

Re-examining Spontaneous Trait Transference from an Attributional PerspectiveMarine Rougier ^aLeonard Heusler ^bJan De Houwer ^a

^a Department of Experimental Clinical and Health Psychology, Ghent University,
Belgium

^b Radboud University

Author Note

This research was made possible by the Methusalem grant BOF22/MET_V/002 of Ghent University to Jan De Houwer and by the LEARVUL H2020 Twinning Project No. 952464/2020. Correspondence should be addressed to Marine Rougier, Department of Experimental-Clinical and Health Psychology, Ghent University, H. Dunantlaan 2, 9000 Ghent, Belgium. Email: Marine.Rougier@UGent.be.

The pre-registration files, materials (stimuli and JsPsych code), data, data codebooks, and analytic (R) scripts for all experiments are made publicly available at https://osf.io/9m7gz/?view_only=b910cced96304ffd8eee044256080472. Authors report having no conflict of interest in publishing this work. Studies received approval (number 2021/39) from the ethical committee of the Faculty of Psychology and Educational Sciences at Ghent University.

Marine Rougier played a lead role in experiments conceptualization, methodology programming, data curation, analyses, data collection, methodology, project administration, visualization, and writing the original draft. Leonard Heusler played an equal role (with MR)

in methodology, programming, data curation, data collection and a supporting role in conceptualization and editing. Jan De Houwer played a lead role in funding acquisition an equal role (with MR) in conceptualization and a supporting role in writing the original draft and editing.

Abstract

In spontaneous trait transference (STT), communicators describing the behavior of others (i.e., actors) are perceived as having the personality traits implied by the behavior they describe. We tested whether this effect relies on moderators that are indicative of rule-based, attributional processes: the communicator-actor relation (friends vs. enemies; Experiment 1), the diagnosticity of the statement for inferring the implied trait to the communicator (Experiment 2), and the validity of the statement (i.e., whether it was actually provided by the communicator; Experiment 3). In line with attributional theories, trait ratings revealed a joined impact of the three moderators. Experiment 4 showed communicators were attributed alternative traits – not implied by the behavior itself, but by the fact that they described the behavior. Together, our results suggest that participants attribute traits to the communicator based on the communicator's behavior (i.e., the act of describing the behavior of someone else).

Keywords: Spontaneous trait transference; Evaluative conditioning; Relational moderator; Attribution; Inference.

Re-examining Spontaneous Trait Transference from an Attributional Perspective

Individuals form detailed impressions of others based on minimal information. For instance, people are judged as sociable when physically attractive (Dion et al., 1972) or as more positive (negative) after being paired with another positive (negative) stimulus (Moran et al., 2023). Impressions are also shaped by what individuals say about others. For instance, providing information about the negative (positive) behavior of others leads to a more negative (positive) impression of the communicator (Wyer et al., 1990). Research on Spontaneous Trait Transference (STT) shows statements about others' behavior can even lead to trait-specific assumptions (Skowronski et al., 1998). For instance, if Judy (communicator) reports on John's (actor) honest behavior ("He found an expensive watch and decided to return it"), people are more likely to attribute honesty to Judy than another positive trait (e.g., intelligence).

In this paper, we further examine the nature and properties of STT. Inspired by the idea that STT stems from rule-based, attributional processes, we combined this with knowledge about evaluative conditioning (EC) processes. In the remainder, we first review assumptions about the processes underlying STT. Afterwards, we discuss research on the role of attributional processes in EC. Finally, we provide an overview of our studies. Our aim is to expand knowledge on STT moderators, thereby constraining STT theories. Moreover, by relating moderators typical of evaluative learning to impression formation, this work also participates in integrating these effects into a broader perspective – a gap that has been explicitly highlighted by previous work (De Houwer et al., 2019; Uleman et al., 2008).

STT: Empirical Evidence and Cognitive Models

In a typical STT procedure, participants are exposed to pairs of a photographed individual (communicator) and a trait-implying statement (e.g., "She watched her neighbor's house to see who came and went") – that is, a behavioral statement implying that the actor of

the behavior possesses a specific trait (e.g., 'nosy'). Instructions specify that the communicator discusses another person's behavior (i.e., the communicator is not the actor of the behavior). Then, direct and indirect measures assess the extent to which participants assume the communicator possesses the implied traits. As a direct measure, participants might rate the communicator on the implied trait (e.g., 'nosy') and evaluatively congruent or incongruent other traits (e.g., 'insecure', 'creative') implied in other trials (Skowronski et al., 1998). A STT effect occurs when the implied trait receives higher ratings than other traits. In the indirect false recognition task, participants indicate if a trait was explicitly mentioned in the sentence presented with the person during exposure ('yes' vs. 'no'; Todorov & Uleman, 2002). None of the crucial traits are actually presented but some are implied in the sentence presented together with the face (e.g., nosy; "correct implied") whereas other traits are implied by a statement of another pair (e.g., insecure; "false implied"). A higher rate of false recognition ('yes' responses) for correct (vs. false) implied traits has been interpreted as an STT effect.

The cognitive processes underlying the STT effect are debated between two theoretical perspectives. The dominant associative perspective posits that STT results from automatic binding between mental representations of the communicator and the implied trait due to their co-activation (Brown & Bassili, 2002). This implies unqualified links that disregard the specific relationship between representations (e.g., *Alice is* vs. *wants to be* clever) and operate independently of the validity of this relationship (i.e., whether Alice is actually clever, e.g., Carlston & Skowronski, 2005).

An alternative explanation for STT effects is grounded in rule-based, attributional processes (e.g., Heider, 1958; Kelley & Michela, 1980). Despite the apparent lack of STT's logical basis (as the communicator is not performing the behavior), several factors may lead participants to assume that communicating on another's behavior reflects traits of the

communicator (Carlston & Skowronski, 2005). For example, the communicator may be seen as highlighting behavior they value or as having a similar personality as the actor, especially when studies specify that the actor is an acquaintance of the communicator. Unlike the associative view, the attributional perspective implies that STT effects rely on inferences (i.e., rule-based reasoning) about the communicator's traits. From this perspective, communicators are evaluated not just for *what* they say, but for *why* they chose to say it. Hence, STT would reflect trait attribution based on the act of sharing behavioral information. Hence, factors influencing these inferences – such as whether the communicator (dis)approves or did (not) formulate the statement – should moderate the STT effect. Notably, given that rule-based, attributional processes can operate under conditions of automaticity (e.g., when little time or mental resources are available; see Moors, 2016), STT effects could still qualify as spontaneous in the sense of being mediated by automatic processes. Moreover, at the procedural level (i.e., regardless of the mediating mental processes), the impact of the actor's behavior on perception of the communicator can be seen as spontaneous in that it arises without the instruction to consider the actor's behavior when judging the communicator.

Empirically, the STT effect persists even with explicit warnings to avoid it (Carlston & Skowronski, 2005) or when participants are asked to detect if the communicator is lying (Crawford et al., 2007). Thus, STT is not eliminated by information likely to prevent/invalidate attributional processes. The effect also occurs under illogical conditions, such as with ostensible random communicator-statement pairings or when the target stimulus is an inanimate object (e.g., a banana; Brown & Bassili, 2002; Goren & Todorov, 2009). Additionally, STT lacks a negativity bias (greater impact of negative behaviors; Carlston & Skowronski, 2005), which challenges attributional explanations that typically predict this bias due to the greater diagnosticity of negative behaviors.

On the other hand, STT is influenced by working memory capacity, contradicting associative theories that assume efficient association formation (Wells et al., 2011; but see Marien et al., 2012). Literature on gossiping shows observers also formulate assumptions about the communicator beyond the described behavior. For instance, communicators discussing others' extreme immoral or competent behaviors are positively perceived (Peters & Kashima, 2015),¹ which challenges the idea of mere communicator-trait associative binding.

In sum, evidence for associative processes is not as clear-cut as often suggested. This is especially true considering that previous research focused on whether STT persists when attributions are unlikely (e.g., random pairing), without comparing this experimental condition with a control condition (e.g., with non-random pairing; Goren & Todorov, 2009; Skowronski et al., 1998). Thus, even if STT seems to occur under conditions that do not favor attributions, understanding how STT changes when attributional logic is altered is essential to understand the mechanisms underlying this effect.

Attributional Processes in Evaluative Conditioning

STT resembles EC at both procedural and theoretical levels. In EC, the evaluation of a neutral stimulus (conditioned stimulus) changes due to pairings with a positive or a negative stimulus (unconditioned stimulus; De Houwer, 2007). Procedurally, both effects involve making (trait-specific or valenced) assumptions about a target stimulus (communicator or conditioned stimulus) based on the pairing with a distinct source information (behavioral statement or unconditioned stimulus; see De Houwer et al., 2019).

Theoretically, EC also involves two distinct perspectives similar to those that have dominated STT research. From an associative perspective, EC is due to the formation and

¹ Gossiping research differs from STT because it targets general positive/negative perception of communicators, not trait-specific assumptions about communicators (i.e., whether the communicator possesses the trait implied by the behavioral statement).

activation of associations between the representations of conditioned and unconditioned stimuli (e.g., Baeyens et al., 1992). More recent accounts contend that EC effects arise from forming and retrieving propositions about relations in the environment (e.g., the proposition “the stimuli go together”; De Houwer, 2009, 2018). Propositions specify how events are related (e.g., “the conditioned stimulus predicts [vs. causes] the unconditioned stimulus”) and can be assessed as true or false (e.g., De Houwer, 2009). Thus, as in attribution theories (e.g., Heider, 1958), individuals apply rule-based reasoning or “inferences” (e.g., inferring similarity in valence from co-occurrence).

Propositional theories inspired research on several previously unexplored moderators of EC (see De Houwer et al., 2020, for a review). First, information on how stimuli are related can moderate EC. For example, standard EC effects (i.e., more positive evaluation of conditioned stimuli paired with positive [vs. negative] unconditioned stimuli) arise when stimuli are described as friends but are reduced or reversed when they are described as enemies (Fiedler & Unkelbach, 2011; Högden & Unkelbach, 2021). Second, diagnosticity also affects EC. For instance, causal descriptions (the conditioned stimulus causes the unconditioned stimulus) lead to larger effects than predictive descriptions (the conditioned stimulus predicts the unconditioned stimulus), likely because they are more diagnostic of the conditioned stimulus’s valence (Hughes et al., 2019). Third, validity of pairings influences EC. For instance, informing participants that the positive/negative behaviors (unconditioned stimuli) consistently paired with men (conditioned stimuli) were accidentally mixed-up reduced the EC effect compared to valid pairings (Moran et al., 2017).² Finally, these moderating effects were typically stronger on direct measures (i.e., ratings) than on indirect measures (e.g., Implicit Association Test; Hu et al., 2017; Moran & Bar-Anan, 2013; Zanon

² This procedure departs somewhat from typical EC in that it involves more socially relevant stimuli (i.e., descriptions of behaviors) instead of more minimal evaluative content (e.g., images).

et al., 2014). Some have interpreted this as evidence for dual-process models that attribute effects on direct measures to inferential processes and effects on indirect measures to associative processes (Rydell & McConnell, 2006). Others have explained these dissociations in terms of inferential processes only (De Houwer, 2018).

In this work, we tested whether STT could similarly be influenced by relation, diagnosticity, and validity information. From an attributional perspective, STT should vary as a function of these factors that affect the perceived warrant for inference. A purely associative account, by contrast, predicts that STT should arise from the mere communicator-statement co-occurrence, regardless of such moderators. To gauge evidence in favor of associative processes, we also examined whether STT persists when logical grounds for inference are absent (e.g., Moran, 2024).

Because both direct ratings and the indirect false recognition task have been used in prior STT research (usually in separate experiments), we also employed both measures in our research. The typical justification for using the indirect false recognition task in STT research is that, compared to direct ratings, it provides a purer index of associative processing (e.g., Goren & Todorov, 2009). From this perspective, one would expect that variables that influence attributional processes (such as the potential moderators that we tested) should have a smaller effect on false recognition performance than on direct ratings. However, we did not formulate this prediction for two reasons. First, from an attributional perspective, it is not necessarily the case that these variables have a smaller effect on indirect than on direct measures (see De Houwer et al., 2021, p. 878, for a discussion). Second, because participants in the false recognition task are instructed to judge whether a trait was mentioned rather than whether it fits the person, it might well be that recognition performance merely reflects memory processes rather than impression formation. We will not discuss this issue further here but will return to it in the General Discussion.

Testing predictions from attributional/associative accounts informs on the nature of STT as an empirical phenomenon (i.e., what facilitates the emergence of STT) and it constrains any cognitive future model of STT effects on which moderators should be considered in the model. Importantly, however, note that we did not aim to settle in a definitive manner the theoretical debate on whether STT is due to associative processes, attributional processes, or both. As can be seen in EC, findings supporting one view (e.g., propositional) can often be explained by the other (e.g., associative; see De Houwer et al., 2013, 2020, 2021; Moran et al., 2023). Instead, our primary aim is to leverage the attributional perspective – a perspective that has been overlooked so far – to generate new predictions about STT moderators.

