UNIVERSIDADE DE LISBOA

Faculdade de Psicologia

New Bottles for New and Old Wine: New Proposals for the Study of Spontaneous Trait Inferences

Diana Orghian

Orientadores: Prof. Doutor Leonel Garcia-Marques
Prof. Doutor Daniel Wigboldus

Tese especialmente elaborada para a obtenção do grau de Doutor em Psicologia, especialidade de Cognição Social

2017

UNIVERSIDADE DE LISBOA

Faculdade de Psicologia

New Bottles for New and Old Wine: New Proposals for the Study of Spontaneous Trait Inferences

Diana Orghian

Orientadores: Prof. Doutor Leonel Garcia-Marques
Prof. Doutor Daniel Wigboldus

Tese especialmente elaborada para a obtenção do grau de Doutor em Psicologia, especialidade de Cognição Social

Júri:
Presidente:
Prof. Doutora Isabel Maria de Santa Bárbara Teixeira Narciso Davide, Professora Associada e Vice-presidente do Conselho Científico da Faculdade de Psicologia da Universidade de Lisboa
Vogais:
- Doutora Rita Isabel Saraiva Jerónimo, Professora Auxiliar
Escola de Ciências Sociais e Humanas do ISCTE-Instituto Universitário de Lisboa;
- Doutora Teresa Maria Freitas Teixeira de Morais Garcia Marques, Professora Catedrática
Unidade de Investigação em Psicologia Cognitiva, do Desenvolvimento e da Educação do ISPA – Instituto Universitário de Ciências Psicológicas, Sociais e da Vida;
- Doutor Leonel Garcia Marques, Professor Catedrático
Faculdade de Psicologia da Universidade de Lisboa;
- Doutor Mário Augusto de Carvalho Boto Ferreira, Professor Associado
Faculdade de Psicologia da Universidade de Lisboa.

Projeto apoiado e financiado pela Fundação para a Ciência e Tecnologia

2017

American Psychological Association (PsycINFO Classification Categories and Codes):

3000 Social Psychology (3040 Social Perception and Cognition)

2340 Cognitive Processes (2343 Learning and Memory)

The present research was conducted with the financial support of Fundação para a Ciência e Tecnologia (FCT; the Portuguese National Foundation for Science and Technology) under the grant with the ID: SFRH/BD/84668/2012.

– That's why I like to listen to Schubert while I'm driving. As I said, it's because all the performances are imperfect. A dense, artistic imperfection stimulates your consciousness, keeps you alert. If I listen to some utterly perfect performance of an utterly perfect piece while I'm driving, I might want to close my eyes and die right then and there. But listening to the D major, I can feel the limits of what humans are capable of – that a certain type of perfection can only be realised through a limitless accumulation of the imperfect. And personally, I find that encouraging. Do you see what I'm getting at?

– Sort of . . .

Haruki Murakami, *Kafka on the Shore*

ACKNOWLEDGEMENTS

I would like to thank my collaborators Dietmar Heike, Jim Uleman and Tânia Ramos for making this work happen, for their wise comments, engaging questions, and rich discussions.

Thanks to Anna Smith, Filipa de Almeida, Tânia Ramos, Catarina Nunes, Melissa Lacro, Ana Lapa and Christine Looser for reviewing my dyslexic writing and thanks to Daniel Wigboldus for his useful feedback.

Would like to thank my dear Flávio for his patience and companionship, my family and friends for supporting me in this journey, my colleagues from A321 for all the great moments, and Cognition in Context Group for being my academic family during these years.

And, or course, a very, very special thank you to my deepest partner in this work, Professor Leonel Garcia-Marques, who is, by far the smartest person I know, and this is not only a spontaneous trait inference, this is a very rule based causal attribution! Thank you for these four years of dedication in teaching me how to think about Psychology while giving me freedom to explore it in my own way, in teaching me how to go after my ideas while helping me to refine them and, last but not least, thank you for teaching me how to enjoy a good Sovina while telling us your brilliant and hilarious stories!

Cambridge, September 2016

Diana Orghian

PREFACE

In this dissertation you will find some of the work I have been doing in the last four to five years. I say some of the work because, as any science worker knows, not everything we do has necessarily a fruitful outcome and not everything we do is "relevant" for a doctoral thesis. Some times what we explore happens to fail, in other words, we are wrong and, of course, we don't talk about our own wrong doings, we only talk about others' wrong doings (and I do some of that in this dissertation). One of the first lessons I was thought by my wise supervisor, right after the first experiment from my original project has failed, was that "Nature talks to us, we just have to learn how to better listen to it". That reminded me of my relationship with God when I was a child... eventually I gave up listening to God during my teen years. I didn't give up Nature, I still go to church, I still believe in the priest, and I have that almost blind faith that, one day, I will get a job ... at the church.

Therefore, this dissertation is about those unique moments when Nature mumbled something back. This dissertation has two very different contributions to the state of art in Spontaneous Trait Inference (STI). The first contribution regards a specific methodological aspect in the way the spontaneous trait inferences are measured. The second aspect regards a theoretical view about the processes underlying STIs.

In the first Chapter, I describe the nature of STI. Here, I try to convince you that STI is cool, that it matters and that it is worth being studied. Hopefully, this will give you a historical perspective about the discovery of this phenomenon, about the main paradigms that were developed in order to better capture its essence and the main debates that followed its discovery.

In Chapter 2 we introduce a debate about the processes underlying STI. Attribution and Association are the two processes that are usually suggested as possible candidates. Our work shows that Attribution is not supported by the empirical evidence and we also discuss how it brings a legacy that does not naturally apply to STI. We used connectionist models to show that STI and

its effects can be mimicked in a purely associative network, meaning that the findings we have up to date do not support the attributional perspective.

Chapter 3 follows from an assumption we make in Chapter 2, the assumption that we pay more attention to the actor that enacts the behavior as opposed to an irrelevant person that might be present at the moment of encoding.

In Chapter 4 we explore the way the information is processed in STI. In particular, we explore the way the behavioral information and the photo of the actor is processed when compared to a control situation where an irrelevant person is paired with the behavioral description.

In Chapter 5 and 6 we explore a methodological aspect of trait inference measurements. There are two ways of activating a trait, the first is via a real inference, in the sense that the trait is a result of reading and comprehending the behavioral description. And a second way, that leads to the activation of the trait due to intra-lexical associations that exist between the target trait and individual words in the behavioral description. We suggest a way of controlling for this confound and we also analyze the effect that word-based priming has on different STI paradigms and on the interpretation of previous findings in the literature.

In Chapter 7, after discussing the limitations of the main paradigms used to study STI, a new measure is presented (and tested) as an attempt to solve those limitations. The measure is a modified free association task that gives us access to the conceptual network of the inferred trait.

Finally, we close the dissertation with Chapter 8, where a short review of the previous chapters is made and a final conclusion is drawn.

RESUMO

Um dos temas que têm sido investigados pela psicologia social em geral e pela cognição social em particular é a forma como as pessoas percebem o comportamento manifesto dos outros e a forma como utilizam essa informação de maneira a categorizar os outros e de maneira a prever os seus comportamentos futuros. Mais especificamente, na presente dissertação, discute-se o tema das Inferências Espontâneas de Traço (IETs), um fenómeno que resulta na inferência ou na extração de traços de personalidade a partir do comportamento dos outros, sem intenção e sem consciência por parte do percipiente (Winter & Uleman, 1984).

Dois aspetos distintos das IETs são explorados na presente dissertação. Um diz respeito aos processos cognitivos subjacentes a este fenómeno e o outro diz respeito às metodologias utilizadas para medir e detetar IETs.

No que toca aos processos responsáveis pelas IETs, na literatura podemos encontrar duas perspetivas, uma que defende que as IETs são simples associações que se estabelecem entre o traço inferido a partir do comportamento e o ator desse mesmo comportamento (*e.g.*, Bassili & Smith, 1986) e uma outra que defende que são atribuições, isto é, que os traços passam a ser vistos como sendo atributos do ator (Carlston & Skowronski, 2005). A oposição entre estas duas perspetivas ganhou especial importância teórica quando os investigadores descobriram que o traço implicado pode ficar associado a alguém que esteja apenas a comunicar o comportamento, mas que não seja o ator do mesmo, um fenómeno chamado Transferências Espontâneas de Traço (TETs; Skowronski, Carlston, Mae, & Crawford, 1998; Brown & Bassili, 2002). Os investigadores parecem concordar quanto ao processo responsável pelas TETs, isto é, que um processo associativo está na sua origem. A discordância diz respeito às IETs. A perspetiva dualista e a sua explicação atribucional das IETs foi inspirada num conjunto de diferenças empíricas que foram encontradas entre as IETs e as TETs (para mais ver Orghian, Garcia-Marques, Uleman, & Heinke, 2015). Na presente dissertação, utilizou-se métodos computacionais de forma a demonstrar que estas mesmas diferenças podem ser replicadas num modelo puramente associativo. Assim, quatro

simulações foram desenhadas de forma a mimicar quatro resultados empíricos que distinguem as IETs das TETs. A primeira diz respeito a um maior efeito IETs comparativamente ao efeito TETs; a segunda pretende reproduzir um estudo em que o ator é apresentado simultaneamente com um comunicador, o que se mostrou interferir com as TETs; a terceira reproduz um efeito de *halo* maior em IETs do que as TETs; e por fim a quarta simulação pretende mimicar uma tarefa inferencial concorrente (a deteção de mentira) que interfere com as IETs mas não com as TETs. Este modelo serve para mostrar que as evidências empíricas que têm sido utilizadas a favor da perspetiva dualista não são suficientes para a sustentar porque podem ser reproduzidos num modelo que postula um único processo para ambos os fenómenos.

Este modelo baseia-se num pressuposto que está na origem da replicação bem-sucedida das quatro diferenças. Foi assumido que, quando a pessoa que acompanha o comportamento é o ator do comportamento, essa pessoa é relevante no contexto do seu próprio comportamento, e, por isso, receberá mais atenção e processamento por parte do percipiente do que no caso do comunicador. De seguida, tentou-se testar esse pressuposto experimentalmente.

Em três estudos foram exploradas as diferenças atencionais entre as IETs e as TETs. Assim, no primeiro estudo utilizou-se um paradigma que utiliza pistas espaciais (*spatial cueing paradigm*) e mostrou-se que os sujeitos precisam de mais tempo para reorientar a sua atenção da cara do ator do comportamento para uma nova localização onde se encontra um alvo (uma seta para a qual o sujeito tem que indicar a direção) do que no caso do comunicador. Nos estudos dois e três mediu-se os movimentos oculares dos participantes enquanto estes estavam a codificar a informação relativa a atores e comunicadores. Os resultados mostram que as pessoas prestam, de facto, mais atenção à cara do ator do que à pessoa irrelevante. Nestes dois estudos utilizou-se o paradigma do falso reconhecimento, e, ao contrário do que está relatado na literatura, não foram encontradas diferenças em termos da magnitude dos efeitos. Foi, no entanto, detetada uma correlação significativa entre o número de fixações na cara do ator e a proporção de falsos reconhecimentos (indicativos de inferências de traços acerca do actor). A mesma correlação não foi encontrada nos ensaios que dizem respeito às TETs. Esse resultado leva-nos a concluir que a atenção pode ter um efeito mediador nas IETs mas não nas TETs.

De seguida, demonstrou-se que as diferenças entre IETs e TETs não são ao nível da inferência do traço a partir da discrição comportamental. A memória das pessoas para as frases apresentadas

é semelhante nos dois casos. No que diz respeito à memória para as caras, mostrou-se que as caras dos atores foram melhor memorizadas do que as caras das pessoas que não eram relevantes para o comportamento descrito na frase. Este resultado está de acordo com os dados atencionais e o pressuposto em que se baseia o modelo acima descrito.

Na segunda parte da dissertação exploramos dois aspetos metodológicos das medidas utilizadas para detetar inferências. O primeiro aspeto está relacionado com o tipo de frases controlo utilizadas. Habitualmente, as frases controlo mais utilizadas são frases neutras (que não implicam nenhum traço em específico), sendo que as vezes não se utilizam frases controlo de todo. O tipo de controlo utilizado é, no entanto, importante pela razão a seguir exposta. Há duas formas de ativar um traço ao codificar uma frase descritiva de um comportamento, uma que resulta do processamento da frase e do comportamento como um todo (primação com base no texto), e uma segunda forma pela qual o traço é ativado como consequência direta da leitura de determinadas palavras da frase (primação com base na palavra; como é o caso da palavra "xadrez" que pode primar o traço "inteligente"). Apenas a primeira forma de ativar traços pode ser considerada uma inferência. Uma forma de saber se a ativação deve-se à primação com base na palavra ou no texto é ter frases controlo em que as mesmas palavras que fazem parte das frases implicativas são rearranjadas em frase novas de forma a não implicar o traço (Keenan, Potts, Golding, & Jennings, 1990). Se se encontrar diferenças em termos da ativação do traço entre as frases implicativas e as frases rearranjadas, estas diferenças só se podem dever à inferência com base no texto porque as duas frases têm aproximadamente as mesmas palavras. Assim, primeiro criamos um conjunto de pares de frases implicativas e rearranjadas e depois testamos parte delas com diferentes paradigmas. Mais especificamente, o que se hipotetizou foi uma maior influência da ativação com base na palavra em paradigmas imediatos do que em paradigmas tardios. Em paradigmas imediatos (como são a decisão lexical e o paradigma do reconhecimento do alvo) a inferência mede-se imediatamente a seguir à leitura da frase implicativa, enquanto que num paradigma tardio (como o paradigma do falso reconhecimento) a inferência é medida depois do sujeito ter lido todas as frases implicativas. Tal como previsto, observou-se uma influência significativa da ativação com base na palavra em medidas imediatas. E como tal, recomendamos que, em estudos futuros, se utilizem frases controlo rearranjadas de forma a certificar que se está a lidar com inferências verdadeiras.

