similar in the three groups. These results replicate the results found with the median split.

B3. EXPERIMENT 3

A mixed ANOVA was conducted after splitting the material in three groups. No main effect of type of sentence was detected, $F < 1$, but a significant interaction between the type of sentence and the group, $F(2, 93) = 4.440$, $p = .014$, $\eta_p^2 = .09$, was found. Next, three separate repeated measures ANOVAs were conducted, one for each group. In Group 1 (weak), and as represented in Figure B2, a significant difference between the two types of sentences was found, $F(1, 31) = 4.236$, $p = .048$, $\eta_p^2 = .12$, with lower reaction times in the trait-implying condition ($M = 613$, $SD = 26$) than in the rearranged one ($M = 624$, $SD = 23$). In Group 2 (intermediate), no significant difference between the two types of sentences was found, $F(1, 31) = .048$, $p = .828$, $\eta_p^2 = .002$, with similar reaction times in the trait-implying condition ($M = 623$, $SD = 30$) and in the rearranged one ($M = 625$, $SD = 40$). Finally, in Group 3 (strong), a significant difference between the two types of sentences was found, $F(1, 31) = 11.293$, $p = .002$, $\eta_p^2 = .27$, with larger reaction times in the trait-implying condition ($M = 623$, $SD = 35$) than in the rearranged condition ($M = 610$, $SD = 25$).

B4. EXPERIMENT 4

Two mixed ANOVAs were performed, one for the trait-implying material and one for the rearranged. For the trait-implying condition, a main effect of pairing was found, $F(1,45) = 9.21$, $p = .004$, $\eta_p^2 = .17$, with a higher false recognition rate in the match condition ($M = .43$, $SD = 26$) than in the mismatch condition ($M = 30$, $SD = 20$). When the same analysis is conducted for the rearranged sentences, no significant main effect, $F < 1$, and no interaction between the pairing and the group were found, $F(1,45) = 1.69$, $p = .195$, $\eta_p^2 = .07$, as suggested by in Figure B3. No other effects were found, $F < 1$.

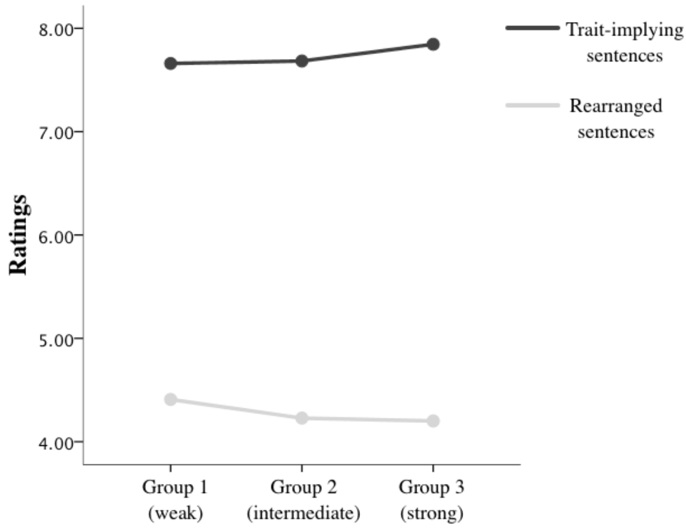

FIGURE B1. Average ratings as a function of type of material (weak vs. intermediate vs. strong pairs of sentences) and type of sentence (trait-implying vs. rearranged) in Experiment 2.

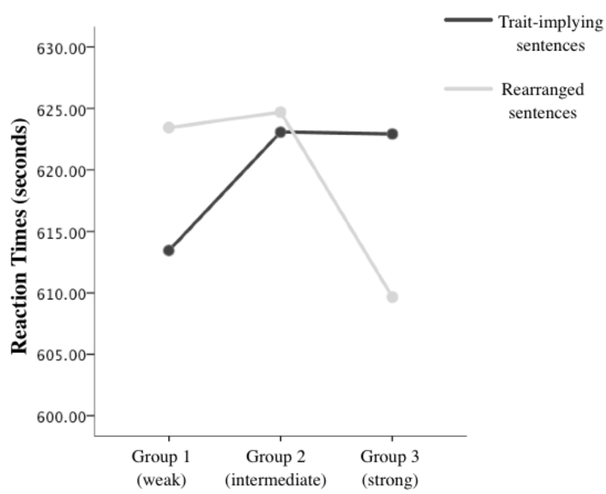

FIGURE B2. Mean reaction time as a function of type of material (weak, intermediate vs. strong pairs of sentences) and type of sentence (trait-implying vs. rearranged) in Experiment 3.