The Present Work

In a Preliminary Experiment, we replicated the typical STT effect using both rating and false recognition tasks (methods and results are available as Supplementary Materials). In Experiment 1, we examined the moderating role of the communicator-actor relation (friends or enemies). Experiment 2 assessed participants' perception of statement diagnosticity (i.e., the extent to which a statement is diagnostic to assume the communicator has the implied trait)³, allowing us to assess its role in the STT effects of the Preliminary Experiment and Experiment 1, as well as the joint impact of relation and diagnosticity in Experiment 1. In Experiment 3, we manipulated validity by specifying whether the communicator formulated the statement (communicator-statement match) or not (mismatch). We tested the impact of validity and the joint impact of validity and diagnosticity. In Experiments 1-3, we used both

³ This type of 'communicator-based' diagnosticity (i.e., whether communicating about someone else's behavior indicates the communicator has the implied trait) departs from what could be coined as 'actor-based' diagnosticity (i.e., whether the behavior clearly indicates the actor has the implied trait). All the behavioral statements we selected to have high actor-based diagnosticity for the implied trait (see Experiment 1), that is, behaviors strongly evoked the implied trait about the actor. This placed us in the best position to observe STT effect if it were driven by automatic trait-communicator binding.

ratings and a false recognition task and thus tested the measure-dependency of effects. We also tested whether STT remained significant under conditions unfavorable to attributional processes. Finally, Experiment 4 followed up on the observation that the STT effect was descriptively reversed for low diagnostic traits (lower ratings for implied than non-implied trait). We tested whether, in case of low diagnostic traits, alternative traits were attributed to the communicator.

All experiments were pre-registered on the Open Science Framework, including a priori theoretical reasoning, hypotheses, power estimations, procedures, and statistical analyses. Sample size determination, all data exclusions, all manipulations, all measures, and deviations from initial pre-registrations are explicitly reported. Pre-registration files, materials (stimuli, study code), data, data codebooks, and analytic (R) scripts for all experiments are publicly accessible at https://osf.io/9m7gz/?view_only=b910cced96304ffd8eee044256080472. Ethic approval was obtained from the ethical committee of the Faculty of Psychology and Educational Sciences at Ghent University.

Experiment 1

In STT, communicators and actors are portrayed as acquaintances or without relational context. Building on findings from EC (Fiedler & Unkelbach, 2011), we tested whether the communicator-actor relationship (friends or enemies) alters STT effects. If STT depends on attributional processes, depicting the actor and communicator as friends (vs. enemies) would facilitate the inference that both are similar and thus that the actor's behavior is informative about communicator's traits.

Method

Deviation from the pre-registration

We pre-registered that relational information would influence both ratings and false recognition but also explored whether the former was influenced more than the latter. For all experiments, we opted for mixed-models as the primary analysis instead of OLS regressions (mixed-models were always pre-registered as an addition). We initially planned OLS regression to remain consistent with prior literature but ultimately decided to rely on mixed models because they offer greater robustness and generalizability and they can handle continuous within-participant variables (e.g., statement diagnosticity; Judd et al., 2012). For transparency and comparability purpose, OLS regression analyses are available in the Supplementary Materials (Table S2). To facilitate comparison with existing literature, we also report the residual STT effects (i.e., STT effects when logical grounds are not met, e.g., STT in the ‘enemies’ condition) in OLS regression (Table S8). We indicate when mixed model and OLS regression analyses lead to distinct conclusion.

Participants and Design

We based our sample size on a recent similar EC research ($dz = 0.50$, Högden & Unkelbach, 2021, Exp. 1). A sample of 200 participants would yield 99.87% power (two-tailed paired t -test, $\alpha = 0.05$) to detect both the smallest STT effect from our Preliminary Experiment ($dz = 0.33$; see Supplementary Materials) and the relational moderation effect ($dz = 0.50$). We recruited 199 participants ($M_{age} = 41.22$, $SD_{age} = 14.71$, 81 women, 54 men, 4 responded “other”) via Prolific Academic (£2.70 retribution).⁴ All were US citizens, native English speakers, first-time participants in our lab’s studies, and had an approval rate of $\geq 98\%$. We used a 2 (*Relation*: friends vs. enemies) x 2 (*Type of trait*: correct implied vs. false implied) design for the false recognition task and a 2 (*Relation*: friends vs. enemies) x 3

⁴ Small discrepancies between the planned and actual number of participants (e.g., 199 actual participants instead of the planned 200 in Experiment 1) were due to recruitment via Prolific Academic, where participants are sometimes counted without generating data (or vice versa). When this happened, counted only participants with data. Due to a programming error, demographic information was missing for 60 participants.

(*Type of trait*: implied vs. evaluatively congruent vs. evaluatively incongruent) design for the rating task. All variables were manipulated within participants.

Materials

We selected 36 behavioral statements, including 24 trait-implying statements and 12 fillers (for a complete list, see Table S1). Trait-implying sentences described a behavior of a third party (e.g., “S/he left a 25% tip for the waitress”) implying, but not mentioning explicitly, a trait (e.g., “generous”; selected from Kruse & Degner, 2023; Uleman, 1988; Van Overwalle, 2012). The behavioral statements were pretested in priorly cited research to consensually evoke the implied trait for the actor of the behavior (i.e., it was listed by participants in > 66% of cases; see Table S1). Filler sentences described a behavior that explicitly mentioned the corresponding trait (e.g., “S/he confidently walked into the interview room”; the trait being “confident”; selected from Kruse & Degner, 2021, 2023). Half of statements referred to positive traits (e.g., “smart”) and half to negative traits (e.g., “nosy”). For communicators, we used photographs of female (18) or male (18) faces selected from the 10k US Adult Faces Database (Bainbridge et al., 2013). Faces were all unknown (i.e., non-celebrity) and varied in age (i.e., from 20-30 years to the 60+ category) and race (30 White faces, three Black faces, one South Asian face, and two Hispanic faces; see OSF repository for full details).

Procedure

Exposure Task. The experiment was programmed on *Lab.js* (Henninger et al., 2022) and was framed as a study on memory (Goren & Todorov, 2009). In the initial “memory task,” participants memorized pairs consisting of a face (communicator) and a statement. Statements were presented as a short excerpt from a longer interview in which communicators described the behavior of someone they knew. Participants were told that communicators were instructed to describe either a very good friend or a serious enemy. Each

face-statement pair was displayed with the label “friends” or “enemies” in a text box above it. To clarify that the communicator was referring to another person’s behavior, the gender of the actor was always opposite to the communicator.

For each participant, 36 photograph-statement pairs (with $N = 24$ trait-implying and $N = 12$ filler statements) were presented once in random order. Each pair appeared for 8 seconds with a 1-second inter-trial interval (Todorov & Uleman, 2003). Within each type (trait-implying and filler) and trait valence (positive/negative), the gender of the paired face and the relation label (friend/enemy) were randomized with a 50/50 distribution constrain. The 24 trait-implying statements could either be correct/false implied in the false recognition task (50/50 distribution). In the rating task, when used as control traits, they could either be congruent/incongruent (50/50 distribution). Because of randomization, minor imbalances may have occurred at the level of individual participants (e.g., more female faces or ‘enemies’ label for correct implied trait) but proportions were fully balanced at the group level.

False Recognition Task. In the “recognition task”, participants viewed faces from the previous phase paired with a word, which was either implied by the behavior shown with that face or implied by another person’s sentence (Todorov & Uleman, 2002). The photograph appeared at the center, the trait below, and “yes” and “no” response options at the bottom. Participants selected “yes” if they believed the word was part of the original sentence and “no” if it was not. For filler statements ($N = 12$), the trait explicitly mentioned always matched the face. For other statements, faces were paired with either the correct implied trait ($N = 12$) or a false implied trait from another face’s statement ($N = 12$).

Rating Task. In the “rating task”, participants rated how much each face possessed three traits on a Likert scale from 1 (*not at all*) to 7 (*extremely*). One trait was always the one implied by the original statement (or the trait explicitly mentioned for fillers), while the other

two were distractors – an evaluatively congruent trait (same valence) and an incongruent trait (opposite valence) that were implied or filler traits from other trials (Carlston & Skowronski, 2005). Trait combinations and order were randomized. Because the false recognition task always preceded the rating task, we tested whether the trait type (correct vs. false implied) in the former influenced STT in the latter (e.g., via priming; see Table S7 in Supplementary Materials). This did not have a significant effect in any experiment.

Memory of relation. Participants reported whether each face had been labeled as “friends”, “enemies”, or whether they did not remember (full results on this exploratory variable are presented as Supplementary Materials Table S3). Finally, they completed exploratory questions (see Supplementary Materials), could leave comments, completed demographic questions, were debriefed, and received payment and contact information.

Results

Following our pre-registration, we excluded three participants with zero variance in their ratings and removed data from filler statements. If any control variable (e.g., trait valence, participants’ gender) significantly moderated the STT effect or other key effects, these are mentioned in the main text. The inclusion of control variables never changed the significance of the main effects of interest. As preregistered, we therefore report the results of the models that do not include these controls. Results on the exploratory variables and correlations between the false recognition and rating tasks are provided as Supplementary Materials (Tables S3 and S4).

Data were analyzed using RStudio, version 1.4.1106 (RStudio Team, 2021). Mixed-model analyses were performed using the *lme4* package version 1.1-30 (false recognition task; Bates et al., 2015) and *lmerTest* package version 3.1-0 (rating task; Kuznetsova et al., 2017). When STT moderation by relational information is non-significant in frequentist analyses, we report the Bayes Factor *BF01* to assess evidence for a null effect (H_0). We used

the JZS default Bayes factor (*ttestBF* function) from the *BayesFactor* R package (version 0.9.12-4.2; Morey et al., 2015), applying a Cauchy prior distribution with the r scale $\sqrt{2}/2$ (Rouder et al., 2009). Bayes factors interpretation follows Lee and Wagenmakers's (2014) classification.

False Recognition Task

We ran a logistic mixed-model with trait type (false implied = -0.5, correct implied = 0.5) and relational information (enemies = -0.5, friends = 0.5) as fixed effects. This model predicted the log-odds of answering ‘yes’ (1) or ‘no’ (0). A STT effect would be a higher rate of ‘yes’ responses (false recognition) for correct implied traits compared to false implied traits. We also estimated the random slopes for trait type and relation, with participants and statements as random factors.

The likelihood of ‘yes’ responses was higher for correct implied ($M = 0.23$, $SD = 0.15$) than for false implied traits ($M = 0.19$, $SD = 0.13$), $B = 0.45$ (0.09), $z = 5.00$, $p < .001^5$, indicating a STT effect. Contrary to our prediction, this effect was very similar in the ‘friends’ ($M_{correct} = 0.23$, $SD_{correct} = 0.15$, $M_{false} = 0.18$, $SD_{false} = 0.13$) and ‘enemies’ conditions ($M_{correct} = 0.23$, $SD_{correct} = 0.15$, $M_{false} = 0.20$, $SD_{false} = 0.13$), $B = 0.19$ (0.15), $z = 1.31$, $p = .19$, $BF_{01} = 8.84$.

Rating Task

We estimated a mixed-model with trait type (implied vs. evaluatively congruent vs. evaluatively incongruent) and relational information (enemies = -0.5, friends = 0.5) as fixed effects, and ratings as the outcome measure. Trait type was coded using two orthogonal contrast codes: quadratic contrast (C1: implied = 2/3, evaluatively congruent = -1/3, evaluatively incongruent = -1/3) and linear contrast (C2: implied = 0, evaluatively congruent

⁵ In the case of logistic mixed-models, B represents the change in log-odds of the outcome per unit increase in the predictor and the value in parenthesis is the standard deviation of this parameter. We did not compute the effect sizes for the mixed-model analyses given the lack of consensus on this matter (Correll et al., 2021). Effect sizes for OLS regression can be found in Table S2 (Supplementary Materials).

= 1/2, evaluatively incongruent = -1/2). While the first contrast (C1) tests the STT effect (i.e., higher ratings for implied than non-implied traits), the second contrast (C2) explores a potential valence congruency effect (e.g., higher ratings for evaluatively congruent than incongruent traits). We estimated random slopes for trait type and relational information, with participants and statements as random factors.

The average STT effect comparing the ratings for implied ($M_{implied} = 3.87$, $SD_{implied} = 1.60$) vs. evaluatively congruent and incongruent traits considered jointly (contrast C1; $M_{cong} = 3.84$, $SD_{cong} = 1.57$, $M_{incong} = 3.78$, $SD_{incong} = 1.58$), did not emerge $B = 0.06$ (0.06), $t(23.21) = 1.10$, $p = .28$. STT emerged when relying on OLS regression, that is, when dismissing variability stemming from statements (see Supplementary Material Table S2).⁶ The difference of rating between evaluatively congruent and incongruent traits was not significant, $B = 0.05$ (0.04), $t(26.06) = 1.36$, $p = .19$ (contrast C2).