Por fim, depois de se realizar uma análise das várias medidas que se utilizam na investigação das IETs, concluímos que há dois aspetos cruciais a ter em conta. Um diz respeito à contaminação das medidas com recordação explícita, um problema reconhecido já há algum tempo na literatura. E um segundo diz respeito ao tipo de processo que é requerido pela tarefa. O tipo de processo pode ser mais conceptual ou mais percetivo. Tendo em conta que as inferências são ativações de conceitos que se baseiam maioritariamente no processamento do significado dos comportamentos, as medidas conceptuais são mais apropriadas em oposição às percetivas que são melhores em casos em que a ativação do conceito se baseia em processamento percetivo. De forma a lidar com o problema da contaminação e com esta característica conceptual das IETs, apresentamos uma medida nova inspirada na tarefa modificada de associação livre apresentada por Hourihan e MacLeod (2007). Nesta tarefa, os sujeitos começam por ler frases implicativas ou rearranjadas. De seguida são lhes apresentados traços um a um. A tarefa consiste em dizerem a primeira palavra que lhes vem à cabeça ao ler o traço. Assume-se que a inferência levará à ativação não apenas do traço, mas também da rede semântica desse traço e por isso no momento da associação livre será mais fácil para o sujeito gerar um associado do traço. A variável dependente é o tempo que o sujeito demora a gerar uma resposta. Este paradigma foi testado no seu formato imediato e tardio e ainda foi testada a sua sensibilidade quanto à diferença entre IETs e TETs. Além de parecer um bom paradigma para detetar inferências, pode ser informativo quanto às diferenças entre IETs e TETs no que diz respeito aos seus processos subjacentes.

Palavras-chave: inferências espontâneas de traço; modelo conexionista; atenção; primação com base em palavras; tarefa de associação livre.

ABSTRACT

An important research topic in social cognition concerns the way people understand others' behaviors and the way they use this information to categorize others and infer causes for their actions. More specifically, in this dissertation, we investigated the Spontaneous Trait Inference (STI), a phenomenon that allow people to infer or extract personality traits from others' overt behaviors and to use those traits to make further judgments. It is a spontaneous mechanism because it occurs without intention or awareness.

The dissertation is organized in two parts that deal will two distinct aspects of STI. The first aspect regards the processes responsible for the occurrence of STI. The second is about the paradigms used to detect STI and their limitations.

In the first part, we discuss the two perspectives that exist in the literature regarding the processes underlying STI. These two perspectives emerged as a reaction to the discovery of a surprising phenomenon, the Spontaneous Trait Transferences (STT). STTs occur when a trait is inferred from a behavior and associated with someone else that not the actor: a communicator, a bystander or any other irrelevant stimulus present in the context at the encoding moment. Based on empirical differences between STI and STT, a dualistic perspective was proposed in which STI are said to result from attributional thinking and STT from simple associations. A different perspective suggests that the same process, an associative one, can be responsible for both.

Our contribution to this debate was to develop a computational model in order to demonstrate that the evidence supporting a dualist view are weak, because a simple associative model can reproduce, not only STI and STT, but also the empirical differences between them. Moreover, as an assumption of the model, we argued that there might be an attentional difference between STI and STT. Thus, next we tested this assumption by using the spatial cueing paradigm and eye-tracking devices, which allowed us to conclude that people pay more attention to the actor of a behavior than to an irrelevant person presented with it. Also in agreement with the attentional difference and with the model, we showed, by using forced recognition paradigm, that in both

STI and STT the trait is inferred in a similar way from the behavior, whereas the memory for the photo is better in STI than in STT.

In the second part of the dissertation, we discuss the main methodologies used to measure STI. We start by examining a confound present in many studies investigating STI, the word-based priming. This confound consists in the activation of the trait based, not on the interpretation of the whole sentence and the behavior in it described, but on the presence of specific words that alone lead to the priming of the trait. Moreover, we showed that this is only a problem for the immediate measures of STI such probe recognition paradigm, but not for delayed, such the false recognition.

A different limitation that affects all the measures based on memory is the contamination with explicit recall of the sentence. The use of online measures can solve, in part, that problem. However, online measures are data-driven, or, in other words, are measures that rely on feature and perceptive processing. This characteristic makes them unsuitable for STI, that is a conceptually-driven mechanism. Thus, we introduce a new implicit conceptual measure, the modified free association task. In this task people first read trait-implying or control material. Afterwards, they perform a free association task were a word (the inferred trait) is presented and the subject is instructed to say the first word that comes to their mind when reading the presented target. We tested this new paradigm in delayed and immediate modes and we also tested its sensitivity do STI and STT difference.

Key-words: spontaneous trait inferences; connectionist model; attention; word-based priming; free association task.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF PUBLICATIONS INCLUDED

This dissertation is written as a compilation of articles published, submitted or ready for submission during the period of the doctoral program. Below is a list of the articles included in this dissertation and the respective chapter.

Note that the articles were slightly modified in terms of format and content in order to better fit to the structure of the dissertation.

CHAPTER 2 *A Connectionist Model of Spontaneous Trait Inference and Spontaneous Trait Transference: Do they have the same underlying processes?*

Social Cognition, 33(1), (2015),

Diana Orghian, Leonel Garcia-Marques, Jim Uleman, and Dietmar Heinke.

CHAPTER 3 *Why "well done is better than well said": The involvement of attention in Spontaneous Trait Inferences and Spontaneous Trait Transferences*

In preparation,

Diana Orghian, Leonel Garcia-Marques, and Dietmar Heinke.

CHAPTER 4 *Information Processing in Spontaneous Trait Inference and Transference*

In preparation,

Diana Orghian, Tânia Ramos, and Leonel Garcia-Marques.

CHAPTER 5 *Acknowledging the Role of Word-Based Activation in Spontaneous Trait Inferences*

Submitted,

Diana Orghian, Tânia Ramos, Joana Reis, and Leonel Garcia-Marques.

CHAPTER 6 *Activation is not always inference: Testing the influence of word-based priming on immediate and delayed tests of spontaneous trait inferences*

Submitted,

Diana Orghian, Tânia Ramos, and Leonel Garcia-Marques.

CHAPTER 7 *Capturing spontaneous trait inferences with the modified word association test*

Submitted,

Diana Orghian, Anna Smith, Leonel Garcia-Marques, and Dietmar Heinke.

1

INTRODUCTION

> *If perceptual experience is ever had raw, i.e., free of categorical identity, it is doomed to be a gem serene, locked in the silence of private experience.*
>
> Bruner, 1957, p. 125

Spontaneous Trait Inference (STI) is the tendency to infer personality traits when observing behaviors other people perform. It is called spontaneous because the trait inference occurs without intention or awareness (*e.g.*, Winter & Uleman, 1984).

In the first Chapter of this dissertation, we review how the topic of STI evolved over decades of research. We don't present an exhaustive literature review; instead, we focus on aspects of STI that are yet to be solved or are more relevant in methodological or theoretical terms.

The Chapter starts with a short description of the research and the ideological context where STI emerged. After describing how the STI was discovered, we review the suggestion regarding STI's automaticity. We conclude that a continuum view of automaticity applies better to inferences than does a traditional dichotomous view.

Next, we examine the empirical evidence supporting the idea that the trait inference is a characterization of the actor performing the behavior, and not a mere categorization of the action. This aspect of STI is crucial because the researchers are mainly concerned in studying inferences people make about others, because those are the ones that will regulate our social interactions and allow us to make predictions about future behaviors of specific people. There are at least two paradigms that show that a trait is specifically linked to the actor, the saving in relearning and the false recognition paradigms.

The considerations that follow refer to the concept of *inference* itself. Because there is almost no systematic dialogue in the STI field regarding what an inference is, and specially what it is not, we turned to text comprehension literature in order to better picture the concept. Crucially, we call attention to the distinction between word-level activation and text-level activation and to the fact that in STI research, these two are usually confounded. A solution to bypass this confound is also presented.

Next, we review the main paradigms used to study STI and their limitations. There are two very different kinds of paradigms, memory-based paradigms that rely on participant's memory in order to measure the inference, and on-line measures where no reference to past events is required. A different way to distinguish measures is by the type of processes required to perform the task. Data-driven tasks rely more on superficial features of the stimuli and conceptually-driven tasks rely more on top-down processing. We suggest that conceptually-driven tasks are more appropriate to investigate STIs. This debate culminates in the proposal of a new conceptually-driven paradigm that is presented as a possible solution for some of the limitations of more traditional paradigms.

Finally, the Chapter ends with a important theoretical debate regarding the processes underlying STIs. In the literature, it is still unclear what is the hallmark process of STI and what makes this phenomenon unique and different from other simpler phenomena. The view that gained more popularity in the field claims that an attributional process underlies STI. Our contribution to this debate is the development of a computational model that demonstrates how the evidence supporting the attributional explanation is scarce. Part of our demonstration is the assumption that attention plays an important role in the STI phenomenon, an assumption that is further tested in Chapter 3.

1.1 A LONG TIME AGO...

From our ancestral times, human survival has depended on the existence of communities. This means our survival has been relying on our ability to understand others, whether these others were friends or foes. There are several reasons why inferring personality traits is an efficient way of understanding others. A very intuitive one is for cognitive economy reasons (*e.g.*, Heider,

1958); we cannot memorize everything about everybody and because of that we need to simplify the information. Inferring traits can be seen as a way of organizing/categorizing behavior (*e.g.*, *cooperative* behavior) and also people (*e.g.*, *trustworthy* person). Second, we not only learn through our own experience, we also learn from other people's experience. Thus, if I had a bad experience with someone, you bet I will share that with my people, so that they can defend themselves against this person. Thus, traits can be used as descriptive of someone's personality (*e.g.*, Hamilton, 1988). Third, by inferring traits, and categorizing behaviors, we can predict future behaviors (McCarthy & Skowronski, 2011b). Knowing what to expect from others gives us some perception of control over our complex social environment. Hence, trait inference is an urgent and important topic in Psychology.

We are innate interpreters, that are programmed to organize and classify information in order to extract meaning. As Bruner (1957) refers to the difference between the visual field (reflected light, shadows, brightness) and the visual world (recognizable objects), we can refer to the "social field" (people, actions, situations) and the "social world" (representations of people with certain personalities, that are part of certain social groups). There is a social field out there that has people in it, and those people perform actions, react to situations and to other people. This field exists independently of our perception. Whereas our social world is something we create, it is our interpretation of the social field we are exposed to. In this world we create, there are people with emotions, goals and desires; there are people we like and people we don't.

Furthermore, from the analogy regarding the visual experience, it immediately follows that by comparing the real thing out there with the representation we create in our mind of that thing, we can measure the accuracy of our perception. Of course, when it comes to inferring others' personalities, this approach might be troublesome, and, curiously, not because of our subjective interpretation, but mainly because measuring the objective social reality is more complicated than measuring the visual reality. We can easily and objectively quantify the length or the height of an object, but we cannot observe and measure the personality in the same way. The same can be said about the behaviors people perform (the actions themselves). An action can be seen and described in objective terms (*e.g.*, "she is vocalizing a melody"), but can a personality trait (*e.g.*, "she is a happy person") be perceived and described in the same way or a trait is already part of the realm of subjectivity? In other words, does the trait exist outside the perceiver's mind?

Until the 50s, social psychologists were concerned with the accuracy of our judgments about others, in particular, they were searching for the perfect judge of personality (Allport, 1937). It was with the critical work done by Cronbach (1955, 1958) that this approach to personality was almost eradicated. The research moved toward the study of our subjective perception of personality, without comparing it to some kind of "truth".