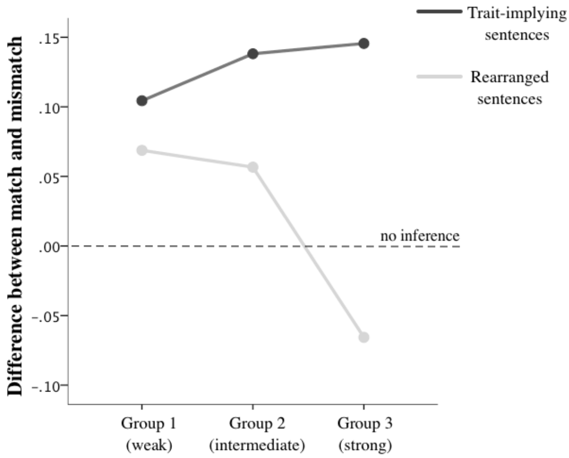

FIGURE B3. Difference between match and mismatch in the false recognition rate as a function of type of material (weak vs. intermediate vs. strong pairs of sentences) and type of sentence (trait-implying vs. rearranged) in Experiment 4.

REFERENCES

Calvo, M. G., Castillo, M. D., & Schmalhofer, F. (2006). Strategic influence on the time course of predictive inferences in reading. *Memory and Cognition, 34*, 68-77. https://doi.org/10.3758/BF03193387

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology, 66*, 840-856. https://doi.org/10.1037/0022-3514.66.5.840

Carlston, D. E., & Skowronski, J. J. (2005). Linking versus thinking: Evidence for the different associative and attributional bases of spontaneous trait transference and spontaneous trait inference. *Journal of Personality and Social Psychology, 89*, 884-898. https://doi.org/10.1037/0022-3514.89.6.884

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: II. On the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology, 69*, 420-436. https://doi.org/10.1037/0022-3514.69.3.429

Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory and Cognition, 2*, 130-138. https://doi.org/10.3758/BF03197504

Crawford, M. T., Skowronski, J. J., Stiff, C., & Scherer, C. R. (2007). Interfering with inferential, but not associative, processes underlying spontaneous trait inference. *Personality and Social Psychology Bulletin, 33*, 677-690. https://doi.org/10.1177/0146167206298567

Domingos, A., & Garcia-Marques, T. (2013). Normas de valência e familiaridade de "não-palavras" portuguesas. *Laboratório de Psicologia, 6*(1), 49-74. https://doi.org/10.14417/lp.693

Dosher, B. A., & Corbett, A. T. (1982). Instrument inferences and verb schemata. *Memory and Cognition, 10*, 531-539. https://doi.org/10.3758/BF03202435

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191. https://doi.org/10.3758/BF03193146

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology, 33*, 465-495. https://doi.org/10.1080/14640748108400804

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692-731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology,* 44, 1355-1361. https://doi.org/10.1016/j.jesp.2008.04.005

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition, 27*, 222-248. https://doi.org/10.1521/soco.2009.27.2.222

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association, 1*, 114-125. https://doi.org/10.1007/s40881-015-0004-4

Gross, D., Fischer, U., & Miller, G. A. (1989). The organization of adjectival meanings. *Journal of Memory and Language, 28*(1), 92-106.

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology, 39*, 434-447. https://doi.org/10.1016/S0022-1031(03)00033-7

Keenan, J. M., & Jennings, T. M. (1995). The role of word-based priming in inference research. In R. F. J. Lorch & E. O'Brien (Eds.), *Sources of coherence in reading* (pp. 37-50). Hillsdale, NJ: Erlbaum.

Keenan, J. M., Potts, G. R., Golding, J. M., & Jennings, T. M. (1990). Which elaborative inferences are drawn during reading? A question of methodologies. In D. A. Balotta, G. B. F. d'Arcais & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 377-402). Hillsdale, NJ: Erlbaum.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*, 163-182. https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W., & Mross, E. F. (1985). Context effects in word identification. *Journal of Memory and Language, 24*, 336-349. https://doi.org/10.1016/0749-596X(85)90032-4

Lucas, M. M. (1987). Frequency effects on the processing of ambiguous words in sentence contexts. *Language and Speech, 30*, 25-46. https://doi.org/10.1177/002383098703000103

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314-324. https://doi.org/10.3758/s13428-011-0168-7

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 82-91. https://doi.org/10.1037/0278-7393.12.1.82

Newman, L. S. (1991). Why are traits inferred spontaneously? A developmental approach. *Social Cognition, 9*, 221-253. https://doi.org/10.1521/soco.1991.9.3.221

Newman, L. S. (1993). How individualists interpret behavior: Idiocentrism and spontaneous trait inference. *Social Cognition, 11*, 243-269. https://doi.org/10.1521/soco.1993.11.2.243

Oden, G. C., & Spira, J. L. (1983). Influence of context on the activation and selection of ambiguous word senses. *The Quarterly Journal of Experimental Psychology, 35*, 51-64. https://doi.org/10.1080/14640748308402116