The STT effect was not significantly larger in the ‘friends’ ($M_{implied} = 3.88$, $SD_{implied} = 1.59$, $M_{cong} = 3.82$, $SD_{cong} = 1.58$, $M_{incong} = 3.79$, $SD_{incong} = 1.59$) than ‘enemies’ condition ($M_{implied} = 3.86$, $SD_{implied} = 1.60$, $M_{cong} = 3.85$, $SD_{cong} = 1.56$, $M_{incong} = 3.78$, $SD_{incong} = 1.57$), $B = 0.03$ (0.06), $t(56.40) = 0.52$, $p = .61$, $BF_{01} = 9.91$. Interestingly, the STT by relation interaction was moderated by the exploratory variable of relation memory when adding this variable as a fixed factor in the model (incorrect = -0.5 vs. correct = 0.5 ; see Supplementary Materials Table S3) so that the expected interaction emerged for participants with correct memory, $B = 0.19$ (0.10), $t(406.30) = 1.97$, $p = .0497$, but not those with incorrect memory, $B = -0.05$ (0.07), $t(123.70) = 0.75$, $p = .45$. Considering only participants with correct memory, the residual STT effect in the ‘enemies’ condition did not emerge, $B = 0.01$ (0.08), $t(101.60)$

⁶ Additional mixed-model analyses revealed that the STT effect significantly varied across statements, $\chi^2 = 37.96$, $p < .001$. This variability was also significant in the Preliminary Experiment, $\chi^2 = 46.07$, $p < .001$.

= 0.11, $p = .91$ – the STT effect emerged in the ‘friends’ condition, $B = 0.20$ (0.09), $t(59.25) = 2.35$, $p = .02$.

Finally, the valence of the traits moderated the STT by relational information interaction so that the interaction was larger for positive than for negative traits, $B = 0.28$ (0.12), $t(83.63) = 2.45$, $p = .016$.⁷ When considering only positive traits, the interaction between STT and relational information was significant, $B = 0.17$ (0.08), $t(87.87) = 2.09$, $p = .040$, while it was not significant for negative traits, $B = -0.11$ (0.08), $t(87.95) = 1.31$, $p = .19$. Considering only positive traits, the residual STT effect in the ‘enemies’ condition did not emerge, $B = -0.05$ (0.08), $t(26.25) = 0.65$, $p = .52$ – nor the STT effect in the ‘friends’ condition, $B = 0.12$ (0.10), $t(22.69) = 1.23$, $p = .23$.

Discussion

STT emerged in false recognition but not in the rating task. The communicator-actor relationship did not significantly influence STT on false recognition. In the rating task, moderation occurred only for accurate recall of relational information and positively valenced statements. Hence, we found only limited evidence for a moderation of STT by relational information. The strong variation in STT across statements suggests that an important statement-level boundary condition, with some statements facilitating STT.

One possible interpretation is that some statements were perceived as more diagnostic than others for attributing the implied trait to the communicator. From an attributional perspective, STT should arise only for statements that are diagnostic for inferring traits of the communicator. Moreover, relational information could matter only when statements allow for trait inference in the first place (high diagnostic). In sum, relational information and diagnosticity might jointly influence STT effects. We assessed statements’ diagnosticity in

⁷ Valence did not emerge as a moderator in later studies.

Experiment 2 to retrospectively assess the role of diagnosticity in the Preliminary Experiment and Experiment 1.

Experiment 2

Experiment 2 involved participants rating the diagnosticity of behavioral statements for inferring the communicators' traits. We also pre-tested the planned validity manipulation for Experiment 3 by varying whether the statement matched the communicator (paired with the communicator who formulated it) or mismatched (paired with a different communicator). We expected higher diagnosticity judgments for 'match' trials.

Method

Deviation from the pre-registration

In the pre-registration of Experiment 2, we focused on the effect of statement diagnosticity on the main STT effect in both the Preliminary Experiment and Experiment 1. After collecting data for Experiment 2, we decided to also test the joint effect of diagnosticity and relational information on the STT (i.e., the three-way interaction).

Participants and Design

A sample of 100 participants provides 80% of power to detect a validity effect (pretested manipulation) on diagnosticity ratings of $dz = 0.25$ (two-tailed t -test for paired samples with a 5% false-positive rate). We ended up with 101 participants ($M_{age} = 35.91$, $SD_{age} = 14.17$, 74 women, 24 men, 3 responded "other"). Recruitment and pre-selection criteria were the same as before except participants were paid £1.20. We relied on a within-participant design with *Validity* (match vs. mismatch) as the only variable.

Materials and Procedure

The experiment was programmed using *jsPsych* (de Leeuw, 2015). Participants were presented with the 36 photograph-statement pairs (generated following the same constraints as before). Each trial included the word "MATCH" (green) or "MISMATCH" (red) displayed in

uppercase above the pair. They were informed that “MATCH” indicated the behavioral statement was formulated by the paired communicator (i.e., valid pair), while “MISMATCH” meant the statement was formulated by a different communicator displayed in another pair during the exposure phase (i.e., invalid pair). Full instructions are available in the Supplementary Materials.

Participants were instructed to carefully examine all the information on the screen, including the communicator, the behavioral description, and the match/mismatch information. They then rated how useful (i.e., diagnostic) the information was for drawing conclusions about the communicator’s personality (e.g., “To what extent is this information useful to draw a conclusion about whether the person in the photograph is [trait]?”) on a scale from 0 (*not useful at all*) to 10 (*very useful*). The question appeared at the bottom of the screen alongside the communicator’s photograph, the statement, and the match/mismatch information.

Results

We excluded three participants with zero variance in diagnosticity ratings. The analysis comprised two parts: (1) testing statement-level diagnosticity as a moderator of prior STT effects and (2) examining the pretested impact of match/mismatch information on diagnosticity ratings.

Re-analysis of Preliminary Experiment and Experiment 1

Ratings from “match” trials were used to calculate the average diagnosticity for each statement.⁸ Higher scores indicated greater diagnosticity for inferring the communicator’s possession of the corresponding trait. Scores ranged from 3.50 (“S/he took the elevator up

⁸ We also computed a score for ‘mismatch’ trials and tested whether the statement diagnosticity, for these trials, moderated the observed effects in the Preliminary Experiment and Experiment 1. No significant moderation of the statement diagnosticity emerged when using this continuous score. This makes sense because, logically speaking, statements that were not given by the communicator should always be low in diagnosticity for inferring traits of the communicator.

one flight” for ‘lazy’) to 5.78 (“S/he watched her/his neighbor's house to see who came and went” for ‘nosy’). We then used this score in re-analyzing STT effects from Preliminary Experiment (presented as Supplementary Materials) and Experiment 1.

Variables used the same contrast coding as before, with continuous diagnosticity mean-centered. Mixed-models were identical to prior analyses but included statement diagnosticity as a fixed effect.

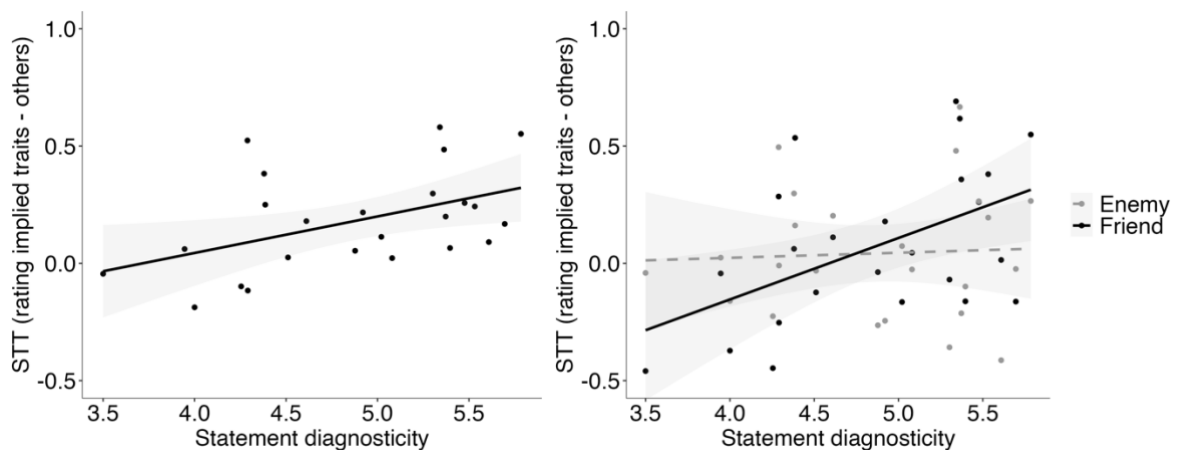
Preliminary Experiment. In the false recognition task, the STT effect was not moderated by statement diagnosticity, $B = -0.03$ (0.10), $z = 0.36$, $p = .72$. In the rating task, the STT effect was moderated by statement diagnosticity, $B = 0.16$ (0.06), $t(22.54) = 2.44$, $p = .023$, so that the more diagnostic the statement the larger the STT effect (see Figure 1, left panel). No residual STT emerged for low diagnostic statements (-1SD), $B = 0.08$ (0.06), $t(29.59) = 1.37$, $p = .18$ – the STT effect emerged for high diagnostic statements (+1SD), $B = 0.28$ (0.06), $t(29.61) = 4.59$, $p < .001$.

Experiment 1. In the false recognition task, we did not observe moderation of the STT effect by statement diagnosticity, $B = -0.03$ (0.14), $z = 0.23$, $p = .82$. Similarly, there was no three-way interaction when considering the relational information, $B = 0.16$ (0.24), $z = 0.66$, $p = .51$. In the rating task, statement diagnosticity did not moderate the STT effect, $B = 0.14$ (0.09), $t(21.99) = 1.64$, $p = .11$, but we did observe a three-way interaction, $B = 0.24$ (0.09), $t(111.89) = 2.63$, $p = .010$ (see Figure 1, right panel). For highly diagnostic traits, the STT effect was larger for ‘friends’ than ‘enemies’ trials, $B = 0.18$ (0.08), $t(155.99) = 2.23$, $p = .030$. For low diagnostic traits, the interaction between the STT effect and the relational information was not significant, $B = -0.12$ (0.08), $t(117.21) = 1.44$, $p = .15$. Moreover, in the ‘friends’ condition, the STT effect was moderated by statement diagnosticity, $B = 0.26$ (0.10), $t(23.97) = 2.62$, $p = .015$, so that the higher the statement diagnosticity the larger the STT effect. As depicted in Figure 1, the STT effect descriptively, but not significantly, $B = -0.09$

(0.09), $t(24.68) = 0.98$, $p = .33$, reversed for low diagnostic statements in the ‘friends’ condition. In the ‘enemies’ condition, this interaction was not significant, $B = 0.02$ (0.10), $t(23.52) = 0.23$, $p = .82$. No residual STT emerged for low diagnostic statements (-1SD), $B = -0.03$ (0.08), $t(22.17) = 0.37$, $p = .72$ (when considering both ‘friends’ and ‘enemies’ conditions)⁹ and STT did also not emerge for high diagnostic statements (+1SD), $B = 0.15$ (0.08), $t(21.89) = 1.97$, $p = .062$. Similarly, no residual STT emerged for the ‘enemies’ condition, $t(24.83) = 0.74$, $p = .47$ (when considering all statements) and STT did also not emerge for the ‘friends’ condition, $B = 0.07$ (0.06), $t(23.86) = 1.21$, $p = .24$.

Figure 1

STT effect in the rating task (rating correct - false implied scores) at the trait level, as a function of the statement diagnosticity in the Preliminary Experiment (left panel) and as a function of relational information (friends vs. enemies) in Experiment 1 (right panel)



Note. The STT effect is represented by the difference in ratings for the implied traits (i.e., traits implied by the behavioral statement of the communicator) and for the non-implied traits (i.e., evaluatively congruent and incongruent traits implied by other behavioral statement of

⁹ This residual effect emerged in the opposite direction (reversed STT) in OLS regression (see Table S8).

other communicators). A difference in ratings above zero represents an STT effect and below zero a reversed STT effect. Grey areas represent the 95% confidence intervals.

Effect of Validity Information on Diagnosticity Ratings

We employed a mixed-model with validity (mismatch = -0.5, match = 0.5) as a fixed effect, including its random slope for participants and traits. Diagnosticity ratings served as the outcome measure. Perceived diagnosticity was higher for ‘match’ ($M = 4.79$, $SD = 3.06$) than ‘mismatch’ trials ($M = 2.43$, $SD = 2.65$), $B = 2.36$ (0.22), $t(103.02) = 10.48$, $p < .001$. Independent of the match/mismatch information, positive (vs. negative) traits were evaluated as more diagnostic, $B = 0.26$ (0.12), $t(34.08) = 2.13$, $p = .040$.

Discussion

Statements diagnosticity moderated STT in the Preliminary Experiment, as measured by ratings. In Experiment 1, we observed a joint impact of diagnosticity and relational information: STT in the rating task was strongest with a ‘friends’ relation and highly diagnostic statements. If either the relation or diagnosticity was suboptimal for making inferences about the traits of the communicator, STT effects were weaker. Note that in the Preliminary Experiment, the communicator and actor were always said to be acquaintances, which is similar to the ‘friends’ condition of Experiment 1. It is therefore encouraging to see statement diagnosticity moderated STT in the rating task in both the Preliminary Experiment and the ‘friends’ condition of Experiment 1.

We did not, however, find any impact of the moderators in the false recognition task. This dissociation will be further addressed in the general discussion section.

Experiment 3

Experiment 3 employed the pretested validity (match/mismatch) manipulation during the exposure phase. According to an attributional perspective, (a) validity should influence

the STT effect by facilitating (valid) or preventing (invalid) rule-based reasoning and (b) this effect should depend on the diagnosticity of the statements. Specifically, STT should be largest when the communicator matches with the statement and when the statement is highly diagnostic for attributing the implied trait.

Method

Deviation from the pre-registration

None.