The coming trend was two folded. On one hand, a genius mind named Solomon Asch, inspired by Gestalt Psychology, was interested in the "basic features ... present in the judgment of actual people" (Asch, 1946, p. 283) and in the way people combine information in order to create unique impressions. On the other hand, Fritz Heider was trying to link person perception to search for causality, *i.e.*, he tried to understand what makes people do things and what causes actions (Heider, 1944). Unlike Asch, Heider did not use experimentation to test his ideas, relying mainly on his personal experience and observations. His main goal was the theoretical analysis and the conceptual clarifications. During the time Asch was bringing experimentation to the field, Heider was bringing theory and rules, both scarce in Social Psychology at the time.

Heider's theorizing gave birth to what we call today the classical attribution theory that was further developed by Jones and Davis's more concrete conceptualizations and experiments (Jones & Davis, 1965) and culminated in Kelley's ideas (1967).

Jones and Davis (1965), proposed the first systematic model of dispositional inference - the theory of *Correspondent Inference*. In their theory, the authors tried to account for trait inference based on what the perceiver understands about an actor's action. *Correspondence* refers to the "extent that the act and the underlying characteristic or attribute are similarly described by the inference." (1965, p. 223). In other words, a correspondence inference occurs when it is assumed with high confidence that the behavior is a direct reflection of actor's intention, that is, in turn, a reflection of his disposition. What leads to the perception of intentionality is 1) whether there are many uncommon effects following the action; more uncommon effects resulting from a behavior will lead to less intentional inference, because there can be multiple intentions responsible for the action, and 2) whether the effect of the action is seen as desirable by the actor, *i.e.*, the perceiver will find the behavior as more diagnostic of actor's intention more desirable the consequence seems to be to the actor. Thus, if someone's action has a number of uncommon effects, it is more difficult to infer which of the many possible intentions behind these effects guided the behavior.

One way of solving this ambiguity is to analyze the desirability of the effects, such that if there is one that is more desirable, the perceiver will assume that that was the effect intended by the actor.

Moreover, Jones and Davis (1965) suggested that the perceiver is interested in isolating the invariant properties that distinguish one person from another. The final result of such operation is a correspondence inference: it is a dispositional attribution of a trait to someone that, in the perceiver's perspective, detains the trait more than an average person. Thus, the perceiver is looking for extraordinary dispositions and not only for common dispositions presented by the majority and that is why the uncommon effects have such a crucial role in this theory. The authors also suggested that if an actor is not constrained by social situations, the perceiver will attribute the action to his intention and consequently to his dispositions. If the actor's actions are constrained by the context, then no such dispositional inference will occur.

Jones and Davis's ideas were disconfirmed, especially by studies showing dispositional inference in situations where the actor had no free choice in deciding his actions and thus, no intention could be inferred (Jones & Harris, 1967; Jones, Worchel, Goethals, & Grumet, 1971; Snyder & Jones, 1974). It was shown, in a now-classic study that when people were randomly assigned, versus given the option to chose to defend a pro-Fidel Castro (or anti-Castro) position in a debate, subjects judged the targets that were randomly assigned to the position as having a pro-Castro attitude even knowing that this position had nothing to do with the actual position of the judged target (Jones & Harris, 1967). We also know, for example, that people randomly assigned to receive bad news are also judged as more chronically depressed than those assigned to receive good news (Gilbert, Pelham, & Krull, 1988).

After Jones and Davis, Harold Kelley suggested that people perform a set of inferential tests that are informative about the covariance of one's perception and the thing perceived. He goes as far as asserting that ordinary people use the same tools as researchers use when doing science (*e.g.*, Kelley, 1967). As the attributional system is envisioned to extract invariance, Kelley suggested that if I am acting nicely toward you, you will ask three critical questions: 1) am I routinely nice to you (consistency)? 2) am I nice to other people as well or only to you (distinctiveness)? 3) are people usually nice to you or it is only me (consensus)? Answering these questions will inform you about the attributional locus of my behavior. Was it due to something about you, something

about me or something about the context? In other words, you are testing for three kinds of covariation (covariation principle): covariation of the behavior with the actor, with the stimuli, and with time. He introduced another two famous principles: the "discounting principle" (Kelley, 1971) that says that if the actor's behavior covaries with more than one effect, the confidence in the inference made decreases, and the "augmentation principle" that states that a cause is perceived as stronger if its effect occurs in the presence of inhibitory causes.

In the following decades the research on person perception proceeded with a whole new generation of thinkers, strongly influenced by the cognitive revolution spreading through the labs world-wide – the social cognitivists (Gilbert, 1998). They were no longer concerned with the accuracy of our perceptions, nor with the logical rules that attributions are made of. They wanted to know what happens in our mind when we make a dispositional inference, logical or illogical, accurate or inaccurate. The implementation of the rules was the main interest.

Another very important shift was with regard to the awareness and the spontaneity of dispositional inferences. Whereas the attributional researchers were working with conscious and effortful operations, the social cognitivists focused on unconscious, effortless processes (Gilbert, 1998). Although, in the early days of person perception, there were already some hints about the inescapability of the impression formation (Asch, 1946; Heider, 1944; Tagiuri, 1958); during the heyday of attribution research, dispositional inference undoubtedly became synonymous with causal thinking. For instance, the most popular perspective those days was that people needed motivation to make attributions. Kelley and Michela for instance, asserted that "a person's interests (. . .) determine when he will become motivated to make attribution at all" (1980, p. 473).

In the eighties researchers like Smith and Miller (1983) started to suggest, that the trait inference could be seen as an inherent part of encoding and comprehension of information. These authors used processing time to study the kind of judgements occurring during the process of comprehension. The trait-implying material was presented before participants knew the type of judgement they would have to make. Smith and Miller found that judgements about intention and trait inferences were not significantly slower than control judgements (*i.e.*, judgements about gender) that were not relying on inference making. The authors interpreted the results as evidence that judgements about intentions and dispositions don't need extra time to be inferred at the moment the judgement is made and suggested that that happens because dispositions were

previously inferred during the comprehension of the event. Moreover, Smith and Miller found that judgements about liking and repetitions of the same actions were slower than control, and questions regarding causal attribution about the person or the situation were the slowest of all. The author asserted that causation judgements happen after trait assignments are made.

However, only with Winter and Uleman (1984) and the implicit measure by them introduced did the field begin to have clear evidence about the spontaneity of trait inferences. These authors adopted the logic of the encoding specificity principle (*e.g.*, Thomson & Tulving, 1970; Tulving, 1972) that says that the encoding event has an important role at retrieval, and as such, a cue from the encoding moment given at retrieval will lead to an effective retrieval. The logic applies to trait inference in the following way. If a trait is inferred when the behavioral sentence is being read under memory instruction, then it is because the inference making is a natural result of comprehending the actor's behavior and not a result of explicit goals. If that is true, then the trait will be encoded in memory alongside the sentence being memorized and, later on, the trait will serve as an effective retrieval cue for that behavioral information. If a participant reads the sentence "The librarian carries the old woman's groceries across the street", he infers that the librarian is helpful, and the word "helpful" will be a good retrieval cue for the sentence. In comparison with no cue condition, the dispositional cue leads to a better recall of the sentence, and in comparison to a strong associate of a word from the sentence it performed equally well or better. The participants in these studies were not instructed to form impressions but they still spontaneously inferred traits. Moreover, participants seemed to be unaware that they were making such inferences. This effect was named Spontaneous Trait Inference (STI).

The new idea that trait inferences are spontaneous inspired Gilbert and colleagues (1988) to create a model claiming that any attempt to understand others started with a corresponding disposition. Only after inferring the implied trait people would check whether other, situational, factors affected the behavior. The authors assert that there is an initial dispositional inference that does not require considerable effort or conscious attention. This initial stage is then followed by a correction stage that is more deliberative and more cognitively demanding. We come back to this model later on in section 1.8.

Regardless of STI's current popularity, it was not accepted straight away as a real psychological phenomenon. A series of papers were published in the following years trying to ameliorate the

methods in order to answer the different concerns raised by the research community. As discussed below, some of the concerns were successfully addressed, while others remain unsolved. In the rest of the introduction, we discuss some of these concerns and the following Chapters in the dissertation are an attempt to address these concerns, especially those that remain unsolved.

1.2 AUTOMATIC VERSUS CONTROLLED STI

One of the first concerns the scientific community addressed was regarding the apparent automaticity of STI. Winter and collaborators suggested that trait inferences might be initiated in an automatic way (Winter, Uleman, & Cunniff, 1985). However, this suggestion was mainly based on two characteristics of STI: lack of intention and lack of awareness. According to the most popular view of automaticity, for a process to be considered automatic, it needs to satisfy four criteria: lack of intentionality, lack of awareness, efficiency, and uncontrollability (*e.g.*, Bargh, 1994).

A trait inference is considered intentional when it is triggered by an explicit impression formation goal, whereas it is unintentional or spontaneous when the inference is a result of processing the behavior with other goals that are not relevant to impression formation (Uleman, 1999). Winter and Uleman (1984) instructed their subjects to memorize the information, and the detection of STI under these instructions was taken as evidence that STI was unintentional. Winter and colleagues (Winter et al., 1985) presented further evidence regarding the lack of intentionality. They presented trait implying sentences as distractors while subjects performed a digit memory task. Even though participants believed that the sentences were completely irrelevant, they inferred traits with very little awareness of doing so. Later on, the traits proved to be good retrieval cues for the sentences and as good as semantic associates. Other researchers applied awareness questionnaires after the behavioral material was presented and they found no awareness and no correlation between awareness and trait-cued recall (Lupfer, Clark, & Hutcherson, 1990; Moskowitz, 1993). Uleman and Moskowitz (1994), found however, that 14 % of the subjects reported trait thoughts and these thoughts correlated with trait-cued recall, suggesting awareness of the inference among some subjects.

Moreover, subsequent studies showed that STIs occur even when people are processing trait-implying sentences under cognitive load (processing difficult digit lists; Lupfer et al., 1990; Winter et al., 1985). However, Wigboldus and collaborators (Wigboldus, Sherman, Franzese, & van Knippenberg, 2004) found differences in the way stereotypical information interacted with trait inference under cognitive load. They found that under high cognitive load, and when stereotypical information incongruent with the to-be-inferred trait was provided about the actor, the trait was inferred less than in the cases where the stereotypical information was consistent with the inference. This result not only shows sensitivity to cognitive load, but also an interaction of trait inference with other relevant information about the actor. These results question the uncontrollability and the efficiency criteria that automatic processes are supposed to present (Bargh, 1994). Other studies have been showing that STIs are sensitive to cognitive load (Uleman, Newman, & Winter, 1992) and furthermore, a correlation was found between STI and individual differences in working memory (Wells, Skowronski, Crawford, Scherer, & Carlston, 2011).

We also know that the STI are affected by people's processing goals, with more STIs being made when the sentences are read under meaning searching goals and with fewer STIs being made under more superficial analysis goals (sentence features; Uleman & Moskowitz, 1994). This suggests that STIs are not completely uncontrollable. Additionally, we know that traits are more likely to be inferred if participants are instructed to infer traits than when they are not (*e.g.*, Bassili & Smith, 1986) and also when subject are primed with affiliation goals (Rim, Min, Uleman, Chartrand, & Carlston, 2013). However, being affected by goals does not mean that the process cannot be automatic. There are studies showing that even goal-pursuit inferences can occur in an automatic way (Lieberman, Jarcho, & Obayashi, 2005; McCulloch, Ferguson, Kawada, & Bargh, 2008). Participants can be instructed to perform dispositional inferences and still perform them in an automatic way. Automatic in the sense that other relevant information such situational information won't be used for discounting or augmentation of the dispositional inference, specially under cognitive load. The opposite can also be true, participant instructed to perform situational inferences can perform them in an automatic way, in the sense that no dispositional explanation will be taken into account (Lieberman et al., 2005). This is consistent with another important result showing that spontaneous situational inferences can be obtained at the same time the trait inferences are obtained and in the same participant (Ham & Vonk, 2003).

There is also evidence that STIs are affected by practice. For example, Bassili (1993) demonstrated that participants who practiced inference making before, are more likely to make STI than are control participants. Furthermore, Bargh and Thein (1985) showed that participants that chronically make more dispositional inferences show more STI.