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition, 9,* 225-236. https://doi.org/10.3758/BF03196957

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition, 33*, 20-66. https://doi.org/10.1521/soco.2015.33.1.20

Orghian, D., Ramos, T., & Garcia-Marques, L. (2018). You are cruel even if he did it: Behavior and face processing in spontaneous trait inference and transference. *Basic and Applied Social Psychology, 40*, 104-114. https://doi.org/10.1080/01973533.2018.1436056)

Orghian, D., Ramos, T., Reis, J., & Garcia-Marques, L. (2018). Acknowledging the role of word-based activation in spontaneous trait inferences. *Análise Psicológica, 36*, 115-131. https://doi.org/10.14417/ap.1319

Orghian, D., Smith, A., Garcia-Marques, L., & Heinke, D. (2017). Capturing spontaneous trait inference with the modified free association paradigm. *Journal of Experimental Social Psychology, 73*, 243-258. https://doi.org/10.1016/j.jesp.2017.07.004

Otten, M., & Van Berkum, J. J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research, 1291*, 92-101. https://doi.org/10.1016/j.brainres.2009.07.042

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language, 27*, 399-415. https://doi.org/10.1016/0749-596X(88)90064-2

Ramos, T., Garcia-Marques, L., Hamilton, D. L., Ferreira, M., & Van Acker, K. (2012). What I infer depends on who you are: The influence of stereotypes on trait and situational spontaneous inferences. *Journal of Experimental Social Psychology, 48*, 1247-1256. https://doi.org/10.1016/j.jesp.2012.05.009

Rim, S., Min, K. E., Uleman, J. S., Chartrand, T. L., & Carlston, D. E. (2013). Seeing others through rose-colored glasses: An affiliation goal and positivity bias in implicit trait impressions. *Journal of Experimental Social Psychology, 49*, 1204-1209. https://doi.org/10.1016/j.jesp.2013.05.007

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*, 489-537. https://doi.org/10.1016/0010-0285(82)90017-2

Sharkey, A. J., & Sharkey, N. E. (1992). Weak contextual constraints in text and word priming. *Journal of Memory and Language, 31*, 543-572. https://doi.org/10.1016/0749-596X(92)90028-V

Skowronski, J. J., Carlston, D. E., & Hartnett, J. (2008). Spontaneous impressions derived from observations of behavior: What a long, strange trip it's been (and it's not over yet). In N. Ambady & J. J. Skowronski (Eds.), *First impressions* (pp. 313-333). New York: Guilford.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18*, 427-440. https://doi.org/10.1016/S0022-5371(79)90237-8

Todorov, A., & Uleman, J. S. (2002). Spontaneous trait inferences are bound to actors' faces: Evidence from a false recognition paradigm. *Journal of Personality and Social Psychology, 83*, 1051-1065. https://doi.org/10.1037/0022-3514.83.5.1051

Todorov, A., & Uleman, J. S. (2003). The efficiency of binding spontaneous trait inferences to actors' faces. *Journal of Experimental Social Psychology, 39*, 549-562. https://doi.org/10.1016/S0022-1031(03)00059-3

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology, 87*, 482-493. https://doi.org/10.1037/0022-3514.87.4.482

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin, 22*, 377-394. https://doi.org/10.1177/0146167296224005

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology, 28*, 211-279. https://doi.org/10.1016/S0065-2601(08)60239-7

Uleman, J. S., Rim, S., Adil Saribay, S., & Kressel, L. M. (2012). Controversies, questions, and prospects for spontaneous social inferences. *Social and Personality Psy-*

chology Compass, 6, 657-673. https://doi.org/10.1111/j.1751-9004.2012.00452.x

Van Overwalle, F., Drenth, T., & Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or to the action? *Personality and Social Psychology Bulletin, 25*, 450-462. https://doi.org/10.1177/0146167299025004005

Wang, M., Xia, J., & Yang, F. (2015). Flexibility of spontaneous trait inferences: The interactive effects of mood and gender stereotypes. *Social Cognition, 33*, 345-358. https://doi.org/10.1521/soco.2015.33.4.1

Whitney, P., & Williams-Whitney, D. (1990). Toward a contextualist view of elaborative inferences. *Psychology of Learning and Motivation, 25*, 279-293.

Wigboldus, D. H., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology, 84*, 470-484. https://doi.org/10.1037/0022-3514.84.3.470

Wigboldus, D. H., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition, 22*, 292-309. https://doi.org/10.1521/soco.22.3.292.35967

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47*, 237-252. https://doi.org/10.1037/0022-3514.47.2.237

Zárate, M. A., Uleman, J. S., & Voils, C. I. (2001). Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition, 19*, 295-323. https://doi.org/10.1521/soco.19.3.295.21469