Participants and Design

The design of Experiment 3 mirrored that of Experiment 1, with the addition of statements diagnosticity (within-participants; mean-centered). We planned 200 participants for 80% power and $d_z = 0.18$ (two-tailed paired t -test, 5% false-positive rate). A total of 201 participants were recruited ($M_{age} = 35.94$, $SD_{age} = 11.74$, 93 women, 100 men, 6 responded “other”, and 2 preferred not to say). Recruitment and pre-selection criteria were consistent with previous experiments (retribution of £2.80).

Materials and Procedure

The materials and procedure were identical to Experiment 1, except that relational information (friends/enemies) was replaced by validity information indicating the match/mismatch between the communicator and the behavioral statement. Prior to the exposure phase, participants received instructions about the match/mismatch information identical to Experiment 2.

Exposure trials followed the same time course as in Experiment 2. Of the 36 photograph-statement pairs, half were randomly assigned the “MATCH” label (green) and the other half the “MISMATCH” label (red). Following the false recognition and rating tasks, participants were asked to recall, for each face, whether it was presented with a “MATCH”

label, a “MISMATCH” label, or if they did not remember. However, due to a programming error, responses for this task were not recorded.

Results

We excluded four participants with zero variance in ratings. We applied the same analytical strategy as in Experiment 1, focusing on validity instead of relational information.

False Recognition Task

The likelihood of “yes” responses was higher for correct implied ($M = 0.23$, $SD = 0.14$) than for false implied traits ($M = 0.18$, $SD = 0.13$), $B = 0.49$ (0.10), $z = 5.13$, $p < .001$. This STT effect was not significantly larger in the ‘match’ ($M_{correct} = 0.23$, $SD_{correct} = 0.13$, $M_{false} = 0.18$, $SD_{false} = 0.12$) than in the ‘mismatch’ condition ($M_{correct} = 0.23$, $SD_{correct} = 0.14$, $M_{false} = 0.18$, $SD_{false} = 0.13$), $B = 0.19$ (0.14), $z = 1.38$, $p = .17$, $BF_{01} = 6.99$. When adding the diagnosticity score in the mixed-model, this variable did not moderate the STT effect, $B = 0.15$ (0.15), $z = 1.05$, $p = .29$, or its interaction with validity, $B = 0.02$ (0.22), $z = 0.10$, $p = .92$.

Rating Task

Ratings for implied ($M_{implied} = 4.03$, $SD_{implied} = 1.63$) vs. evaluatively congruent and incongruent traits considered jointly (contrast C1; $M_{cong} = 3.94$, $SD_{cong} = 1.60$, $M_{incong} = 3.91$, $SD_{incong} = 1.60$), did not differ $B = 0.11$ (0.07), $t(30.69) = 1.46$, $p = .15$. As in Experiment 1, the STT effect did emerge when relying on OLS regression (see Supplementary Material).¹⁰ The difference of rating between evaluatively congruent and incongruent traits was not significant, $B = 0.03$ (0.04), $t(33.37) = 0.88$, $p = .38$ (contrast C2).

STT was not significantly larger in the ‘match’ condition ($M_{implied} = 4.04$, $SD_{implied} = 1.65$, $M_{cong} = 3.95$, $SD_{cong} = 1.60$, $M_{incong} = 3.91$, $SD_{incong} = 1.60$) as compared to the

¹⁰ As in Experiment 1, additional mixed-model analyses revealed that the STT effect significantly varied across traits, $\chi^2 = 76.89$, $p < .001$.

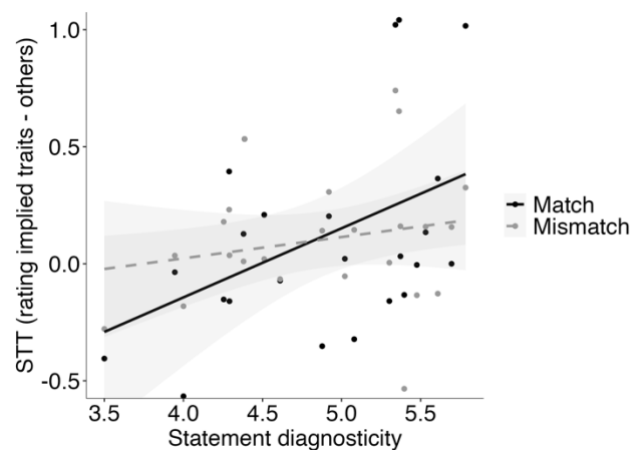
‘mismatch’ condition ($M_{implied} = 4.02$, $SD_{implied} = 1.62$, $M_{cong} = 3.92$, $SD_{cong} = 1.59$, $M_{incong} = 3.90$, $SD_{incong} = 1.60$), $B = 0.01$ (0.07), $t(37.10) = 0.19$, $p = .85$, $BF_{01} = 12.37$. When adding the diagnosticity score in the mixed-model, it did not significantly moderate the STT effect on its own, $B = 0.19$ (0.10), $t(22.09) = 1.84$, $p = .08$. However, there was a three-way interaction between STT, match/mismatch information, and statement diagnosticity, $B = 0.22$ (0.10), $t(45.76) = 2.23$, $p = .03$, indicating a joint impact of both moderators on STT (see Figure 2). A first way to interpret this is that the interaction between STT and validity depend on statement diagnosticity. For high diagnostic traits, STT was descriptively but not statistically larger for ‘match’ than ‘mismatch’ trials, $B = 0.15$ (0.09), $t(47.50) = 1.71$, $p = .094$. For low diagnostic traits, the interaction between STT and match/mismatch information was also not significant, $B = -0.13$ (0.09), $t(47.19) = 1.42$, $p = .16$, but descriptively in the opposite direction. As can be seen in Figure 2, the STT effect descriptively, but not significantly, $B = -0.07$ (0.12), $t(25.73) = 0.62$, $p = .54$, reversed for low diagnostic statements in the ‘match’ condition. Another way to present the three-way interaction is to say that the interaction between STT and diagnosticity depended on validity. In the ‘match’ condition, STT was moderated by statement diagnosticity, $B = 0.29$ (0.13), $t(22.12) = 2.21$, $p = .037$, so that the higher the statement diagnosticity the larger the STT. In ‘mismatch’ condition, this interaction was not significant, $B = 0.08$ (0.09), $t(25.29) = 0.83$, $p = .41$.

Finally, no residual STT emerged for low diagnostic statements (-1SD), $B = -0.01$ (0.10), $t(26.22) = 0.13$, $p = .90$ (when considering both ‘match’ and ‘mismatch’ conditions) while STT did emerge for high diagnostic statements (+1SD), $B = 0.23$ (0.10), $t(21.87) = 2.37$, $p = .02$. Similarly, no residual STT emerged in the ‘mismatch’ condition, $B = 0.10$

(0.06), $t(31.03) = 1.60$, $p = .12$ (when considering all statements) and it also did not emerge in ‘match’ condition, $B = 0.11$ (0.08), $t(22.32) = 1.37$, $p = .18$.¹¹

Figure 2

STT effect in the rating task (rating correct - false implied scores) at the trait level, as a function of the validity information (match vs. mismatch) and statement diagnosticity in Experiment 3



Note. The STT effect is represented by the difference in ratings for the implied trait (i.e., the trait implied by the behavioral statement of the communicator) and for the non-implied traits (i.e., evaluatively congruent and incongruent traits implied by other behavioral statement of other communicators). A difference in ratings above zero represents an STT effect and below zero a reversed STT effect. Grey areas represent the 95% confidence intervals.

Discussion

Again, we did not observe clear effects of each moderator separately but a clear joint impact in the rating task: STT was strongest when conditions for making inferences about the

¹¹ In OLS regression, the residual STT effect did emerge in the opposite direction (reversed STT) for low diagnostic statements and it also emerged in the typical direction for the ‘mismatch’ condition (see Table S8).

traits of the communicator were optimal, that is, when both validity and diagnosticity were high. Once more, we found no impact of the moderators on STT in the false recognition task.

Interestingly, for low-diagnostic statements in the ‘match’ condition, the implied trait descriptively (but not significantly) received lower ratings than non-implied traits (reversed STT effect; see Figure 2). Such reversal was also (descriptively, not significantly) present in the ‘friends’ condition of Experiment 1 (cf. re-analysis in Experiment 2; see Figure 1). This suggests that when some logical bases are met for inferring the implied trait – the communicator and actor are friends, the behavioral description was provided by the communicator – but the behavioral statement is not diagnostic to attribute the implied trait to the communicator, participants might infer other traits. Consistent with the gossiping literature, alternative attributions might reflect a “what kind of person would say this” reasoning. For example, if a communicator describes someone behaving in a lazy way, participants may not attribute the trait ‘lazy’ but rather the trait ‘judgmental’, as judgmental individuals are more likely to share information about others’ laziness. Hence, participants may infer the traits of the communicator not based on the behavior of the actor but based on the behavior of the communicator, that is, the fact that the communicator intentionally described a specific behavior. This possibility was tested in Experiment 4.

Experiment 4

We first ran two pilot studies to identify possible alternative traits that can be inferred from low-diagnosticity statements. Put differently, we relied on statements that do not allow for an inference about the implied trait but that do allow for an inference of another trait. We then presented these statements in Experiment 4 and examined whether participants attribute alternative traits in the STT context. We only included the rating task to shorten the task.

Method

Deviation from the pre-registration

We pre-registered a test comparing the alternative vs. non-implied trait but we additionally tested the alternative vs. implied traits comparison and the effect of the continuous variable of statements diagnosticity.

Participants and Design

The design mirrored that of Experiment 1, with the addition of a fourth trait type (implied vs. alternative vs. non-implied positive vs. non-implied negative; within-participants). The smallest STT effect size observed in previous studies for the rating task was $dz = 0.19$. Anticipating smaller effects for alternative traits, we aimed for 300 participants, providing 80% power to detect an effect of $dz = 0.14$ (paired t -test, 5% false positive rate). We ended up with 304 participants ($M_{age} = 39.58$, $SD_{age} = 12.39$, 149 women, 154 men, and 1 responded “other”). Recruitment and pre-selection criteria remained the same, except participants were compensated with £1.40.

Materials and Procedure

The experiment was programmed using *jsPsych* (de Leeuw, 2015). The procedure followed Experiments 1-3 but differed in three ways: (1) only the rating task was included; (2) we presented the 36 statements in the exposure task but only used low-diagnosticity statements for ratings (i.e., 12 lowest diagnostic statements from Experiment 2); (3) participants rated faces on four traits: implied, non-implied positive, non-implied negative, and “alternative”.¹²

The alternative trait for each statement was identified through two pilot studies, using procedures similar to those employed for identifying implied traits attributed to actors (Van Overwalle et al., 2012; see Supplementary Materials). In the first study ($N = 41$), participants listed traits they would attribute to the communicator for the 12 lowest-diagnosticity face-

¹² We used non-implied “positive” and “negative” terminology because a non-implied trait could be evaluatively congruent with an implied trait (e.g., cautious) but evaluatively incongruent with the corresponding alternative trait (e.g., critical).

statement pairs (see Table S5). In the second study ($N = 50$), participants rated communicators on the four most cited traits (excluding the implied trait, if cited). The highest-rated trait was selected as the “alternative” trait.

Results

We excluded one participant with zero rating variance and used the same analytical strategy with adapted contrast codes. First, we tested the crucial effect of the alternative trait compared to non-implied congruent and incongruent traits (C1: implied = 0, alternative = 2/3, non-implied congruent = -1/3, non-implied incongruent = -1/3) and the typical STT effect (C2: implied = 2/3, alternative = 0, non-implied congruent = -1/3, non-implied incongruent = -1/3). Second, we explored valence effects by comparing ratings for non-implied congruent vs. incongruent traits (C3: implied = 0, alternative = 0, non-implied congruent = 1/2, non-implied incongruent = -1/2). Congruency tests were conducted separately for implied and alternative traits, as some traits were congruent for one but incongruent for the other. Third, we explored whether participants attributed higher ratings to the alternative vs. implied traits (C4: implied = 1/2, alternative = -1/2, non-implied positive = 0, non-implied negative = 0) and tested if this effect varied as a function of the continuous diagnosticity score.

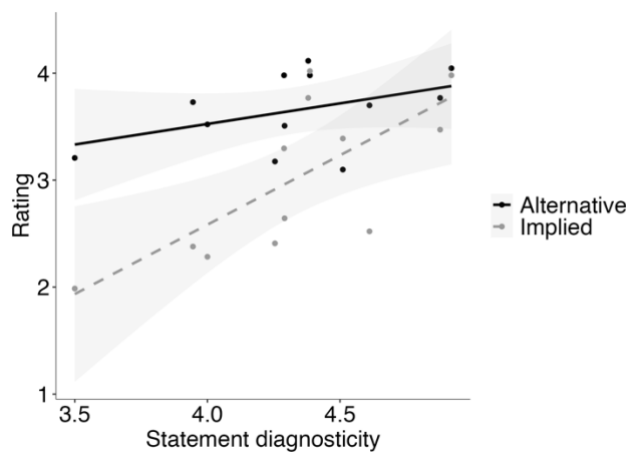
Ratings for alternative traits ($M_{alternative} = 3.65$, $SD_{alternative} = 1.37$) were higher than for non-implied congruent and incongruent traits considered jointly (contrast C1; $M_{pos} = 3.53$, $SD_{pos} = 1.34$, $M_{neg} = 3.73$, $SD_{neg} = 1.58$), $B = 0.52$ (0.09), $t(11.00) = 5.69$, $p < .001$. Ratings for implied traits ($M_{implied} = 3.01$, $SD_{implied} = 1.61$) did not significantly differ from ratings for non-implied congruent and incongruent, $B = -0.12$ (0.19), $t(11.49) = 0.61$, $p = .56$.¹³ The difference between evaluatively congruent and incongruent traits (contrast C3) was not significant when considering evaluative congruence for both alternative traits, $B = 0.44$ (0.21), $t(11.20) = 2.11$, $p = .06$, and implied traits, $B = -0.05$ (0.24), $t(11.07) = 0.19$, $p = .85$.