Thus, if we look at the automaticity as a continuum and not a dichotomy, STI would be, in some aspects, closer to the more automatic pole (in terms of intentionality and awareness) and closer to the more controlled pole in others (in terms of control and efficacy). Indeed, by employing the Process Dissociation Procedure (PDP; Jacoby, 1991; Hassin, Uleman, & Bargh, 2004; McCarthy & Skowronski, 2011a), it was demonstrated that there are both automatic and controlled processes taking place in STI. In the PDP, it is assumed that there are no process-pure phenomena, and given the right experimental design it is possible to quantify the contribution of controlled and automatic processes in a certain task. A typical experiment needs to have two conditions: an inclusion condition where both automatic and controlled processes work together and an exclusion condition where the two types of processes work in opposition. In the inclusion condition, it is usually assumed that the performance is due to controlled processes (C) plus the automatic processes when the controlled fail ($A(1-C)$; $inclusion\ condition = C + A(1-C)$). The performance in the exclusion condition is due to automatic processes and the failure of controlled ($exclusion\ condition = A \times (1-C)$). Note that for this to be true controlled and automatic processes are assumed to be independent. This procedure allows us to calculate the contribution of each type of process, the controlled by subtracting the exclusion condition to the inclusion condition ($C = inclusion\ condition - exclusion\ condition$) and the automatic by dividing the performance in the exclusion condition by the term representing failure of controlled processes ($A = exclusion\ condition \div (1-C)$). This procedure was used in combination with the false recognition paradigm (McCarthy & Skowronski, 2011a). The inclusion condition consisted of trials where the sentence included the trait ("I am *grumpy* because people just annoy me. It seems like wherever I go that I am surrounded by idiots.") and the exclusion condition consisted of trials where the paragraph only implies the trait ("People just annoy me. It seems like wherever I go that I am surrounded by idiots."). After a phase where subjects memorized trait-implying paragraphs, they were asked to indicate whether a word (the trait) was part of the paragraph seen before. In the inclusion, both the explicit remembering of the sentence and the inference led to

the correct "yes" response, whereas in the exclusion only the inference (in the absence of remembering) was leading to "yes" responses (those are called false recognitions). The authors found that both processes are taking part in STI. Moreover, these two types of processes are affected by different variables (McCarthy & Skowronski, 2011a).

These findings suggest that categorizing STIs as automatic versus controlled (following the requirements of awareness, intention, efficiency and control; Bargh, 1994) is not an adequate approach, whereas positioning it onto a continuum might be a better approach that would allow us to observe how the dynamics of the controlled and automatic processes evolve when exposed to different and relevant manipulations.

1.3 INFERENCE ABOUT THE ACTOR VERSUS THE BEHAVIOR

The second concern raised by the researchers in the field was related to the link between the trait and the actor. Was the trait an inference made about the behavior or was it about the actor performing the behavior?

In the studies using cued-recall (Winter & Uleman, 1984; Winter et al., 1985), the traits were very effective cues to recall the whole sentence, but they were more effective in retrieving the predicate of the sentence than the actor. Leading one to wonder whether the trait is actually about the actor (Bassili, 1989; Higgins & Bargh, 1987). Traits can be used to categorize people or their behaviors, and while categorizing a person in terms of a trait implies that their behavior was also categorized, categorizing a behavior does not mean the person is also categorized in terms of the same trait. Indeed, Whitney and collaborators (Whitney, Davis, & Waring, 1994) showed, by using sentence reading time, that spontaneously inferred traits might very well be categorizations of behaviors. They had people read four passages in which the last sentence implied a specific trait. The first sentence was a behavioral description of a neutral, a congruent or an incongruent behavior in relation to the trait implied in the last description. The authors found that the reading time of the fourth sentence was faster when it was congruent with the first sentence than when it was neutral, and, crucially, this was true even when the actor of the fourth sentence was different from the one in the first sentence.

Two subsequent paradigms were developed to disentangle this matter:

The first was the saving in relearning paradigm (Carlston & Skowronski, 1994), which was an adaptation of Ebbinghaus's relearning paradigm (Ebbinghaus, 1964). The paradigm consists of a sequence of phases. In the first phase, participants were presented with pairs of paragraphs describing behaviors and photos while instructed to get familiarized with or memorize the material. After a confusion task, that had the purpose of making the recall of the behaviors less likely, subjects had to memorize pairs of photos and traits (phase 2). There were three types of trials in this phase: 1) trials where the photo was paired with a trait that was implied in the paragraph presented with that photo in phase 1, meaning that if the trait was inferred in phase 1 this was a relearning opportunity of the same pairing – relearning pairs; 2) trials where both the photo and the trait were "old", in the sense that they were seen or inferred in phase 1, but were not paired together - priming pairs and 3) trials where the photo and the trait were "new", so no possible relearning was taking place here - control pairs. In a third phase participants had to perform a cued-recall task where a photo was presented as a cue and the trait presented with it in phase 2 had to be recalled. Surprisingly, the recall was significantly better for the relearning pairs than for the priming and control pairs, meaning that there was an association being relearned between that actor and that trait.

A different way of studying this same question was developed by Todorov and Uleman via the false recognition paradigm (Todorov & Uleman, 2002). A typical experiment using this paradigm starts with the presentation of behavioral descriptions paired with the actors performing those behaviors – the learning phase. After participants are exposed to these pairs, the same actors are presented with traits in a test phase. Sometimes the actor is presented with a trait that was implied in the sentence paired with that actor in the learning phase (match trials). Other times the actor is paired with a trait that was implied before with a different person (mismatch trials). Participants are instructed to indicate if the trait was part of the sentence presented alongside that actor in the learning phase. The dependent variable is the rate of false recognitions, *i.e.*, wrong "yes" responses. In this test phase, both the photos and the traits are familiar to participants, whereas the pairing between the photo and the trait can be new or old. If the trait was linked to the actor during the learning phase, more false recognitions are expected in the match condition than in the mismatch.

Note, however, that the link between the person and the trait isn't very specific because almost any kind of stimulus presented with the behavior (Carlston & Skowronski, 2005) can get attached to the trait, a phenomenon called transference (this phenomenon will be discussed in section 1.6).

1.4 WHAT IS AN INFERENCE AND HOW IT OCCURS?

In the realm of trait inference there are important questions still unanswered, and some of them are surprisingly fundamental in their nature. One of those questions regards the very nature of the inference concept.

Research done in text and discourse comprehension literature can provide us with useful insights regarding this question. An assumption underlying text comprehension regards the involvement of inferential thinking (*e.g.*, M. K. Johnson, Bransford, & Solomon, 1973). Because it is such a vast field, most of the problems found in the STI literature have a close reflection in this literature and, consequently, a possible solution.

To comprehend a sentence or a text means to be able to combine information from two sources, explicit information given by the text and knowledge the perceiver has stored in their memory. This interaction gives birth to new information, information that is not part of the text but that is implied in the text. As straightforward and intuitive as this general definition might sound, when it comes to details, there is not much consensus about what an inference is.

McKoon and Ratcliff define inference as "any piece of information that is not explicitly stated in the text" (McKoon & Ratcliff, 1992, p. 440). This broad definition includes very simple inferences like the relation between a pronoun and its referent or more elaborate inferences (*e.g.*, inferences regarding consequences or instrumental inferences). Moreover, McKoon and Ratcliff suggest that there are only two types of inference that are drawn automatically, those that are necessary for local coherence, and those that are highly accessible. To further clarify the idea of *highly accessible* concepts, McKoon and Ratcliff propose that each element in the text (from basic elements like words to more complex structures as propositions) cue additional information from memory in a passive and efficient way (McKoon, Gerrig, & Greene, 1996; McKoon & Ratcliff, 1995; Ratcliff & McKoon, 2008).

These same authors introduced a gradual view of inference, that is, they suggest that an inference can be only partially encoded. In one of the studies, the authors found that if a relevant prime (a word that was part of the implying sentence) was presented before the target word at test, the participants presented more errors in indicating whether the target was part of the sentence in the implying condition than in the control condition (McKoon & Ratcliff, 1986). This difference did not reach significance when a neutral prime was used. McKoon and Ratcliff interpreted this result as evidence that the inference was only minimally encoded when the preceding prime was neutral, whereas with the relevant prime some more strength was added to the inference, making it detectable (McKoon et al., 1996; McKoon & Ratcliff, 1989c, 1989a, 1990a; Potts, Keenan, & Golding, 1988). They also claimed that an inference is more encoded when the information the inference is based on is well-known.

Kintsch's model (1988, 1998) also fits with the gradual idea of inference making. In the construction-integration model, it is suggested that the inferential process starts with a construction of a network of concepts that gets activated without any restrictions. This means that even inconsistent concepts can become activated at this stage, as long as there is some association going from the text to knowledge stored in memory. In the second phase, the integration phase, the representational system will apply cycles of filtering over the activated network in such a way that at the end only contextually relevant concepts are maintained activated. Thus, an inference weakly activated at the beginning of this process can become highly activated over the time if it is contextually relevant.

This view of inference is a more dynamic one, where context, search for meaning, and coherence interact. Suggestions have been made about how this gradual and flexible view of inferences can be applied to the STI research (Ramos, 2009). The STI has been described by default as an all-or-none process, and approaching it from a continuum perspective might be more appropriate and more fruitful. Specifically, because there is evidence that the magnitude of the effect vary depending on the processing goal (Uleman & Moskowitz, 1994) or on the consistency with the remaining contextual information (Wigboldus, Dijksterhuis, & Van Knippenberg, 2003).

However, while the process of how inference making occurs is relatively examined, the question of what an inference *is*, is sometimes left unexplored. Perilously, can a simple association

like activating the word "smart" due to the processing of the word "library" be considered an inference?

McKoon and Ratcliff were more concerned with the automaticity of inference making and did not address this question in a systematic way. Their very broad definition of inference includes any information activated based on the text and that is not part of it. So, we may assume that a simple word-to-word association can also be included in this definition.

On the other hand, Kintsch distinguishes between "proper" inferences that are "problem-solving" processes and "knowledge retrievals" in which a gap in the text is bridged by pieces of preexisting knowledge retrieved from memory (Kintsch, 1993, 1998). Moreover, he suggests that both types can occur in an automatic or controlled way. Knowledge retrieval occurs in an automatic way when information in short-term memory cues strong and relevant information in long-term memory expanding the capacity of working memory. An example is the activation of "cars have doors" by "A car stopped. The door opened". A control process, however, takes place when the short-term cues do not lead to the retrieval of the information needed to fill the gaps in the text. An example is finding, in a deliberate and controlled way, the causal link between the following propositions: "Danny wants a new bike. He worked as a waiter.". Both of these examples are not considered inference by Kintsch, because via knowledge retrieval the person is merely accessing information available in long-term memory. An inference takes place when *new* information is generated, *based* on the text and information from long-term memory. However, it is only considered a real inference when this new information is generated deliberately. The automatic activation of "the turtle are above the fish" by the sentence "Three turtles rested on a floating log, and a fish swam beneath them." is not an inference. However, contrary to "cars have doors", "the turtle are above the fish" is not information that already existed in the long-term memory, it is newly generated information during the comprehension process. Kintsch argues that there are statements that have particularities that lead to strong representations (*e.g.*, strong mental images) and this might lead to the automatic generation of new information. However, only a deliberate generation of new information can be called an inference. An inference, according to this author, is a result of effortful and explicit reasoning. This process only comes into action when the network is not capable of integrating the information, and there is no other way to fill

the gaps. It is when the "reasoning is called for as the ultimate repair procedure" (Kintsch, 1998, p. 192).

In an attempt to apply this same logic to STI, inferring a trait about a specific person can be considered as generating new information, because that specific person being categorized as "smart" does not exist, *a priori*, in the long-term memory of our participant (unless he is doing the same experiment for the second time). But, as it has been discussed in previous sections of this introduction, the STIs aren't deliberate and so, they do not qualify as a real inference in Kintsch's taxonomy. At most they can be considered an automatic generation of new information. In regard to word-to-word activations, it is clear that, by using Kintsch's strict criterion, these are not inferences, even though this specific aspect is not directly addressed by the author.

Keenan and collaborators (Keenan, Potts, Golding, & Jennings, 1990; Keenan & Jennings, 1995) focused specifically on the word-to-word activation issue. For that matter, they distinguish two ways of activating a trait. One way is via the comprehension of the meaning of the sentence as a whole, via text-based priming. This kind of activation is considered to be a real inference, since not only the individual words in the sentence are processed, but the general meaning of sentence is processed and integrated with information stored in memory.

In terms of STI, the behavior as a whole is processed/comprehended and relevant knowledge such as categories of behaviors, intentions, characteristics of people get activated and integrated into the representation of the encoded event. However, the second way of trait activation, is via simple word-to-word activation, such as "chess" or "library" activating "smart". Information, that is not part of the sentence, is also being retrieved from memory in this case. The processes taking place, however, are very different from text-based inferences. In word-based priming the general meaning of the sentence as a whole might not even be processed at all. Keenan and colleagues do not consider the word-based activation as a real inference and this dissertation embraces this same perspective. They also claim, as opposed to McKoon and Ratcliff, that elaborative inferences are implicitly drawn and do occur at the moment of the encoding, the main difficulty, from their point of view, being encountering a precise and uncontaminated measure of inferences (see section 1.5).

A possible solution for word-to-word associations, suggested by Keenan and colleagues, is by using control sentences that contain the same words as the trait-implying sentences but rear-

ranged in such a way that the traits are not implied anymore (Keenan et al., 1990). This topic is discussed in more detail in Chapters 5 and 6.

We should add that in STI, the story does not end here. For a real dispositional inference to occur, the trait inferred from the behavior has to be integrated into the representation of the actor that performs that behavior. The representation on the actor must change, must become richer than before the inference is done because a new property is now part of this representation.

1.5 PARADIGMS AND THEIR LIMITATIONS

Regardless of the infancy of the STI research, there are already a considerable amount of paradigms developed. In this section, we are going to discuss the main paradigms, their advantages and their limitations.

Regarding the moment the inference is measured, paradigms can be classified into two categories: those where the inference is measured while the participants are reading the behavior or immediately after and those where other intervening, unrelated material are processed between the reading of the sentence and the moment the inference is tested (Keenan et al., 1990).

A different way to categorize inference measures is by type, as opposed to the timing the inference is tested. There are memory measures, where the participants have to contrast their memory of the behavioral description with a target, and there are activation measures, where the activation of the inferred concept is measured (Keenan et al., 1990). Cued-recall and recognition are memory measures. Lexical decision, reading time (naming) and Stroop tasks are activation measures. For example, the before mentioned cued-recall is a delayed measure because all the sentences are seen first and the cues are presented for recall only afterwards, and it is a memory test because the participants are instructed to recall the behaviors.

One difference between activation and memory measures is that activation measures are thought to access on-line inferences, whereas memory measures to access more stable inferences that are incorporated into the final representation of the event (Kintsch, 1988). The timing here might be critical if we assume that for an inference to be incorporated into the final representation more time is needed. Indeed, there are authors suggesting that some inferences take time to develop and thus, immediate measures may yield no evidence of inference not because the subjects do not

draw inferences but because the test occurs before the inference in complete (Calvo & Castillo, 1998; Till, Mross, & Kintsch, 1988; McKoon & Ratcliff, 1989b). Calvo and Castillo's studies (1998), where inferences about predictive events were investigated, support this idea. In these studies, after presenting a predictive or a control context to subjects, a target that could confirm or disconfirm the predictive event followed. When the naming of the target happened 1500 ms after the onset of the last word from the inferential (as opposed to control) context, a predictive inference was observed. Only a tendency was detected when the naming happened after 1000 ms and no inference was detected when the naming of the target was 500 ms after reading the inference material.

There are also other differences between the paradigms, one important difference being the type of material used. Some researchers use paragraphs with multiple sentences (*e.g.*, Carlston & Skowronski, 2005) whereas others use more elementary contexts like single sentences (*e.g.*, Todorov & Uleman, 2002). Finally, another important difference is the type of control sentence used. Some researchers use neutral sentences as controls (*e.g.*, Van Overwalle, Drenth, & Marsman, 1999; Wigboldus et al., 2003) and only a few use rearranged sentences (Uleman, Hon, Roman, & Moskowitz, 1996) in order to control for word-to-word associations.

In this section we analyze each paradigm individually.

1.5.1 *Cued-Recall*

In the cued-recall paradigm (*e.g.*, Winter & Uleman, 1984; Claeys, 1990; Uleman, Winborne, Winter, & Shechter, 1986), after participants are exposed to trait-implying descriptions, they are instructed to recall the description while provided with different types of cues. What is usually found is that the trait is an effective cue to recall the description and just as effective as strong semantic associates.

However, researchers like D'Agostino and Beegle (1996) argued that part list cueing effects can easily explain the result (recalling some items from a memorized list inhibits the recall of the rest of the items). To demonstrate their point, these authors conducted a experiment similar to Winter and Uleman (1984) but additionally to the within-subject manipulation of the type of cue, they also included a between-subject manipulation of the cue. Corroborating their hypothesis that the

effect resulted from part list cueing, they verified that the effect only occurs in the within-subject manipulation. Another problem related to this and other memory measures is the uncertainty that the effect detected is due to encoding processes or rather to reconstructive processes taking place during retrieval. To better illustrate this limitation, Singer (1978) used sentences containing actions, such as "stir the soup" and he gave, as cues, two different instruments (ladle or spoon) that were asymmetrically related with the action. The forward association between "stir the soup" and "spoon" is strong but the backward association is rather weak, while the backward association of this action with "ladle" is stronger than the forward association. If the instrumental inference occurs at encoding, then spoon should be a better cue than ladle. If the cued-recall reflects reconstructive processes at retrieval, then the ladle should lead to better recall. Singer found that the second scenario was true, meaning the cued-recall reflected mostly reconstructive processes. In similar ways, when the trait is given as cue, if the backward association between the trait and the behavior is strong, this reconstructive process might be an alternative explanation for Winter and Uleman's findings (1984). Wyer and Srull also defended that traits might trigger typical behaviors, that, in turn, might cue the recall of the seen behavior (1986).

Finally, a limitation we are going to further explore in Chapters 5 and 6 is related to how the trait gets activated during the reading of the sentence. In the cued-recall paradigm there is no way to know whether the trait inference is a result of word-based priming or text-based priming because no appropriate control material is used.

1.5.2 *Recognition probe paradigm*

The probe recognition paradigm is another very commonly used paradigm to study trait inferences (Newman, 1991, 1993; Uleman et al., 1996; Van Overwalle et al., 1999; Wigboldus et al., 2003, 2004). In this paradigm, participants read implying and non-implying paragraphs and after reading the last sentence in the paragraph they are shown a probe word. The task is to say whether or not the word occurred in the preceding paragraphs. Participants are expected to be slower in correctly rejecting trait probes and to make more inaccurate responses following trait-implying paragraphs than non-implying controls (Uleman et al., 1996). The inference makes it

harder to accurately reject the probe as having been presented because once inferred, the concept becomes part of the overall mental representation of the text (McKoon & Ratcliff, 1986).

For example, the probe "cautious" is easier to reject when preceded by the non-implying sentence "Everyone started off before checking their seat-belts" than when preceded the trait-implying version "He checked everyone's seat-belts before setting off". Note that in the controls used in this task, an effort is made to keep the same words from the trait-implying sentences, so that the two versions are equated for word-based activations (Uleman et al., 1996).

Forming an inference in this paradigm not only does not aid performance, but it leads to more incorrect responses and slower ones. This is an improvement comparing to the cued-recall paradigm. The main problem in this paradigm is that there is no easy way to know whether the inference occurs at encoding or at retrieval. There are two ways for the inference to occur only at retrieval (Keenan et al., 1990). One way is by drawing the inference while checking the probe against the sentence. For example, the subject might have not activated the trait during the reading of the sentence, but when the trait is presented one can conclude that it would have been an appropriate trait to infer. Making this inference during the test would slow the RTs and would lead to a similar result as the one expected if the trait would have been inferred at encoding. Thus, presenting the probe word at test is a second opportunity for participants to (mentally) review the sentence and thus, make the inference at retrieval rather than encoding. McKoon and Ratcliff (1986) suggested the use of a deadline at the test moment so that the possibility of making the inference at the retrieval is eliminated or decreased. Unfortunately, the use of deadline in STI studies was not adopted on a regular basis (but see *e.g.*, Van Overwalle et al., 1999). Another way to obtain a similar result is without making an inference either at encoding or at retrieval, but due to the greater compatibility of the trait-implying sentence with the probe than the compatibility of the control sentence with the probe. Context checking occurs when participants compare a probe to their internal memory of the text (Forster, 1981). The inevitable closer fit between the trait-implying sentence and probe increases the likelihood of stating that the probe featured in the original sentence, a response that perfectly mimics an inference made at encoding. This limitation is, in part, solved by control sentences that have the same words as the trait-implying ones. However, this only makes the compatibility identical at the word level and not at the level of higher order structures such proposition and schemas.

1.5.3 *Saving in relearning*

Saving in relearning (*e.g.* Carlston & Skowronski, 1994; Skowronski, Carlston, Mae, & Crawford, 1998) is a paradigm Carlston and Skowronski (1994) adapted from the saving paradigm developed by Ebbinghaus (1964). As already described in this introduction, this paradigm has an initial exposure phase where participants are shown actor photos paired with trait-implying paragraphs while instructed to familiarize themselves with the material. Next, in the learning phase participants are presented with pairs of traits and photos. Some are relearning pairs since the photo is presented with the trait that was implied in the paragraph seen with that picture, and some are control pairs meaning the traits and the photos are new. In a later testing phase, participants are presented with photos from the learning phase and are instructed to recall the trait presented with it. If the trait is inferred during the exposure phase, then the learning of the trait will be easier in the relearning trials than in the control ones, where the trait was not previously activated by a behavioral description.

These results can also be explained by retrieval processes. Upon the presentation of the photo-trait pair in the learning phase, the participant may try to remember the behavior to help memorize the trait. Here too, the compatibility between probe and sentence may contaminate the results as the trait word is naturally a good match to the trait-implying sentence and no control of the word-based priming is included.

In the saving in relearning experiments, instead of presenting a single sentence in each trial, multiple sentence paragraphs are presented to the subjects. These paragraphs are rich and consistent with each other. Park's work showed that pairing multiple trait-implying sentences with faces is enough to trigger impression formation goal (1989). This is in agreement with the fact that it has been difficult to find differences between formation impression instructions and mere familiarization instructions with this paradigm (Carlston & Skowronski, 1994).

1.5.4 *False recognition paradigm*

In the false recognition task, participates, in an initial phase, are presented with pairs of photos of people and trait-implying sentences under memory instructions (*e.g.*, Todorov & Uleman,

2003; Goren & Todorov, 2009). Some of the sentences in this phase explicitly mention the implied trait. Later on, in the test phase, the same photos from the first phase are paired with traits implied in the sentence presented with those photos - match trials, or new pairs are presented. In the new pairs, previously seen photos are paired with traits previously implied alongside different photos (also called mismatch trials) or the photos are paired with new traits. Participant's task is to indicate whether the trait was part of the sentence presented in the first phase together with that photo. There are usually more false recognitions in the match than in the mismatch/new trials, suggesting that inferences occurred upon reading trait-implying sentences and that the traits are bound to the actors.

Just like in the saving in relearning paradigm, the advantage of this paradigm is in showing that the trait is actually linked to the actor and is not a mere categorization of the behavior. This conclusion is impossible to make in studies using cued recall or probe recognition.

Because this is a memory measure, the criticism that the inference can result from retrieval processes also applies here. The trait has to be contrasted against the memorized sentence, a period during which the person can draw the inference. On the other hand, we know that the same pattern is obtained even when the number of trials to be memorized goes as high as 120 photo-sentence pairs or when the presentation time is very reduced (Todorov & Uleman, 2002), making it very difficult for the participants to actually recall the behaviors. However, the inference can be made at retrieval even without an actual recall of the behavior because the presentation of the trait can lead to the generation of behaviors representative of the trait which might cue the recall of the learned behavior. The same can be said about context checking, especially since in this case (contrary to what happenes in some studies using probe recognition) no control for word-based priming is used. When the participant is trying to contrast the information in the test with the one in the learning phase (or the information from the learning phase that the participant remembers), the relationship between these two pieces of information might lead to more false recognitions. However, while the word-based activation can be a serious problem at the retrieval moment because it increases the compatibility of the trait with the sentence, the word-based activation of the trait at encoding might tend to dissipate with time and thus its role might be residual in delayed measurement of inference as this one is (see more in Chapter 6).

There are other variations of this recognition task. For example, D'Agostino (1996) reports two studies where a delayed recognition test is also used without presenting the actors' photos. The author presented trait-implying sentences, from which half implied the trait while the other half explicitly mentioned the trait in the sentence. Half the participants studied the material for a memory test and the other half to form impressions. The recognition test contained implied traits, explicitly mentioned traits and unrelated new traits. The error rates were higher for implicit than explicitly mentioned traits. This measure too, suffers from explicit retrieval contamination and context checking limitations.

The contamination of implicit measures with explicit recall is a recurrent problem in implicit memory literature. The implicit memory refers to a situation where previously learned information affects performance in a posterior task with no intentional or conscious recollection of that previous material. Implicit memory is said to be contaminated when the participants adopt an intentional retrieval strategy or if they are aware that the tested material was part of the study list. These are usually referred to as the intention and awareness problems, respectively (Butler & Berry, 2001). Intention and awareness can be easily controlled in patients with explicit memory impairments, but are very difficult to control in healthy participants. Roediger and McDermott (1993) outlined a number of precautions in order to control for the contamination problem. To avoid contamination, the test must not require any explicit revision of the learned material, meaning explicit memory measures (based on recognition and recall) should be avoided because all of them require the subject to contrast a test target or a cue against the implying context. However, even when this aspect is controlled, the subject can become aware of the relationship between the material tested and the one studied. One way to control for this is by directly questioning the subject (using a structured interview or a questionnaire) at the end of the experiment and exclude those subjects that are aware of this relation. Also, it is possible to obscure this relationship by using a large amount of fillers aside from the experimental material.