¹³ The residual STT effect emerged in the opposite direction (reversed STT) in OLS regression (see Table S8).

Finally, alternative traits received higher ratings as compared to implied traits, $B = 0.64$ (0.16), $t(11.99) = 3.89$, $p = .002$ (contrast C4). When adding the diagnosticity score from Experiment 2, statement diagnosticity moderated the observed difference between alternative and implied traits, $B = -0.91$ (0.34), $t(10.00) = 2.68$, $p = .023$. Specifically, the lower the diagnosticity for the implied trait, the larger the advantage for alternative traits (see Figure 3). For very low diagnostic traits (-1SD), ratings for alternative traits were higher than for implied traits, $B = 0.98$ (0.18), $t(10.69) = 5.32$, $p < .001$. For higher diagnostic traits (+1SD), this difference was not significant, $B = 0.29$ (0.18), $t(10.70) = 1.59$, $p = .14$.

Figure 3

Average ratings for each trait, as a function of the type of trait (alternative vs. implied) and statement diagnosticity for the implied trait in Experiment 4



Note. Grey areas represent the 95% confidence intervals.

Discussion

In Experiment 4, communicators were attributed traits not directly implied by the actor's behavior but by the act of communicating about that behavior (e.g., a communicator discussing someone's smart behavior was seen as "proud"). Ratings for these alternative traits exceeded those for the implied trait (e.g., "smart") and other non-implied traits (e.g.,

“creative”). The main STT effect, considering only low-diagnostic traits, did not emerge, though ratings for the implied trait increased with its diagnosticity. Higher diagnosticity did not reduce effects for alternative traits, suggesting that a single behavioral statement can evoke multiple trait attributions. In sum, impressions of the communicator extend beyond the actor’s behavior, reflecting attributions in terms of “what kind of person would say this”.

General Discussion

We explored several implications of the idea that STT effects are due to attributional processes. In Experiments 1-3, we tested three moderators: relational information, diagnosticity, and validity. Contrary to our expectations, none had a clear, replicable effect in isolation. However, STT in the rating task consistently depended on their joint influence. This aligns with an attributional perspective: participants infer that the communicator possesses the implied trait only when they are similar to the actor (i.e., friends), the statement is diagnostic, and the communicator formulated it. If any of these conditions are unmet, logical grounds for making trait inferences are undermined. When statements had low diagnosticity, the STT effect descriptively (but not significantly) reversed (see ‘friends’ and ‘match’ conditions in Experiments 1 and 3). Experiment 4 showed that in such cases, participants attributed alternative traits – derived not from the described behavior but from the communicator’s act of sharing it.

Across Experiments 1-4, no residual STT effects appeared in the rating task under conditions unfavorable to attribution processes (e.g., ‘enemies’ condition, low-diagnostic statements, or ‘mismatch’ condition) while STT still emerged under these illogical conditions in the false recognition task. Below, we discuss empirical and theoretical contributions.

Empirical contributions

A key contribution is clarifying STT as an empirical phenomenon. On the one hand, the rating task showed that attributing the implied trait to the communicator – the STT effect

– relies on the combination of relational information, diagnosticity, and validity.

Diagnosticity was pivotal, with high-diagnosticity statements amplifying STT and strengthening the effect of other attributional moderators. Future STT research should prioritize such statements to enhance methodological power.

On the other hand, trait attribution in STT stems from the act of communication itself. Impressions depend on the communicator's choice to highlight specific behaviors. When a described behavior suggests a trait applicable to both the actor and communicator (e.g., "S/he watched their neighbor's house" implying nosiness for both actor and communicator), classic STT occurs. However, some behaviors do not logically transfer the implied trait to the communicator (e.g., "S/he took the elevator up one flight" implies laziness to the actor but not the communicator). In this case, an alternative trait (e.g., judgmental) are attributed, based on the communicator's intent (e.g., criticizing the actor). Overall, our findings suggest that STT is an instance of *spontaneous attribution based on a communicative intent*, rather than a fixed *trait transference* involving the trait directly implied by the described behavior.

The fact that STT is a function of the communicator's behavior suggests STT closely resembles other attributional effects like gossiping and Spontaneous Trait Inference (STI). In gossiping, discussing someone's immoral behavior enhances the communicator's positive impression (Peters & Kashima, 2015). In STI, actors describing their own actions (e.g., "I helped a lady cross the street") create trait-based impressions of themselves (e.g., helpful). Both effects stem from logical reasoning about the target's intentions (e.g., a communicator warning about others' immorality; an actor helping others). Thus, STT involves attributions based on the communicator's intent, as in gossiping, but we extend this by showing that communicative acts can elicit specific trait attributions. Experiment 4 confirmed these attributions were not merely valence-based, as no valence congruency effect emerged. Like

STI, STT involves inferring traits from behavior, but it focuses on communication about others' behavior.

Notably, attributional moderators influenced STT in the direct rating task but not the indirect false recognition task. On the one hand, the moderation effects observed in the rating task further support the idea that STT, as EC, is not simply a product of co-occurrence, but depends heavily on context. On the other hand, the direct/indirect dissociation mirrors EC research, where relational information typically has a larger effect on direct measures (e.g., self-reported liking) than on indirect ones (e.g., Implicit Association Tests), which would primarily capture stimulus co-occurrence (Hu et al., 2017; Moran & Bar-Anan, 2013; Zanon et al., 2014; but see Hughes et al., 2019; Moran, 2024). The robustness of our analyses and large sample sizes reinforces these findings. We revisit this point in the next section.

Finally, our data did not show a residual STT effect in the rating task, unlike previous studies where such effects persisted despite instructions to ignore them (Carlston & Skowronski, 2005) or when communicator-statement pairings were said to be random (Skowronski et al., 1998). This discrepancy may reflect methodological differences. For example, earlier participants might have doubted or reinterpreted instructions (e.g., that communicators were assigned randomly to statements), allowing rule-based attributions to influence STT. Of note, the use of mixed models in our work, compared to OLS regression in prior work, does not seem to explain the difference in residual STT (see Table S8). Overall, while we found no residual STT effect, given the evidence for residual STT effects in earlier studies, more research is needed to settle this issue.

Theoretical Contributions

Results from the rating task, showing moderator effects without residual STT, challenge the view that STT stems solely from unqualified associations between the communicator and implied trait (e.g., Skowronski et al., 1998). Stronger attributions for

alternative traits further suggest participants infer communicator traits based on the act of communication itself. When a rule-based reasoning in terms of communicator-actor similarity is plausible (high-diagnostic statements; heuristic: “people discuss behaviors they approve of or exhibit”), the implied trait is attributed to the communicator. When this reasoning fails (low-diagnostic traits), participants rely on alternative reasoning (heuristic: “what kind of person would say that?”). One might argue that effects for the alternative traits stem from different mechanisms than effects for the implied traits. Hence, the former might not inform us about the latter. However, such an account is more parsimonious in that a single attributional mechanism (i.e., inferences based on the communicator’s behavior) can explain both classic STT in case of high-diagnostic statements (and how it is moderated) and reversed STT in case of low-diagnostic statements (with attributions on alternative traits).

The discrepancy between false recognition and rating tasks can be explained in at least two ways. First, from an attributional perspective, propositional information may be less relevant for indirect tasks like false recognition than for direct tasks like rating (Bading et al., 2019). Indeed, moderators affecting logical attribution should only influence tasks requiring such reasoning. In false recognition, participants judge whether a trait was linked to a communicator, not whether they possess it, making relational and truth information irrelevant. One could even argue that false recognition does not index impression formation at all. Research on false memories shows that implied lures (e.g., unrepresented but implied words; Roediger & McDermott, 1995) can distort recall, suggesting that false recognition effects may stem from memory processes rather than impression formation. Thus, while the false recognition task is a valuable tool for assessing spontaneous encoding of a trait (i.e., remembering that the person *goes together* with the trait), it may be inefficient to capture impression formation (i.e., whether the communicator *possesses* the trait). As a result, task differences likely reflect variations in how well they capture impression formation vs. mere

trait memory. Second, STT effects may involve both propositional and associative processes. Dual-process models suggest that direct tasks reflect propositional (attributional) reasoning, while indirect tasks rely on associative mechanisms (e.g., McConnell & Rydell, 2014). For STT, this implies that rating task effects stem from rule-based, attributional reasoning, whereas false recognition effects result from associative mechanisms.

One way to distinguish between these explanations is to replicate our findings with indirect measures known to capture relational or validity information (e.g., Cummins & De Houwer, 2022). If propositional moderators influence other indirect impression formation measures, their absence in false recognition may be task-specific. Alternatively, studies could examine whether false recognition reflects impression formation or trait memory by testing whether performance aligns with false memory moderators (e.g., Todorov & Uleman, 2003) rather than impression formation moderators.

Finally, our findings suggest a similarity between processes underlying STT and EC, as both appear – at least partly – driven by attributional/propositional processes. This theoretical alignment bridges research domains often treated separately, such as impression formation and evaluative learning (De Houwer et al., 2019; Rougier et al., 2023).

Limitations

Table 1 summarizes our work’s limitations, addressing issues from individual experiments and proposing solutions or future research directions.

Table 1

List of limitations and solutions as a function of experiment

Exp.	Limitation	Way(s) to address the limitation
1, 3	The main STT effect in the rating task emerged only when between-traits variability was controlled for in the statistical analysis (i.e., OLS regression).	This was likely due to strong variation in STT as a function of other trait-related information (diagnosticity). To facilitate the emergence of the main STT effect, these variables should be absent (e.g.,

- Preliminary Experiment) or they should be set at the level most likely to produce a STT effect (e.g., highly diagnostic traits).
- 3 The moderation of the STT effect by validity information was larger for highly diagnostic statements, however it did not reach significance. Future work should test more directly whether the moderation of validity on STT can be found when increasing both methodological power (i.e., by only using highly diagnostic traits) and statistical power (i.e., by increasing the number of participants).
- 1,3 The relational and validity manipulations lacked realism (i.e., information was provided via labels during the exposure phase), questioning the ecological validity of the observed results. Future work should improve the ecological validity of relational and validity manipulations, as well as the exposure phase (i.e., pairing procedure). For instance, participant could see an individual (in person or on video) talking about the behavior of a friend or an enemy.
- 4 We tested the effect on alternative traits while only including low diagnostic statements. Future work could investigate if the observed effect on alternative traits (as well as the larger effect on alternative traits as compared to implied traits) is also observed on high diagnostic statements. We anticipate the effect on alternative traits to also emerge in the rating task for high diagnostic statement, but more likely to the same extent as the effect for implied traits. Indeed, results of Experiment 4 (see Figure 3) suggest that effects on distinct traits can emerge in parallel, with the difference between alternative and implied traits decreasing when statement diagnosticity increases.
- 1, 3, 4 When considering exploratory variables, we did not consistently observe an effect of self-reported use of rule-based reasoning (e.g., similarity heuristic between the actor and the communicator) on STT or the moderation of STT (Table S3 in Supplementary Materials). Experiment 3 suggests that participants rely on similarity heuristics in that the STT effect was larger when participants thought that the communicator and the actor were alike. Future work should test more systematically whether similarity/endorsement heuristics can account for STT effects. We anticipate this to be the case for high-diagnostic traits but not for low-diagnostic traits where other heuristics should be used.

1-4	We relied on samples of English-speaking, US citizens participants recruited via Prolific Academic.	Future work should try to replicate our results to test if both STT and its moderators can be generalized to different cultural contexts and languages. It could be that heuristics and inferences underlying trait attribution are not universally shared.
2, 3	Some of the key interactions yielded p -values in the range of .01 to .05 (e.g., STT by diagnosticity and STT by relation by diagnosticity interactions reported in Experiment 2). While these meet the conventional threshold for statistical significance, p -values in this range should be interpreted with appropriate caution.	We encourage future work to replicate these effects – ideally increasing statistical power (larger samples of participants and behavioral statements) and methodological power (by increasing variability in statements diagnosticity with more extreme high/low diagnostic statement).

Open Questions and Future Directions

First, we did not consistently observe the negativity bias typical of attributional processes. While the Preliminary Experiment showed a stronger STT effect for negative statements in false recognition, this pattern was absent elsewhere. Interestingly, in attributional terms, positive statements should facilitate trait attribution, as reporting positive behavior implies approval, suggesting the communicator shares the trait. In contrast, negative behaviors may be seen as criticism, reducing trait attribution. This aligns with Experiment 1, where STT and relational information interacted based on statement valence – positive statements enhanced STT more in the ‘friends’ than ‘enemy’ condition. Future research could directly test how communicator approval of statements influences these effects.