In summary, memory measures of inferences can inadvertently encourage inference making at retrieval due to contamination from explicit and conscious retrieval. As a result, Keenan and colleagues stated that the best way to measure inferences that occur during encoding is by using "a test that does not require subjects to evaluate the probe against the text" (1990, p. 389). One way to achieve this is by using activation measures. Activation measures of inferences, as the

name indicates, measure the activation of a concept, without any explicit need to retrieve past information. Despite the general benefits of activation over memory measures, the activation tasks differ from each other in a variety aspects.

1.5.5 *Lexical Decision*

The lexical decision is one of the fewest activation measures used in the STI literature. It is a common measure for inferences in text comprehension literature (*e.g.*, Potts et al., 1988; Ratcliff & McKoon, 1988; Lucas, Tanenhaus, & Carlson, 1990). However, in STI literature, as far as we know, there is only one published paper that uses lexical decision.

Zárate and colleagues (2001) investigated cultural differences in the use of traits under the hypothesis that less collectivist cultures make more use of traits. Participants in these studies were instructed to memorize sentences. Each sentence was followed by a lexical decision where the participants had to indicate as quickly and accurately as possible if the presented string of letters was a word or not. There were two types of critical trials and in both the stings presented for lexical decision were trait words. In one of the critical condition the sentences preceding the traits implied those traits and in the second critical condition control sentences, unrelated to the traits, preceded the lexical decision.

The main limitation Zárate and colleagues' studies present is the control material. Their control condition does not control for word-based priming. This is a crucial aspect because lexical decision is an immediate measure, which does not allow for the word level activation to dissipate. As a consequence, we don't know if the trait inference is a result of processing the meaning of the behavioral description or a result of word-based activations. Even if an effective control for word-based priming would have been included, there is another problem. Lexical decision can be decomposed into two elements: the lexical access and the decision process. At a first glance, the subjects in this task don't have a reason to compare the string to the sentence, suggesting that the relatedness between the two shouldn't contribute to the effect. However, there are studies showing that such a comparison occurs (e.g, West & Stanovich, 1982; Balota & Chumbley, 1984; Neely, Keefe, & Ross, 1989). These studies showed that the backward associations between the target and prime affect lexical decision latencies. Context checking seems to happen after the

lexical access in order to facilitate the decision process and if the target is somehow related with the prime/sentence the decision to say “yes is a word” will be encouraged.

Finally, if the lexical decision is to be applied in a delayed fashion, that is, after reading all the sentences, the context checking would be more difficult to occur because subjects would not have the learned material present in their minds. This modification would also allow us to detect inferences that need more time to develop.

1.5.6 *Word Stem Completion*

Finally, there is one last paradigm from STI literature that we would like to describe - word stem completion. Whitney and Williams - Whitney (1990) used constrained word stems to detect on-line (*i.e.*, drawn at encoding) inferences. Participants started by reading trait-implying or control paragraphs and then completed word stems (*e.g.*, "C L _ _ _ _" could be completed with the trait CLUMSY) with the first words that came to their mind that would fit into the blanks. The stems were completed with the relevant traits more often when following trait-implying paragraphs than control paragraphs, allowing the researchers to conclude that inferences occurred online. In this activation measure, an effective control for word-based priming is used (with rearranged versions of the trait-implying paragraphs). Also, the measure was shown to be unaffected by backward associations, and because of that it can be considered a better measure than lexical decision (Whitney, Waring, & Zingmark, 1992).

In a variation of this task, word fragments (*e.g.*, "C _ U _ S _" for CLUMSY) are used instead of constrained word stems, but the logic behind it is similar (Whitney et al., 1992; Bassili & Smith, 1986).

Although we think this measure is more effective in measuring on-line inferences when compared to the previous ones, there is one limitation we would like to point out, a limitation that affects also all the previously mentioned paradigms. The limitation is related to the fact that the answer the participant is required to give is a direct reaction to a specific trait (previously tested). It is a direct measure of the activation of that particular trait. However, it is not obvious that all the participants infer the tested trait and even if they do, we don’t know whether they do it always. For instance, in Whitney and colleagues’ pilot study (1992), the behavioral descriptions

were presented to participants for a trait generation task, and the traits used as targets in the stem completion task were given by the subjects in average 54% of the time (with the least consensual being generated 35% and the most consensual 90%). This clearly shows that the participants do not all infer the same traits. However, the same target traits (or stems) are presented for all the participants. When a specific trait is used in the test phase (making the effect depend on the lexical activation of that specific word) and if, for example, a synonym is activated at encoding instead, the lexical facilitation is not expected to occur anymore (or inhibition might occur due to competition between lexical formats). However, at the conceptual level, the same representation might be activated. In more extreme case, the activated representation might not even overlap with the meaning of the target expected to be given as a stem completion.

Also, according to the minimalist hypothesis (McKoon & Ratcliff, 1986, 1989c, 1990a), inferences that are not necessary for coherence (and we think trait inferences are not) can just be minimally or partially encoded. Thus, the trait might not be totally inferred, because it is encoded "minimally by a set of features or propositions that do not completely instantiate the inference" (McKoon & Ratcliff, 1992, p. 458). In this case, we believe that it is more adequate to talk about the activation of a semantic area or network and not of a specific trait, since the trait itself might need some more aid to be inferred (McKoon & Ratcliff, 1986). In other words, for trait inference to occur, some more help might be necessary to make the activations converge to a specific and relevant trait.

Conceptually-driven tasks that are more dependent on top-down mechanisms, instead of being uniquely sensitive to the activation of a specific trait are more capable of detecting the activation of a network. The use of a conceptually-driven task might help us to overcome this limitation.

The distinction between data-driven and conceptually-driven tasks is not something new. Some researchers, instead of focusing on the distinction between explicit and implicit memory, changed their focus to the type of processing. They started to distinguish between tasks in which subjects rely more on physical features of the stimulus, data-driven tasks, and those with minimal focus on the physical features and where conceptual based, top-down processes are used, conceptually-driven tasks (Roediger & Blaxton, 1987). This perspective is presented as a continuum rather than a dichotomy. As such, it is assumed that any task can rely on both types of processes to some degree. This perspective is framed in the processing framework that stresses that memory

can benefit if the process at encoding is similar with the one required at retrieval (Roediger, Weldon, & Challis, 1989). This is an important framework for trait inferences since an inference can be considered a conceptually-driven process. Also, the perceptual features of the trait are not important during encoding because the trait is not actually presented. Thus, the best way to test a conceptually-driven phenomenon is with a conceptually-driven test. Measures like recognition and free recall are usually considered conceptual tests, but as discussed above they suffer from the contamination problem. While activation measures solve in part the contamination problem, they rely heavily on data-driven processes, which are not appropriate to measure top-down mechanisms.

Roediger and collaborators (Roediger et al., 1989) also proposed an operational definition to classify tasks along this continuum. They present the generation effect (Slamecka & Graf, 1978) to illustrate the comparison between data-driven and conceptually-driven tasks. In a typical generation effect study, two conditions are compared, one where the critical stimulus has to be generated by the participant as a response to a semantic cue, and a second condition where the subject only reads the critical stimulus. Reading is assumed to involve data-driven processes, whereas generating the stimulus is thought to involve conceptually-driven processes, just like in STI studies where the trait is generated (with the exception that in STI subjects are not instructed to do so). The generation effect is presented as the benchmark of conceptual processing.

In this dissertation we argue that all the activation measures mentioned so far are closer to the data-driven pole of the continuum than to the conceptual pole (*e.g.*, Hamann & Squire, 1996; Roediger, Weldon, Stadler, & Riegler, 1992) because they rely on the activation of a specific trait form.

Some memory measures, especially recall and cued-recall, rely more on the right type of processing, a conceptually-driven one, even though the dependency on specific trait formats also apply to memory measures. Moreover, they are contaminated by explicit retrieval processes. Activation measures are less affected by contamination but are heavily data-driven, meaning they are more reliant on superficial features of the stimuli. Another characteristics of activation measures is that they are usually applied immediately after encoding. McKoon and Ratcliff suggested that activation measures are not capable of detecting inference because the computation of an inference might take longer than the time window that these immediate measures rely on (1990b;

1983). In Kintsch's model, time is also crucial for the inference to be integrated and incorporated in the final representation of the event (1988). In this sense, delayed measures should be preferred to immediate ones, so that a trait inference that needs more time to be integrated is also detected.

To deal with all these limitations presented by both activation and memory measures, we suggest a new activation measure, that, contrary to the previous activation measures, is a purely conceptually-driven task. In this task, subjects are not required to retrieve learned material, so contamination from explicit processes is less likely to occur. It is a measure that can be applied in a delayed manner, so more stable and late representations are also accessed and it is a measure that, if applied with the correct control material, controls for word-based priming.

The modified word association paradigm was developed by Hourihan and MacLeod and presented as a conceptual implicit task (2007). In their studies, they contrasted the effect of generating words from meaningful cues (*e.g.*, "the piece of furniture used for sitting?") with reading the words (*e.g.*, "chair"). In the test phase participants had to produce an associate to words they had generated versus read in the learning phase, as fast as possible. Subjects were instructed to say out loud the first word coming to their mind upon seeing the prompt. The logic of the task is that generating a word in the learning phase will conceptually prime that word in implicit memory, and this priming will spread to neighboring words. Thus, the word and its associates will become more easily available when compared to the words that were just read. Consequently, during the free association task, the subject will access the given word's neighbors and give an answer faster. The same logic can be applied to trait inference. If during the reading of the sentence the trait is spontaneously generated as a consequence of processing the meaning of the behaviors, then the trait is being conceptually primed and so are its neighbors. As a consequence and because the semantic net of the inferred trait gets strongly activated, the free association performed later on should be facilitated upon the presentation of this same trait. The subject is never instructed to think back to the memorized material and the two phases (memorizing the material and the free association) are presented as unrelated. This paradigm is presented in detail in Chapter 7.

Finally, in the remaining part of this introduction we review an error that is a byproduct of trait inference - the spontaneous trait transference (STT) - and how it opened doors to an important theoretical debate in the field.

1.6 SPONTANEOUS TRAIT TRANSFERENCE

As in many other cognitive processes, errors can occur and STI is not an exception. STT occurs when an inferred trait is associated to the wrong person, that is, the trait is transferred to someone that is not the actor of the behavior. The transference of traits is not something completely new in social cognition. It is known that a trait (*e.g.*, "kind") can be transferred from the person that initially activated the trait (*i.e.*, the actor) to a third party that is similar in some way with the actor (even if the similarity is completely irrelevant to considerations regarding personality, *e.g.*, long hair) (Lewicki, 1985, 1986). The so-called kill-the-messenger effect, also tells us that the transmitter of a message often becomes associated with the valence of the message he conveys (for a review, see Walther, Nagengast, & Trasselli, 2005). For example, in a study conducted by Manis, Cornell, and Moore (1974), it was shown that listeners evaluated a transmitter of a message more favorable when the message supported listener's view than when it did not. The effect emerged even though the listeners knew that the transmitters did not necessarily agreed with the messages.

In 1995, Carlston and colleagues (Carlston, Skowronski, & Sparks, 1995), using the saving in relearning paradigm, showed that just by presenting a trait-implying description and an irrelevant person on the same screen, these two become associated. They presented participants with trait-implying descriptions paired with photos of communicators offering descriptions of other unseen people instead of pairing them with the actors of those behaviors (as it is usually done in STI trials). They found saving in relearning, that is, the trait inferred about the actor was being associated to the communicator, even though the STT effect was almost half the effect of STI.

Later on, Skowronski and collaborators (Skowronski et al., 1998) showed that STT also affects explicit ratings of the communicator in regard to the transferred trait. In this same paper, the authors showed that even when any plausible similarity between the personality of the actor and of the communicator is eliminated, by saying that the description was randomly paired with the person, STT still occurred (Experiment 3). Furthermore, Brown and Bassili (2002) showed that the trait can be associated even with inanimate objects. These results suggest that this is an associative error because no logical reason can be attributed to the established link.

Another question that needed an answer was whether the STT effect could result from a poor memory about who is the actor and who is the communicator. If the participant would mistakenly remember a STT trial as being a STI trial, this could explain the effect. This confusion between STI and STT trials could happen due to two different reasons: the lack of attention during the encoding of the trials, or the forgetting taking place during retrieval. In order to minimize the confusion at encoding, Carlston and Skoworonski (2005) manipulated the gender in such a way that the gender of the communicator was different from the gender of the actor whose behavior was being described. Moreover, the authors gave extra time to process each trial (20 ms). None of these manipulations significantly affected the STT effect, which suggests that if there is some confusion between the trials, it has to be residual. By the same token, it was shown that STTs are verified in between-subject manipulations (Carlston & Skowronski, 2005; Crawford, Skowronski, Stiff, & Scherer, 2007) that also suggests that the hypothesis of erroneous encoding does not hold. In order to discard the possibility of STT being a result of forgetting the nature of the trial (actor versus communicator), during the retrieval, Carlston and Skowrosnki (2005, Experiment 2) used an on-line rating scale, that is, the rating was performed trial by trial. This manipulation makes the role of the person less likely to be forgotten. This experimental setting also led to the detection of the STT effect.