Second, other forms of statement diagnosticity may moderate STT or other impression formation effects. For instance, the extent to which a behavior implies that the *actor* possesses a trait (‘actor-based’ diagnosticity; see Footnote 3) should not influence STT but it should moderate STI. Although all the behavioral statements we used were high in actor-based diagnosticity – based on prior pretests (e.g., Uleman, 1988; see Table S1) – we tested in an exploratory analysis whether actor-based diagnosticity moderated STT. It did not

(see Table S6), supporting the view that STT is not driven by mere communicator-trait binding, but reflects attributional reasoning based on whether the communicative act warrants trait inference. In STI, however, higher actor-based diagnosticity should increase STI, as this one directly arises from the actor's behavior.

Conclusion

In their review, Uleman et al. (2008) posed the challenge of integrating STI, STT, EC, and spontaneous meta-inferences. This work takes initial steps toward this integration by testing predictions of attributional/propositional processes across these domains. Our findings suggest that STT effects arise from rule-based attributions about the communicator, based on their act of communicating about others' behavior.

References

- Bading, K., Stahl, C., & Rothermund, K. (2019). Why a standard IAT effect cannot provide evidence for association formation: the role of similarity construction. *Cognition and Emotion, 34*(1), 128–143. <https://doi.org/10.1080/02699931.2019.1604322>
- Baeyens, F., Eelen, P., Crombez, G., & Van Den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30*(2), 133–142. [https://doi.org/10.1016/0005-7967\(92\)90136-5](https://doi.org/10.1016/0005-7967(92)90136-5)
- Bainbridge, W. A., Isola, P., & Oliva, A. (2013). The intrinsic memorability of face photographs. *Journal of Experimental Psychology: General, 142*(4), 1323–1334. <https://doi.org/10.1037/a0033872>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious Mixed Models*. <https://doi.org/10.48550/ARXIV.1506.04967>
- Brown, R. D., & Bassili, J. N. (2002). Spontaneous trait associations and the case of the superstitious banana. *Journal of Experimental Social Psychology, 38*(1), 87–92. <https://doi.org/10.1006/jesp.2001.1486>
- 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*(6), 884–898. <https://doi.org/10.1037/0022-3514.89.6.884>
- Correll, J., Mellinger, C., & Pedersen, E. J. (2021). Flexible approaches for estimating partial eta squared in mixed-effects models with crossed random factors. *Behavior Research Methods, 54*(4), 1626–1642. <https://doi.org/10.3758/s13428-021-01687-2>
- 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(5), 677–690.

<https://doi.org/10.1177/0146167206298567>

Cummins, J., & De Houwer, J. (2022). Are Relational Implicit Measures Sensitive to Relational Information? *Collabra: Psychology*, 8(1), 38621.

<https://doi.org/10.1525/collabra.38621>

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology*, 10(2), 230–241.

<https://doi.org/10.1017/s1138741600006491>

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, 37(1), 1–20.

<https://doi.org/10.3758/LB.37.1.1>

De Houwer, J. (2018). Propositional models of evaluative conditioning. *Social Psychological Bulletin*, 13(3), e28046. <https://doi.org/10.5964/spb.v13i3.28046>

De Houwer, J., Gawronski, B., & Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *European Review of Social Psychology*, 24(1), 252–287. <https://doi.org/10.1080/10463283.2014.892320>

De Houwer, J., Richetin, J., Hughes, S., & Perugini, M. (2019). On the assumptions that we make about the world around us: A conceptual framework for feature transformation effects. *Collabra: Psychology*, 5(1), 43. <https://doi.org/10.1525/collabra.229>

De Houwer, J., Van Dessel, P., & Moran, T. (2020). Attitudes beyond associations: On the role of propositional representations in stimulus evaluation. In *Advances in Experimental Social Psychology* (Vol. 61, pp. 127–183). Elsevier.

<https://doi.org/10.1016/bs.aesp.2019.09.004>

- De Houwer, J., Van Dessel, P., & Moran, T. (2021). Attitudes as propositional representations. *Trends in Cognitive Sciences*, 25(10), 870–882.
<https://doi.org/10.1016/j.tics.2021.07.003>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12.
<https://doi.org/10.3758/s13428-014-0458-y>
- Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3), 285–290. <https://doi.org/10.1037/h0033731>
- Fiedler, K., & Unkelbach, C. (2011). Evaluative conditioning depends on higher order encoding processes. *Cognition & Emotion*, 25(4), 639–656.
<https://doi.org/10.1080/02699931.2010.513497>
- Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, 27(2), 222–248.
<https://doi.org/10.1521/soco.2009.27.2.222>
- Heider, F. (1958). *The Psychology of Interpersonal Relations* (0 ed.). Psychology Press.
<https://doi.org/10.4324/9780203781159>
- Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, 54(2), 556–573. <https://doi.org/10.3758/s13428-019-01283-5>
- Högden, F., & Unkelbach, C. (2021). The role of relational qualifiers in attribute conditioning: Does disliking an athletic person make you unathletic? *Personality and Social Psychology Bulletin*, 47(4), 643–656.
<https://doi.org/10.1177/0146167220945538>
- Hu, X., Gawronski, B., & Balas, R. (2017). Propositional versus dual-process accounts of evaluative conditioning: I. The effects of co-occurrence and relational information on

- implicit and explicit evaluations. *Personality and Social Psychology Bulletin*, 43(1), 17–32. <https://doi.org/10.1177/0146167216673351>
- Hughes, S., Ye, Y., Van Dessel, P., & De Houwer, J. (2019). When People Co-occur With Good or Bad Events: Graded Effects of Relational Qualifiers on Evaluative Conditioning. *Personality and Social Psychology Bulletin*, 45(2), 196–208. <https://doi.org/10.1177/0146167218781340>
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54–69. <https://doi.org/10.1037/a0028347>
- Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology*, 31(1), 457–501. <https://doi.org/10.1146/annurev.ps.31.020180.002325>
- Kruse, F., & Degner, J. (2021). Spontaneous state inferences. *Journal of Personality and Social Psychology*, 121(4), 774–791. <https://doi.org/10.1037/pspa0000232>
- Kruse, F. & Degner, J. (2023). Do faces matter? On the interaction of face- and behavior-inferences in spontaneous impression formation. *Unpublished Manuscript*.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13). <https://doi.org/10.18637/jss.v082.i13>
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. (Cambridge: Cambridge University Press.).
- Marien, H., Custers, R., Hassin, R. R., & Aarts, H. (2012). Unconscious goal activation and the hijacking of the executive function. *Journal of Personality and Social Psychology*, 103(3), 399–415. <https://doi.org/10.1037/a0028955>

- McConnell, A. R., & Rydell, R. J. (2014). The systems of evaluation model: A Dual-systems approach to attitudes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 204–218). The Guilford Press.
- Moors, A. (2016). Automaticity: Componential, Causal, and Mechanistic Explanations. *Annual Review of Psychology*, *67*(1), 263–287. <https://doi.org/10.1146/annurev-psych-122414-033550>
- Moran, T. (2024). The effect of irrelevant pairings on evaluative responses. *Journal of Experimental Social Psychology*, *112*, 104602. <https://doi.org/10.1016/j.jesp.2024.104602>
- Moran, T., & Bar-Anan, Y. (2013). The effect of object–valence relations on automatic evaluation. *Cognition & Emotion*, *27*(4), 743–752. <https://doi.org/10.1080/02699931.2012.732040>
- Moran, T., Bar-Anan, Y., & Nosek, B. A. (2017). The effect of the validity of co-occurrence on automatic and deliberate evaluations: Validity and automatic evaluations. *European Journal of Social Psychology*, *47*(6), 708–723. <https://doi.org/10.1002/ejsp.2266>
- Moran, T., Nudler, Y., & Bar-Anan, Y. (2023). Evaluative conditioning: Past, present, and future. *Annual Review of Psychology*, *74*(1), 245–269. <https://doi.org/10.1146/annurev-psych-032420-031815>
- Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2015). Package ‘bayesfactor’. URL [Http://Cran/r-Project.org/Web/Packages/BayesFactor/BayesFactor](http://Cran/r-Project.org/Web/Packages/BayesFactor/BayesFactor).
- Peters, K., & Kashima, Y. (2015). Bad habit or social good? How perceptions of gossip morality are related to gossip content. *European Journal of Social Psychology*, *45*(6), 784–798. Portico. <https://doi.org/10.1002/ejsp.2123>

- Peters, K. R., & Gawronski, B. (2011). Are we puppets on a string? Comparing the impact of contingency and validity on implicit and explicit evaluations. *Personality and Social Psychology Bulletin, 37*(4), 557–569. <https://doi.org/10.1177/0146167211400423>
- Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814. <https://doi.org/10.1037/0278-7393.21.4.803>
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. <https://doi.org/10.3758/PBR.16.2.225>
- Rougier, M., De Houwer, J., Richetin, J., Hughes, S., & Perugini, M. (2023). From halo to conditioning and back again: Exploring the links between impression formation and learning. *Collabra: Psychology, 9*(1). <https://doi.org/10.1525/collabra.84560>
- Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91*(6), 995–1008. <https://doi.org/10.1037/0022-3514.91.6.995>
- Skowronski, J. J., Carlston, D. E., Mae, L., & Crawford, M. T. (1998). Spontaneous trait transference: Communicators take on the qualities they describe in others. *Journal of Personality and Social Psychology, 74*(4), 837–848. <https://doi.org/10.1037/0022-3514.74.4.837>
- 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*(5), 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*(6), 549–562. [https://doi.org/10.1016/S0022-1031\(03\)00059-3](https://doi.org/10.1016/S0022-1031(03)00059-3)

Uleman, J. S. (1988). Trait and gist inference norms for over 300 potential trait-implying sentences. *Unpublished raw data*.

Uleman, J. S., Adil Saribay, S., & Gonzalez, C. M. (2008). Spontaneous Inferences, Implicit Impressions, and Implicit Theories. *Annual Review of Psychology*, *59*(1), 329–360.
<https://doi.org/10.1146/annurev.psych.59.103006.093707>

Unkelbach, C., & Fiedler, K. (2016). Contrastive CS-US relations reverse evaluative conditioning effects. *Social Cognition*, *34*(5), 413–434.
<https://doi.org/10.1521/soco.2016.34.5.413>

Van Overwalle, F., Van Duynslaeger, M., Coomans, D., & Timmermans, B. (2012). Spontaneous goal inferences are often inferred faster than spontaneous trait inferences. *Journal of Experimental Social Psychology*, *48*(1), 13–18.
<https://doi.org/10.1016/j.jesp.2011.06.016>

Wells, B. M., Skowronski, J. J., Crawford, M. T., Scherer, C. R., & Carlston, D. E. (2011). Inference making and linking both require thinking: Spontaneous trait inference and spontaneous trait transference both rely on working memory capacity. *Journal of Experimental Social Psychology*, *47*(6), 1116–1126.
<https://doi.org/10.1016/j.jesp.2011.05.013>

Wyer, R. S., Budesheim, T. L., & Lambert, A. J. (1990). Cognitive representation of conversations about persons. *Journal of personality and social psychology*, *58*(2), 218. <https://doi.org/10.1037//0022-3514.58.2.218>

Zanon, R., De Houwer, J., Gast, A., & Smith, C. T. (2014). When does relational information influence evaluative conditioning? *Quarterly Journal of Experimental Psychology*, *67*(11), 2105–2122. <https://doi.org/10.1080/17470218.2014.907324>

Supplementary Material

Preliminary Experiment (STT replication)

This preliminary experiment sought to replicate the STT effect based on previous work (e.g., Goren & Todorov, 2009). We relied on the false recognition paradigm (indirect task; Todorov & Uleman, 2002) and on a rating task (direct task requiring participants to form an impression of individuals; Carlston & Skowronski, 2005; McCarthy et al., 2018).

Method

Transparency and Openness. We slightly deviated from the pre-registration by opting for mixed-models as the primary analysis instead of OLS regressions (mixed-models were pre-registered as an addition).

Participants and Design. To estimate our sample size, we relied on a low estimate of the effect size reported in previous studies using a similar design and measures ($d_z = 0.30$; Carlston & Skowronski, 2005; Goren & Todorov, 2009; Skowronski et al., 1998; Wells et al., 2011). We opted for 345 participants, which provided us with 80% of power to detect a STT effect of similar size (two-tailed t -test for two paired samples) with a 5% false-positive rate. Participants ($M_{age} = 35.21$, $SD_{age} = 13.45$, 220 women, 110 men, 13 responded “other”, 1 preferred not to say, and 1 responded “none of the previous options”) were recruited via Prolific Academic (www.prolific.co) and took part in exchange for £2.35. Inclusion criteria were the same as in Experiment 1. We relied on a 2 (*Type of trait in the false recognition task*: correct implied vs. false implied) x 3 (*Type of trait in the rating task*: implied vs. evaluatively congruent vs. evaluatively incongruent) within participant design.

Materials and Procedure. Materials and procedure were the same as in Experiment 1 except for the following exceptions. First, we did not manipulate the relation between the communicator and the actor – the communicator and actor were always said to be acquaintances. Thus, we did not provide any information in this regard in the instructions, we

did not present any ‘friends’ or ‘enemies’ labels in the exposure phase or include any ‘memory of relation’ or exploratory questions related to the relation. Second, photograph-statement pairs were presented for 6 seconds (instead of 8). Third, because of a programming error, some statements were only presented with a specific gender or under a specific type of trait condition. Moreover, statements corresponding to filler traits were only presented with female faces, and statements corresponding to false implied traits with male faces. Although we see no reasons why this could have explained the observed pattern of results, this programming error was corrected for the following experiments. Finally, as an additional exploratory question, we asked participants to what extent they thought the person on the picture and the acquaintance were alike (e.g., have similar personalities) on a Likert scale from 1 (*not at all*) to 7 (*extremely*).

Results

We excluded nine participants having zero variance in their responses in the rating task (final sample $N = 336$). Contrast codes and models were the same as in Experiment 1 (except there was no ‘relation’ variable).

False Recognition Task. The likelihood of ‘yes’ responses was higher for correct implied ($M = 0.46$, $SD = 0.22$) than for false implied traits ($M = 0.34$, $SD = 0.22$), $B = 0.62$, $SE = 0.07$, $z = 8.50$, $p < .001$. This effect was larger for negative ($M_{correct} = 0.21$, $SD_{correct} = 0.13$, $M_{false} = 0.14$, $SD_{false} = 0.13$) than for positive traits ($M_{correct} = 0.25$, $SD_{correct} = 0.13$, $M_{false} = 0.20$, $SD_{false} = 0.12$), $B = -0.24$, $SE = 0.11$, $z = 2.21$, $p = .027$. Moreover, the STT effect was also larger when the false recognition task came first ($M_{correct} = 0.39$, $SD_{correct} = 0.20$, $M_{false} = 0.25$, $SD_{false} = 0.18$) rather than second ($M_{correct} = 0.53$, $SD_{correct} = 0.21$, $M_{false} = 0.43$, $SD_{false} = 0.21$), $B = 0.33$, $SE = 0.13$, $z = 2.57$, $p = .010$.¹⁴

¹⁴ The STT effect remained significant when controlling for both the traits valence, $B = 0.62$, $SE = 0.07$, $z = 9.07$, $p < .001$, and the task order, $B = 0.62$, $SE = 0.07$, $z = 8.52$, $p < .001$.

Rating Task. We observed a higher rating for implied traits ($M = 3.97$, $SD = 1.63$) than for evaluatively congruent and incongruent traits considered jointly, $B = 0.18$, $SE = 0.05$, $t(35.60) = 3.68$, $p < .001$ (contrast C1; $M_{cong} = 3.78$, $SD_{cong} = 1.57$, $M_{incong} = 3.80$, $SD_{incong} = 1.59$). Ratings for evaluatively congruent and incongruent traits did not significantly differ, $B = -0.02$, $SE = 0.06$, $t(23.66) = 0.36$, $p = .72$ (contrast C2).

Discussion

This preliminary experiment successfully replicated the STT effect through false recognition and rating tasks, revealing participants' tendencies to attribute the trait implied by the behavioral information to the communicator. However, variable crossings (between traits, type of traits, and faces' gender) were uneven. Experiment 1 addressed this issue while also exploring potential moderating effects of relational information on the STT phenomenon.

Instructions used in Experiment 2 and 3 (validity manipulation)

“Study on person perception. Before the study, we asked a number of people to describe the behavior of somebody they know (i.e., an acquaintance). During the study, on each trial, you will see a **photograph of a person and the description of a behavior**. On some trials, you will also see the word ‘**MATCH**’. In this case, **the behavioral description is about an acquaintance of the person on the photograph**. It means that the person on the photograph described his/her acquaintance in that way. On other trials, you will see the word ‘**MISMATCH**’. In that case, the behavioral description was **not provided by the person on the photograph** but by one of the other persons who was asked to describe an acquaintance.

The match/mismatch information is displayed at the top. The photographed person is presented below on the left, together with the behavioral description below on the right.

‘**MATCH**’ means that the behavioral description was provided by the person on the photograph about an acquaintance of that person. ‘**MISMATCH**’ means that the behavior description was not provided by the person on the photograph but by somebody else.”

Procedure and data analysis of Pilot Studies used in Experiment 4

Two pilot studies aimed at collecting alternative trait inference for low-diagnostic traits (i.e., statements for which the implied trait is not diagnostic to be inferred in the communicator). In a first pilot study we asked participant to list traits that could be inferred on the communicators (open-ended format) and in a second pilot study participants rated the communicator on the traits that were the most frequently listed in the first pilot. We recruited 41 participants ($M_{age} = 40.73$, $SD_{age} = 15.65$; 26 women and 15 men) in Pilot Study 1 and 50 participants ($M_{age} = 37.86$, $SD_{age} = 12.69$; 34 women and 16 men) in Pilot Study 2. Participants were recruited on Prolific Academic and were retributed 0.70£ and 0.60£, respectively. Participants were all Americans, spoke English as a first language, and had a minimum of 98% approval rate.

Pilot Study 1

The study was programmed and administered via JsPsych. After signing an informed consent, participants received similar instructions as in a typical STT procedure: they were told that we would see a photograph of a person (the communicator) and the description of an acquaintance's behavior (the actor). Participants were asked to imagine what the communicator could be like and, specifically, the personality traits the person might possess. Participants were asked to be as creative as possible and avoid using the same personality trait at different trials. Participants were then presented with 12 trials involving the 12 trait-implying statements that received the lowest diagnosticity scores in Experiment 3 (see Table S5). Half of trials involved photograph of female faces, the other half of male faces and statements were gender-reversed. For each statement, participants were asked to indicate the personality traits the communicator might possess. They could list four traits and were asked to indicate at least one trait. At the end of the study, participants answered exploratory questions and demographics.

After collecting participants' responses, we aimed at gathering the most frequently listed traits that differed from the implied ones (e.g., we did not consider "smart" or intelligence-related traits for the statement "He took his first calculus course when he was 12 years old."). To this aim, we relied on ChatGPT-4 (OpenAI). For each statement, we provided ChatGPT the listed traits and asked to "indicate the traits that are the most often listed in the following list by regrouping the traits that are conceptually close". Then, we selected the four most frequently listed traits for each statement while excluding the implied traits.

Pilot Study 2

Participants received similar instructions as in the first pilot study except that this time they were asked to rate the communicator on a series of four personality traits. Participants were exposed to 12 photograph-statement pairs following the same structure as in Pilot Study 1. At each trial, they were asked to indicate the extent to which the communicator possess each of the four traits listed (from 0 = *not at all* to 10 = *very much*). Then, they answered the same exploratory and demographic questions as before.

Following on participants' ratings we computed the average rating per trait and for each statement. Based on these average values we selected the trait that received the highest rating (e.g., "proud" for the statement "He took his first calculus course when he was 12 years old."; see Table S5). The traits were then used in Experiment 4 to test whether, in an actual STT experiment, communicators were attributed those alternative traits.

Exploratory Questions in Experiments 1-4

Experiment 1. Participants answered exploratory questions to assess potential attribution processes (i.e., logical reasoning adopted) and response faking. First, they reported whether they relied on the content of the statements when answering questions about the photographed persons (response options: *yes, no, I don't know*). Second, they rated how much the “friend” or “enemy” label influenced their responses on a scale from 1 (*not at all*) to 7 (*extremely*). Third, they indicated whether they answered truthfully or faked responses to align with perceived researcher expectations (response options: *yes, no, I don't know*). Finally, participants could leave comments, completed demographic questions, were debriefed, and received payment and contact information.

Experiment 2. First, participants described what they thought the researchers aimed to achieve in the experiment (open-ended question) and answered the faking question. Second, participants indicated whether they believed the match/mismatch labels influenced their judgments about the usefulness of the information for assessing the communicator (response options: *yes, no, I don't know*). If they answered “yes,” they specified how (open-ended question). Third, participants reported whether they thought that, when the match was correct (i.e., the communicator formulated the statement), the behavioral statement was generally useful for drawing conclusions about the communicator’s personality (response options: *yes, no, I don't know*). If they answered “yes,” they specified how. Finally, participants rated, in the case of a correct match, 1) how similar they thought the communicator and the described acquaintance were (e.g., similar personalities) and 2) to what extent they believed the communicator agreed with the behavior described. Both questions were rated on a scale from 0 (*not at all*) to 10 (*totally*).

Experiment 3. Participants completed the same set of exploratory questions as in Experiment 2, with two modifications. They were not asked to elaborate (when answering

“yes”) on how the match/mismatch labels influenced their judgments or how the behavioral statement was useful for drawing conclusions. Additionally, the final question regarding the extent to which participants believed the communicator agreed with the described behavior was removed.

Experiment 4. Participants answered the same exploratory questions as in Experiment 3, excluding those related to match/mismatch labels.

Table S1

Behavioral statements used in Experiments 1-3, trait implied by each statement, type of statement, pretest frequency, valence, and reference from which the statement was selected

Reference	Behavioral statement	Implied trait	Type of statement	Pretest frequency	Valence
Uleman, 1988	S/he asked where the stars come from.	curious	Trait implying	0.94	Positive
Uleman, 1988	S/he always drove a little slower than the speed limit.	cautious	Trait implying	0.91	Positive
Uleman, 1988	S/he took the elevator up one flight.	lazy	Trait implying	0.83	Negative
Uleman, 1988	S/he was afraid the new employees wouldn't like her/him.	insecure	Trait implying	0.82	Negative
Uleman, 1988	S/he watched her/his neighbor's house to see who came and went.	nosy	Trait implying	0.80	Negative
Uleman, 1988	S/he stories made people laugh so hard they held their sides.	funny	Trait implying	0.80	Positive
Uleman, 1988	S/he took her/his first calculus course when s/he was 12 years old.	smart	Trait implying	0.80	Positive
Uleman, 1988	S/he picked out the best chocolates before the guests arrived.	selfish	Trait implying	0.78	Negative
Uleman, 1988	S/he thought s/he didn't deserve their award and praise.	modest	Trait implying	0.78	Positive
Uleman, 1988	S/he told the cashier that s/he got too much change.	honest	Trait implying	0.77	Positive
Uleman, 1988	S/he dusted and vacuumed her/his room every day.	neat	Trait implying	0.75	Negative
Uleman, 1988	S/he left a 25% tip for the waitress.	generous	Trait implying	0.74	Positive
Uleman, 1988	S/he stepped on her/his boyfriend's/girlfriend's feet during the foxtrot.	clumsy	Trait implying	0.72	Negative
Uleman, 1988	S/he didn't smoke at home while her/his roommate was trying to quit.	considerate	Trait implying	0.71	Positive
Uleman, 1988	S/he couldn't get herself/himself to greet her/his new neighbor.	shy	Trait implying	0.69	Negative
Uleman, 1988	S/he lost track of the two year old.	irresponsible	Trait implying	0.68	Negative

Uleman, 1988	S/he held a full-time job while being a full-time student.	ambitious	Trait implying	0.68	Negative
Uleman, 1988	S/he phoned for help while the others just screamed.	calm	Trait implying	0.68	Positive
Uleman, 1988	S/he took 15 minutes to find her/his car in the parking lot.	forgetful	Trait implying	0.66	Negative
Kruse & Degner, 2023	S/he needed at least two packs of handkerchiefs when s/he watched a sad movie.	emotional	Trait implying	0.83	Negative
Kruse & Degner, 2023	S/he made sure that each kid received the same amount of Halloween candy.	fair	Trait implying	0.76	Positive
Kruse & Degner, 2023	Every day s/he invented a new game for her/his niece.	creative	Trait implying	0.75	Positive
Van Overwalle et al., 2012	S/he goes every Sunday to church and says her/his prayers every night before bedtime.	religious	Trait implying	> 0.70	Positive
Van Overwalle et al., 2012	S/he never says "thank you".	impolite	Trait implying	> 0.70	Negative
Kruse & Degner, 2021	S/he was very competent at her/his job.	competent	Filler	NA	Positive
Kruse & Degner, 2021	When all her/his friends surprised her/him for her /his birthday, s/he felt loved.	loved	Filler	NA	Positive
Kruse & Degner, 2021	S/he was very competitive, especially when her/his brother was around.	competitive	Filler	NA	Positive
Kruse & Degner, 2021	S/he left the room because s/he was offended by the joke.	offended	Filler	NA	Negative
Kruse & Degner, 2021	S/he was naive enough to think that superman really could fly.	naive	Filler	NA	Negative
Kruse & Degner, 2021	When s/he could answer all the questions on the test, s/he felt proud	proud	Filler	NA	Positive
Kruse & Degner, 2021	S/he confidently walked into the interview room.	confident	Filler	NA	Positive
Kruse & Degner, 2023	S/he knew that her/his friend had lied to her/his, but s/he remained passive and did nothing about it.	passive	Filler	NA	Negative

Kruse & Degner, 2023	S/he finished university but s/he was completely lost and didn't know what to do with her/his life.	lost	Filler	NA	Negative
Kruse & Degner, 2023	S/he looked inappropriate when s/he had a laughing fit at the funeral.	inappropriate	Filler	NA	Negative
Kruse & Degner, 2023	S/he swore to be disciplined and did not eat chocolate while s/he fasted.	disciplined	Filler	NA	Negative
Kruse & Degner, 2023	S/he is chaste and didn't get intimate with him/her before they got married.	chaste	Filler	NA	Positive

Note. Full references can be found in the Reference section of the manuscript. The gender form used in the sentence (i.e., feminine or masculine) was always opposite to the gender of the communicator. The behavioral sentences were pretested within each specific referenced work to evoke the implied trait: the 'pretest free' column indicates the reported frequency of listing of the implied trait for each trait-implying behavioral sentence (filler sentences were not pretested as they explicitly mention the trait). For instance, in Van Overwalle et al. (2012), a statement was selected as a trait-implying statement if the same trait (or a close synonym) was provided by at least 70% of the participants.