Skowronski and collaborators (1998) explain the occurrence of STT via three different steps: 1) The trait is spontaneously activated during the reading of the description, 2) an erroneous association is created between the trait and the wrong person, and 3) the associated trait influences the subsequent judgements about the person.

In subsequent studies, STT has been shown to be a robust effect. STT seems to be relatively resistant to processing goals (Skowronski et al., 1998; Carlston & Skowronski, 2005; Crawford, Skowronski, Stiff, & Scherer, 2007), occurs when participants are warned about the existence of such an effect and are asked to avoid it (Skowronski et al., 1998; Carlston & Skowronski, 2005, Experiment 3), is observed under cognitive load (Crawford, Skowronski, & Stiff, 2007), and is detected even when the association is tested with two days delay (Skowronski et al., 1998). Moreover, it was also shown that STT does not depend on the recall of the behavior, since there is saving in relearning even in the absence of both recognition and recall of the behaviors (Carlston & Skowronski, 1994; Carlston et al., 1995). These results demonstrate how efficient STTs are

and also how independent they are from the characteristics of the stimuli the trait gets attached to. However, they do have boundaries.

Crawford and colleagues (Crawford, Skowronski, & Stiff, 2007) presented both the face of the actor and the face of the communicator simultaneously together with the behavior and saving in relearning was observed only for the actor of the behavior, eliminating the STT effect. Todorov and Uleman (2004) conducted a set of similar experiments, using the false recognition paradigm, where again, the actor and a randomly paired person were presented simultaneously together with the behavior. Along five studies, they showed that STI was always stronger than STT. However, the STT was never reduced to zero by the presence of the actor's face within this paradigm.

1.7 SPONTANEOUS TRAIT INFERENCE VERSUS TRANSFERENCE

The research about STT brought a new urgency to the field, the urgency to distinguish STI from STT, in both empirical and theoretical terms.

Empirically, there are clear differences between STI and STT. First, the magnitude of the STI effect is usually larger than the magnitude of the STT effect (Goren & Todorov, 2009; Skowronski et al., 1998). Second, studies support that STIs are sensitive to implicit theories of personality, and as such, a *halo* effect (generalization from a trait to others that are congruent in valence) was detected in STI and not in STT (Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski et al., 1998; Wells et al., 2011). Third, sometimes a negativity effect (more inferences of negative traits than positive traits) is observed in STI and not in STT (Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007). Fourth, if participants have to perform a concurrent inferential task, such as detecting if the communicator is lying (about their own behavior or someone else's), STIs are reduced while no decrease is observed for the STTs.

Finally, there are studies where the STT effect is eliminated while the STI effect is kept unchanged (Todorov & Uleman, 2004; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007; Goren & Todorov, 2009). One manipulation that leads to the elimination/decrease of STT is presenting both the actor and the communicator simultaneously to the participant while the behavior is being processed (Crawford, Skowronski, & Stiff, 2007; Todorov

& Uleman, 2004). The second manipulation that eliminates STT is by presenting the behavior and the actor/communicator separately (Goren & Todorov, 2009, Experiment 3). If the photo is presented first (without any reference to whether it is an actor or a communicator) and the behavior together with the information regarding the relevance of the face is presented only after the disappearance of the face, the STT effect is eliminated while the STI is not affected.

1.8 ASSOCIATION VERSUS ATTRIBUTION

After showing that the trait can become associated with any stimulus presented during the encoding context, the scientific community started to reveal a generalized concern about the processes responsible for STI and, in particular, how are they different from the ones responsible for STT. The critical aspect in this discussion is the kind of link that is created between the trait and the actor. Is it an inferential link, meaning that the trait is integrated into the representation of the actor, or is it an associative link, *i.e.*, the trait is blindly linked to the stimulus salient in the context (that can be the actor or not)? Answering this question has been challenging, mainly due to lack of conceptualization and clear theoretical distinction between inference and association.

The empirical differences between STI and STT described in the section above, are usually used to argue that STIs are different from STTs in terms of their underlying processes. In regard to this, two main perspectives have been proposed. The first follows the parsimony principle, and suggests that both phenomena could be explained by associative processes (Bassili & Smith, 1986; Bassili, 1989, 1993). The second perspective suggests that while the STT results from an associative link between the irrelevant person and the trait, the STI results from attributional processes (Carlston & Skowronski, 2005). Moreover, it is proposed that the associative link responsible for STT is a consequence of incidental spatial and temporal contiguity. The resulting link is said to be unlabeled, meaning it does not contain any information regarding the relation between the representations linked. Attributional processes are, in contrast, said to depend on more elaborative and deep processing that requires causal thinking. This link is stronger and defines the relationship between the trait and the actor; the trait is seen as an attribute of the actor.

At a first glance, the differences presented in the section above might appear to be in agreement with the dualist perspective (the one defending different processes for STI and STT). However, most of these differences can be explained by an associative account. For example, stronger STI, in comparison to STT can result from stronger associations between actors and traits because the actors are more salient than communicators in the context of the behavior (for more see Chapter 2 or Orghian, Garcia-Marques, Uleman, & Heinke, 2015).

A tenet of attributional theory is that people are motivated to understand their social environment. In his book, Heider, considered by many the father of attributional theory, writes as follows while introducing the subject of his book: "Our concern will be with "surface" matters, the events that occur in the everyday life on a conscious level (. . .)." (1958, p. 1). Of course, unconscious processes at that time had a different connotation (related to psychoanalysis) than it has in our days, but what is clear from this statement is that Heider intends to describe conscious processes. Using these same attributional rules introduced by Heider, to describe STI, a spontaneous and unintentional phenomenon, seem inappropriate to us.

Moreover, the use of attributional theories to explain STI is in contradiction with previous research about the distinction between dispositional inferences and causal attributions. At first, causal attribution was hand in hand with dispositional inferences (*e.g.*, Jones & Davis, 1965; Heider, 1958; Kelley, 1967). Later on, person perception models started to incorporate inference and causal attributions into different processing stages (Trope, 1986; Gilbert et al., 1988; Quattrone, 1982). These models define dispositional inference as an earlier stage, whereas the causal attribution, *i.e.*, the adjustment (Quattrone, 1982), the subtracting rule (Trope, 1986), or the correction (Gilbert et al., 1988), as happening later on. The distinction between dispositional inference and attribution is not, however, very well described in these models.

Attribution is usually seen as the explanation or the judgement about the cause of some event, and in the social realm, it is the judgement people make about the behaviors of others (Krull, 2001). Some research shows that this search for a cause might be distinct from dispositional/correspondent inference and might rely on different mechanisms (*i.e.*, Bassili, 1989; Krull, 2001; Erickson & Krull, 1999; Hamilton, 1988; Smith & Miller, 1983). For example, Johnson, Jemmot and Pettigrew (1984) found that knowing that a behavior was caused by situational forces did not prevent subjects from drawing correspondence inferences. This finding impelled the re-

searchers to suggest that trait inference might be independent of causality. Bassili, after finding that participants instructed to allocate causality showed little trait inference when compared to formation impression instructions, stated that "causal allocation places very little emphasis on trait concepts" (1989, p. 293). Moreover, Erickson and Krull (1999) found that the extremity of a behavior predicted the dispositional correspondence bias more than it predicted the causal attribution.

Not only does causal attribution seem to be different from dispositional inferences, it is also thought to be less common than dispositional inferences and to occur in very specific conditions. For example, Hamilton (1988) suggested that causal attributions are drawn when the behavior of the actor is unexpected. Thus, the trait inference is seen as a relatively quick and effortless categorization of the behavior that contributes to the the general impression of the actor. On the other hand, causal attributions might be slower and trigger more controlled judgements when needed. In other words, causal attributions are triggered when unexpected information has to be integrated in to the actor's representation. It was also shown that the time to make an inference is smaller than the time to make causal attributions (Smith & Miller, 1983).

This distinction between causal attributions and dispositional inferences is in agreement with the stage models mentioned before. However, these models vary in the way the dispositional inference is thought to occur. For example, in Trope's model, the dispositional inference is assumed to happen in a controlled fashion and after the behavior is categorized. The categorization of the behavior is assumed to be automatic, whereas using the trait to describe the actor might or might not occur (Trope, 1986).

Gilbert and colleagues' model (1988) is in more agreement with the current view on STI. They claim that person perception happens in three stage: 1) the behavior is categorized in terms of traits, *categorization*; 2) the person is categorized in the same terms, *characterization* and finnaly 3) a correction can or can not be applied to this categorization based on additional situational information, *correction*. The last stage is the most effortful, less spontaneous and a much more elaborate stage than the STI is. The first two operations are said to occur outside awareness and to be efficient. The researchers provided evidence regarding the efficiency of dispositional inference and the need of resources to correct these dispositions. They showed that under low cognitive load, participants are able to use contextual information to correct the dispositional

inferences. The participants had to rate the anxiety of an anxious looking women from a video while she is discussing relaxing versus anxious topics. If under low cognitive load, participants took into account the topic she was discussing when rating her anxiety, while in the high load condition, they failed to consider this situational information, basing their decisions merely on dispositional inferences (Gilbert et al., 1988).

Note, however that this pattern can be easily reversed. In a different study, participants were informed that the woman in the video was either anxious or calm, and they had to indicate how anxious of a topic she was discussing (Erickson & Krull, 1999; Krull, 1993). The results showed that subjects under cognitive load rated the anxiety of the topic without taking into account the disposition of the actor. This result is interpreted based on processing goals. If the participant has a situational processing goal, that will be the most immediate inference he will draw and correction will be needed to incorporate the disposition in order to create an accurate impression of the event. Thus, this perspective puts more emphasis on the salience of different cues in the environment, such that, if the emphasis of the experiment is dispositional, the anchoring will be the disposition, and if the requirement is situational, the anchoring is situational. Additionally, more recent studies like the one presented by Ham and Vonk (2003), demonstrated that many inferences, situational *and* dispositional are simultaneously drawn at an initial stage and they can even be incongruent. In a second stage this information will be integrated and thus, irrelevant information will be inhibited.

Regardless of this possibility in reversing the effect depending on the requirement of the experiment or context, we see Gilbert and colleagues' model and its first two stages as a close theoretical approximation to STI, and a better one than attribution is.

However, this model does not tell us how STI differs from STT. A question that is still open for debate. We believe that the way the trait information is integrated into the representation of the actor is different in STI and STT. The trait becomes part of the representation of the actor whereas for the irrelevant person, no such integration takes place. In both, STI and STT, the behavior is categorized and the trait is activated. However, in the STI, the person is also characterized (using Gilbert's terminology) in agreement with the behavior categorization (a friendly behavior, thus a friendly person). In STT, this second categorization might not occur, meaning that the trait is indiscriminately linked to any stimulus present in the context due to mere spatial and temporal

contiguity. STT can be seen as a pure consequence of the encoding specificity principle (Tulving, 1972), a result of the trait being extracted from the sentence while the photo of the irrelevant person is part of the context. Thus, the photo and the trait constitute effective retrieval cues for each other, because they are stored together; they are part of the same memory trace of the encoded event.

If the first stage in Gilbert and collaborators' model (1988) is common to STI and STT, then we should expect no difference between STI and STT in the way the behavior is processed. However, different levels of elaboration are expected around the person the behavior is being linked with. If the person is the actor, the person will be characterized in term of the same trait and thus, a representation of the actor detaining that trait will be created. If the person is not relevant to the behavioral description, no such representation is expected. Only an associative link is expected and this link is not exclusive to that person (Brown & Bassili, 2002). If this second idea is true, then the way the faces are processed should vary according to the relevance of the person to the behavior. We expect this to translate into different strengths of the the memory trace, with a stronger trace for the actor and a weaker one for the irrelevant person. These two hypotheses are tested in Chapter 4.

Another difference between STI and STT regards the salience of the stimuli, the actors are more salient stimuli than irrelevant persons. This might lead to different amounts of attention paid to each of them. Some stimuli are selected for more intense scrutiny, while other are attended to more superficially (Posner, 1994). Which stimuli to attend to more and which to attend to less is one of the problems any cognitive system has to solve in order to distribute its attention in an efficient way. When we see the actor of a certain behavior, it is obvious that this actor is more relevant in this context than if it would be presented with a randomly paired behavior or if it would be describing someone else's behavior. This salience is motivated by top-down considerations about the role of the person presented. This extra salience of the actor might translate into more attention paid to it than to the irrelevant actor. This hypothesis is tested in our Chapter 3.

There is one last aspect that distinguishes STI from STT. In STI trials, a richer and a single representation is being created at encoding with two pieces of information representing the same person: the physical image of the person and the behavior of that same person. In STT trials, the

information is more dispersed because there are two different people being represented, the one in the photo and the one that performs the action. Because there are two people being represented, it makes sense to assume that each of them receives less attention and as a consequence are less activated individually. In the STI, we actually have more information about one person, whereas in the STT, we have the same information distributed among two representations. In STI, the two elements are seen as part of the same entity, they are organized together in the same representation. And this is the representation that social psychologists studying trait inference are searching for.