Table S2

Results for the pre-registered analyses (OLS regression) for the false recognition and rating tasks as a function of the type of trait and the moderation by relation or validity (when applicable)

Experiment	Task	Effect	df	<i>t</i>- statistic	<i>p</i>	<i>Cohen's d</i>
Preliminary Experiment	False recognition	Type of trait	335	9.17	< .001	<i>dz</i> = 0.50, 95% CI [0.39; 0.61]
		Rating	C1	335	6.08	< .001
	C2		335	0.94	.35	<i>dz</i> = 0.05, 95% CI [-0.06; 0.16]
Experiment 1	False recognition	Type of trait	195	6.32	< .001	<i>dz</i> = 0.45, 95% CI [0.30; 0.60]
		Rating	Type of trait by relation	195	0.84	.40
	C1		195	2.62	.009	<i>dz</i> = 0.19, 95% CI [0.05; 0.33]
	C1 by relation		195	0.69	.49	<i>dz</i> = 0.05, 95% CI [-0.09; 0.19]
	C2	195	0.60	.55	<i>dz</i> = 0.04, 95% CI [-0.10; 0.18]	
Experiment 3	False recognition	Type of trait	196	6.89	< .001	<i>dz</i> = 0.49, 95% CI [0.34; 0.64]
		Rating	Type of trait by validity	196	1.09	.28
	C1		196	2.83	.005	<i>dz</i> = 0.20, 95% CI [0.06; 0.34]
	C1 by validity		196	0.18	.86	<i>dz</i> = 0.01, 95% CI [-0.13; 0.15]
	C2	196	1.47	.14	<i>dz</i> = 0.11, 95% CI [-0.04; 0.25]	

Note. For the false recognition task, we used as a dependent variable the rate of “yes” responses per participant and per type of trait. The variable of type of trait in the false recognition task, relation, and validity were contrast coded (false implied = -0.5, correct implied = 0.5; enemies = -0.5, friends = 0.5; match = -0.5, mismatch = 0.5) and the type of trait in the rating task was coded via two orthogonal contrast codes (quadratic contrast C1: implied = 2/3, evaluatively congruent = -1/3, evaluatively incongruent = -1/3; linear contrast C2: implied = 0, evaluatively congruent = 1/2, evaluatively incongruent = -1/2).

Table S3

Moderation of exploratory variables as a function of the effect of interest (STT vs. moderation of STT) for the false recognition and rating tasks (Preliminary Experiment and Experiments 1, 3, and 4)

Experiment	Task	Exploratory moderator	Effect moderated	df	statistic	p		
Preliminary Experiment	False recognition	Reliance statement	STT	N/A	$z = 0.86$.39		
		Reliance similarity	-	-	$z = 1.82$.07		
		Demand compliance	-	-	$z = 0.03$.98		
	Rating	Reliance statement	-	336.50	$t = 0.03$.98		
		Reliance similarity	-	336.60	$t = 0.58$.56		
		Demand compliance	-	336.53	$t = 0.95$.34		
		Experiment 1	False recognition	Memory of the relation	STT	N/A	$z = 1.54$.12
				Memory of the relation	STT by relation	-	$z = 0.41$.68
				Reliance statement	STT	-	$z = 0.33$.74
Reliance statement	STT by relation			-	$z = 0.61$.54		
Influence relation	STT			-	$z = 0.23$.82		
Influence relation	STT by relation			-	$z = 0.39$.69		
Experiment 1	Rating	Demand compliance	STT	-	$z = 1.64$.10		
		Demand compliance	STT by relation	-	$z = 1.13$.26		
		Memory of the relation	STT	13400.00	$t = 1.23$.22		
		Memory of the relation	STT by relation	11570.00	$t = 2.14$.03		
		Reliance statement	STT	671.55	$t = 0.91$.36		
		Reliance statement	STT by relation	598.40	$t = 0.25$.81		
		Influence relation	STT	13570.00	$t = 0.80$.94		
		Influence relation	STT by relation	593.60	$t = 0.41$.68		

Experiment 3	False recognition	Demand compliance	STT	671.46	$t = 0.70$.48
		Demand compliance	STT by relation	605.37	$t = 0.55$.58
		Influence labels	STT	N/A	$z = 1.38$.17
		Influence labels	STT by validity	N/A	$z = 1.12$.26
		Statement useful	STT	N/A	$z = 1.29$.20
		Statement useful	STT by validity	N/A	$z = 0.20$.84
		Communicator actor alike	STT	N/A	$z = 1.68$.093
		Communicator actor alike	STT by validity	N/A	$z = 0.11$	0.91
		Demand compliance	STT	N/A	$z = 0.45$.65
		Demand compliance	STT by validity	N/A	$z = 0.84$.40
	Rating	Influence labels	STT	196.40	$t = 2.96$.003
		Influence labels	STT by validity	761.50	$t = 0.39$.70
		Statement useful	STT	196.30	$t = 1.77$.08
		Statement useful	STT by validity	874.90	$t = 1.80$.07
Experiment 4	Rating	Communicator actor alike	STT	196.40	$t = 2.16$.03
		Communicator actor alike	STT by validity	807.10	$t = 1.03$.30
		Demand compliance	STT	196.48	$t = 0.13$.90
		Demand compliance	STT by validity	785.01	$t = 0.51$.61
		Statement useful	STT	301.00	$t = 1.62$.11
		Statement useful	Alternative effect	822.44	$t = 0.83$.41
		Communicator actor alike	STT	300.99	$t = 0.21$.83
		Communicator actor alike	Alternative effect	13750.00	$t = 0.63$.52
		Communicator agrees with statement	STT	301.03	$t = 0.33$.74
		Communicator agrees with statement	Alternative effect	820.50	$t = 0.05$.96
	Demand compliance	STT	300.99	$t = 0.27$.79	

	Demand compliance	Alternative effect	821.81	$t = 0.32$.75
--	----------------------	-----------------------	--------	------------	-----

Note. N/A means non-applicable. The variable of type of trait in the false recognition task, relation, and validity were contrast coded (false implied = -0.5, correct implied = 0.5; enemies = -0.5, friends = 0.5; match = -0.5, mismatch = 0.5) and the type of trait in the rating task was coded via the main contrast code of interest (quadratic contrast C1: implied = 2/3, evaluatively congruent = -1/3, evaluatively incongruent = -1/3).

Table S4

Correlation between the false recognition and rating tasks as a function of the type of effect (STT, STT by relation, and STT by validity) in the Preliminary Experiment and Experiments 1 and 3

Experiment	Effect	<i>r</i>	<i>df</i>	<i>t</i>-statistic	<i>p</i>
Preliminary Experiment	STT	0.24, 95% CI [0.14; 0.34]	334	4.52	< .001
Experiment 1	STT	0.03, 95% CI [-0.11; 0.17]	194	0.43	.67
	STT by relation	0.08, 95% CI [-0.06; 0.22]	194	1.17	.24
Experiment 3	STT	0.21, 95% CI [0.07; 0.34]	195	2.98	.003
	STT by validity	0.05, 95% CI [-0.09; 0.19]	195	0.73	.47

Note. The variable of type of trait in the false recognition task, relation, and validity were contrast coded (false implied = -0.5, correct implied = 0.5; enemies = -0.5, friends = 0.5; match = -0.5, mismatch = 0.5) and the type of trait in the rating task was coded via the main contrast code of interest (quadratic contrast C1: implied = 2/3, evaluatively congruent = -1/3, evaluatively incongruent = -1/3).

Table S5

The low diagnostic behavioral statements, trait implied by each statement, the alternative trait (and its average rating) used in Experiment 4, as established by Pilot Studies 1 and 2

Behavioral statement	Implied trait	Alternative trait	Mean value (and SD) for alternative trait
S/he always drove a little slower than the speed limit.	cautious	critical	6.56 (1.98)
S/he took the elevator up one flight.	lazy	judgmental	6.28 (2.88)
S/he was afraid the new employees wouldn't like her/him.	insecure	compassionate	6.18 (1.77)
S/he stories made people laugh so hard they held their sides.	funny	friendly	7.66 (1.88)
S/he took her/his first calculus course when s/he was 12 years old.	smart	proud	7.78 (1.95)
S/he picked out the best chocolates before the guests arrived.	selfish	outgoing	6.34 (1.96)
S/he told the cashier that s/he got too much change.	honest	friendly	6.42 (1.78)
S/he stepped on her/his boyfriend's/girlfriend's feet during the foxtrot.	clumsy	humorous	6.10 (2.48)
S/he lost track of the two year old.	irresponsible	attentive	5.7 (2.42)
S/he took 15 minutes to find her/his car in the parking lot.	forgetful	critical	6.92 (2.10)
S/he needed at least two packs of handkerchiefs when s/he watched a sad movie.	emotional	observant	6.68 (1.94)
Every day s/he invented a new game for her/his niece.	creative	proud	7.28 (1.86)

Table S6

Effect of the Pretest Frequency (values from Table S1) on STT and critical STT moderation effects when controlling for Pretest Frequency as a function of the task (false recognition vs. rating) in the Preliminary Experiment and Experiments 1-3

Experiment	Task	Effect	B (SE)	df	z- or t-statistic	p
Preliminary Experiment (+ reanalysis in Exp. 2)	False recognition	STT by Pretest Freq.	0.03 (0.82)	NA	0.03	.97
		STT by diagnosticity	0.03 (0.10)	NA	0.34	.73
	Rating	STT by Pretest Freq.	0.90 (0.54)	21.23	1.68	.11
		STT by diagnosticity	0.18 (0.06)	21.23	2.86	.009
Experiment 1 (+ reanalysis in Exp. 2)	False recognition	STT by Pretest Freq.	-0.18 (1.22)	NA	0.15	.88
		STT by relation by diagnosticity	0.21 (0.25)	NA	0.83	.41
	Rating	STT by Pretest Freq.	0.10 (0.76)	20.00	0.13	.90
		STT by relation by diagnosticity	0.25 (0.10)	65.41	2.65	.010
Experiment 3	False recognition	STT by Pretest Freq.	0.93 (1.26)	NA	0.74	.46
		STT by validity by diagnosticity	<0.01 (0.23)	NA	<0.01	>.99
	Rating	STT by Pretest Freq.	0.38 (0.96)	20.08	0.40	.69
		STT by validity by diagnosticity	0.24 (0.10)	48.30	2.35	.023
Experiment 4	Rating	STT by Pretest Freq.	-2.79 (1.91)	8.01	1.46	.18
		STT by diagnosticity	-0.87 (0.35)	8.01	2.51	.036
		STT for alternative traits	0.52 (0.09)	10.21	5.40	<.001

Note. The variable of type of trait in the false recognition task, relation, and validity were contrast coded (false implied = -0.5, correct implied = 0.5; enemies = -0.5, friends = 0.5; mismatch = -0.5, match = 0.5) and the type of trait in the rating task was coded via the main contrast code of interest (quadratic contrast C1: implied = 2/3, evaluatively congruent =

-1/3, evaluatively incongruent = -1/3). Pretest Frequency and statement diagnosticity were mean-centered.

Table S7

Moderation effect of the correct vs. false implied manipulation in the false recognition task on the STT effect in the Preliminary Experiment and Experiments 1 and 3

Experiment	<i>df</i>	<i>t</i>-statistic	<i>p</i>
Preliminary Experiment	22.81	0.01	.99
Experiment 1	23.14	0.03	.98
Experiment 3	23.26	0.31	.76

Note. The variable of type of trait in the false recognition task was contrast coded (false implied = -0.5, correct implied = 0.5) and the type of trait in the rating task was coded via the main contrast code of interest (quadratic contrast C1: implied = 2/3, evaluatively congruent = -1/3, evaluatively incongruent = -1/3).

Table S8

Residual STT effect in the rating task when using OLS regression as a function of the target condition (low diagnostic statements, enemies, mismatch) and the experiment

Experiment	Target condition	Df	t-statistic	p	Cohen's d
Preliminary Experiment	Low diagnostic statements	335	0.84	.40	$dz = 0.05$, 95% CI [-0.06; 0.15]
Experiment 1	Low diagnostic statements	195	2.29	.02	$dz = 0.16$, 95% CI [0.02; 0.30]
	Enemies	195	1.31	.19	$dz = 0.09$, 95% CI [-0.05; 0.23]
Experiment 3	Low diagnostic statements	196	3.07	.002	$dz = 0.22$, 95% CI [-0.08; 0.36]
	Mismatch	196	2.50	.013	$dz = 0.18$, 95% CI [0.04; 0.32]
Experiment 4	Low diagnostic statements	302	2.98	.003	$dz = 0.17$, 95% CI [0.06; 0.28]

Note. The variable of the type of trait in the rating task was coded via the main contrast code of interest (quadratic contrast C1: implied = 2/3, evaluatively congruent = -1/3, evaluatively incongruent = -1/3). Low diagnostic statements were defined as the statements situating one SD below mean value (4.87). Significant residual STT effects for low diagnostic statements emerged in the opposite direction (reversed STT effect) whereas residual STT effect for the mismatch condition emerged in the direction of a typical STT effect.