1.9 OVERVIEW

The debate regarding the processes responsible for STI and STT and the evidence supporting the single process view and the dualistic view are discussed in Chapter 2. This debate is approached in a computational way by demonstrating that, both STI and STT can be replicated in an associative model. The chapter also suggests that attention might have a role in the way links between traits and actors are created.

Next, we follow-up on this suggestion, and in Chapter 3, the hypothesis that different attention is paid to the actor versus a control person (non-actor) is tested by using eye-tracking devices and a modified spatial cueing paradigm.

In Chapter 4 we test the idea that STI and STT do not vary in the way the behavior is processed. This idea is in agreement with the mechanism advocated in the model presented in Chapter 2, where we claim that the difference resides in the way the actor and the irrelevant person is attended to. Accordingly, in this chapter, we also found that participants show a better memory for actors' faces than for irrelevant faces.

In Chapters 5 and 6, we explore the word-based priming problem that is present in many paradigms used in the field. The influence of word-based priming on immediate and delayed measures is also explored. It is argued that immediate measures are more affected by word-based priming than the delayed measures. Thus, the use of efficient control material in these measures becomes crucial to detect real inferences.

In Chapter 7, we continue exploring methodological issues of the paradigms used in the field, focusing on the distinction between memory and activation measures and on the distinction between data-driven and conceptually-driven tasks. Moreover, we discuss how conceptually-driven tasks fit well with the mechanisms underlying trait inferences. Finally, we suggest a new conceptually-driven and implicit measure in which some of the limitations of memory measures and of the activation measures are fixed.

In Chapter 8, a general conclusion is presented where we summarize the work presented in this dissertation and its weaknesses.

1.10 REFERENCES

Allport, G. W. (1937). *Personality*. Holt New York.

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258–290.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? the role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency and control in social cognition. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1, pp. 1–40). Hillsdale, NJ: Erlbaum.

Bargh, J. A., & Thein, R. D. (1985). Individual construct accessibility, person memory, and the recall-judgment link: The case of information overload. *Journal of Personality and Social Psychology*, *49*(5), 129–146.

Bassili, J. N. (1989). Traits as action categories versus traits as person attributes in social cognition. In J. N. Bassili (Ed.), *On-line cognition in person perception* (pp. 61–89). Hillsdale, NJ: Erlbaum.

Bassili, J. N. (1993). Procedural efficiency and the spontaneity of trait inference. *Personality and Social Psychology Bulletin*, *19*(2), 200–205.

Bassili, J. N., & Smith, M. C. (1986). On the spontaneity of trait attribution: Converging evidence for the role of cognitive strategy. *Journal of Personality and Social Psychology*,

50(2), 239–245.

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.

Bruner, J. S. (1957). On perceptual readiness. *Psychological Review*, *64*(2), 123–152.

Butler, L. T., & Berry, D. C. (2001). Implicit memory: Intention and awareness revisited. *Trends in Cognitive Sciences*, *5*(5), 192–197.

Calvo, M. G., & Castillo, M. D. (1998). Predictive inferences take time to develop. *Psychological Research*, *61*(4), 249–260.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*(5), 840–856.

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.

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: Ii. on the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, *69*(3), 420–436.

Claeys, W. (1990). On the spontaneity of behaviour categorization and its implications for personality measurement. *European Journal of Personality*, *4*(3), 173–186.

Cottrell, G. W., & Small, S. L. (1983). A connectionist scheme for modelling word sense disambiguation. *Cognition and Brain Theory*, *6*, 89–120.

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*(3), 466–472.

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.

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity.". *Psychological Bulletin*, *52*(3), 281–302.

Cronbach, L. J. (1958). Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrillo (Eds.), *Person perception and interpersonal behavior* (Vol. 353,

pp. 353–379). Stanford, CA: Stanford University Press.

D'Agostino, P. R., & Beegle, W. (1996). A reevaluation of the evidence for spontaneous trait inferences. *Journal of Experimental Social Psychology*, *32*(2), 153–164.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York, NY: Dover (Original work was publiched in 1885).

Erickson, D. J., & Krull, D. S. (1999). Distinguishing judgments about what from judgments about why: Effects of behavior extremity on correspondent inferences and causal attributions. *Basic and Applied Social Psychology*, *21*(1), 1–11.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, *33*(4), 465–495.

Gilbert, D. T. (1998). Ordinary personology. In S. T. F. D. T. Gilbert & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 2, pp. 89–150). New York, NY: Oxford University press.

Gilbert, D. T., Pelham, B. W., & Krull, D. S. (1988). On cognitive busyness: When person perceivers meet persons perceived. *Journal of Personality and Social Psychology*, *54*(5), 733-740.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, *27*(2), 222–248.

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*(5), 434–447.

Hamann, S. B., & Squire, L. R. (1996). Level-of-processing effects in word-completion priming: A neuropsychological study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 933–947.

Hamilton, D. L. (1988). Causal attribution viewed from an information-processing perspective. In D. Bar-Tal & A. Kruglanski (Eds.), *The social psychology of knowledge* (pp. 369–385). Cambridge, England: Cambridge University Press.

Hassin, R. R., Uleman, J. S., & Bargh, J. A. (2004). Implicit impressions. In . J. A. B. R. Hassin J. S. Uleman (Ed.), *The new unconscious.* New York, NY: Oxford University Press.

Heider, F. (1944). Social perception and phenomenal causality. *Psychological Review*, *51*, 358–373.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Higgins, E. T., & Bargh, J. A. (1987). Social cognition and social perception. *Annual Review of Psychology*, *38*(1), 369–425.

Hourihan, K. L., & MacLeod, C. M. (2007). Capturing conceptual implicit memory: The time it takes to produce an association. *Memory and Cognition*, *35*(6), 1187–1196.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541.

Johnson, J. T., Jemmott, J. B., & Pettigrew, T. F. (1984). Causal attribution and dispositional inference: Evidence of inconsistent judgments. *Journal of Experimental Social Psychology*, *20*(6), 567–585.

Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology*, *98*(1), 203–205.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. *Advances in Experimental Social Psychology*, *2*, 219–266.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Psychology*, *3*(1), 1–24.

Jones, E. E., Worchel, S., Goethals, G. R., & Grumet, J. F. (1971). Prior expectancy and behavioral extremity as determinants of attitude attribution. *Journal of Experimental Social Psychology*, *7*(1), 59–80.

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

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

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15). Lincoln.

Kelley, H. H. (1971). Attributions in social interaction. In E. Jones, D. E. Kanouse, H. H. Kelley,

R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the cause of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.

Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology*, *31*(1), 457–501.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*(2), 163–182.

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes*, *16*(1-2), 193–202.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Pressniversity press.

Krull, D. S. (1993). Does the grist change the mill? the effect of the perceiver's inferential goal on the process of social inference. *Personality and Social Psychology Bulletin*, *19*(3), 340–348.

Krull, D. S. (2001). On partitioning the fundamental attribution error: Dispositionalism and the correspondence bias. In *Cognitive social psychology: The princeton symposium on the legacy and future of social cognition* (pp. 211–227).

Lewicki, P. (1985). Nonconscious biasing effects of single instances on subsequent judgments. *Journal of Personality and Social Psychology*, *48*(3), 563–574.

Lewicki, P. (1986). Processing information about covariations that cannot be articulated. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 135–146.

Lieberman, M. D., Jarcho, J. M., & Obayashi, J. (2005). Attributional inference across cultures: Similar automatic attributions and different controlled corrections. *Personality and Social Psychology Bulletin*, *31*(7), 889–901.

Lucas, M. M., Tanenhaus, M. K., & Carlson, G. N. (1990). Levels of representation in the interpretation of anaphoric reference and instrument inference. *Memory and Cognition*, *18*(6), 611–631.

Lupfer, M. B., Clark, L. F., & Hutcherson, H. W. (1990). Impact of context on spontaneous trait and situational attributions. *Journal of Personality and Social Psychology*, *58*(2), 1239–1249.

Manis, M., Cornell, S. D., & Moore, J. C. (1974). Transmission of attitude relevant information

through a communication chain. *Journal of Personality and Social Psychology*, *30*(1), 81–94.

McCarthy, R. J., & Skowronski, J. J. (2011a). The interplay of controlled and automatic processing in the expression of spontaneously inferred traits: A pdp analysis. *Journal of personality and social psychology*, *100*(2), 229–240.

McCarthy, R. J., & Skowronski, J. J. (2011b). What will phil do next?: Spontaneously inferred traits influence predictions of behavior. *Journal of Experimental Social Psychology*, *47*(2), 321–332.

McCulloch, K. C., Ferguson, M. J., Kawada, C. C., & Bargh, J. A. (2008). Taking a closer look: On the operation of nonconscious impression formation. *Journal of Experimental Social Psychology*, *44*(3), 614–623.

McKoon, G., Gerrig, R. J., & Greene, S. B. (1996). Pronoun resolution without pronouns: some consequences of memory-based text processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 919–932.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, memory, and cognition*, *12*(1), 82–91.

McKoon, G., & Ratcliff, R. (1989a). Assessing the occurrence of eleborative inference with recognition: Compatibility checking vs compound cue theory. *Journal of Memory and Language*, *28*(5), 547–563.

McKoon, G., & Ratcliff, R. (1989b). Inferences about contextually defined categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1134–1146.

McKoon, G., & Ratcliff, R. (1989c). Semantic associations and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 326–338.

McKoon, G., & Ratcliff, R. (1990a). Dimensions of inference. *Psychology of Learning and Motivation*, *25*, 313–328.

McKoon, G., & Ratcliff, R. (1990b). Textual inferences: Models and measures. *Comprehension processes in reading*, 403–421.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, *99*(3), 440–466.

McKoon, G., & Ratcliff, R. (1995). The minimalist hypothesis: Directions for research. In

C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of walter kintsch* (pp. 97–116). Hillsdale, NJ: Erlbaum.

Moskowitz, G. B. (1993). Person organization with a memory set: are spontaneous trait inferences personality characterizations or behaviour labels? *European Journal of Personality*, *7*(3), 195–208.

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1003–1019.

Newman, L. S. (1991). Why are traits inferred spontaneously? a developmental approach. *Social cognition*, *9*(3), 221–253.

Newman, L. S. (1993). How individualists interpret behavior: Idiocentrism and spontaneous trait inference. *Social Cognition*, *11*(2), 243–269.

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition*, *33*(1), 20-66.

Park, B. (1989). Trait attributes as on-line organizers in person impressions. In J. N. Bassili (Ed.), *On-line cognition in person perception* (pp. 39–59). Hillsdale, NJ: Erlbaum.

Posner, M. I. (1994). Attention: the mechanisms of consciousness. *Proceedings of the National Academy of Sciences*, *91*(16), 7398–7403.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, *27*(4), 399–415.

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology*, *42*(4), 593–607.

Ramos, T. M. (2009). *A flexible view of spontaneous trait inference* (Unpublished doctoral dissertation). ISCTE, Instituto Universitário de Lisboa, Lisboa, Portugal.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review*, *95*(3), 385–408.

Ratcliff, R., & McKoon, G. (2008). Passive parallel automatic minimalist processing. In C. En-

gel & W. Singer (Eds.), *Better than conscious: decision making, the human mind, and implications for institutions* (pp. 176–189). Cambridge, MA: MIT Press.

Rim, S., Min, K. E., Uleman, J. S., Chartrand, T. L., & Carlston, D. E. (2013). Seeing others through rose-colored glasses: An affiliation goal and positivity bias in implicit trait impressions. *Journal of Experimental Social Psychology*, *49*(6), 1204–1209.

Roediger, H. L., & Blaxton, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory and Cognition*, *15*(5), 379–388.

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 8, pp. 63–131).

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of endel tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.

Roediger, H. L., Weldon, M. S., Stadler, M. L., & Riegler, G. L. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1251–1269.

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.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592–604.

Smith, E. R., & Miller, F. D. (1983). Mediation among attributional inferences and comprehension processes: Initial findings and a general method. *Journal of Personality and Social Psychology*, *44*(3), 492–505.

Snyder, M., & Jones, E. E. (1974). Attitude attribution when behavior is constrained. *Journal of Experimental Social Psychology*, *10*(6), 585–600.

Tagiuri, R. (1958). Social prefereces and its perception. In R.Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior* (pp. 316–336). Stanford, CA: Stanford University Press.

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of Experimental Psychology*, *86*(2), 255–262.

Till, R. E., Mross, E. F., & Kintsch, W. (1988). Time course of priming for associate and inference words in a discourse context. *Memory and Cognition*, *16*(4), 283–298.

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.

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.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.

Trope, Y. (1986). Identification and inferential processes in dispositional attribution. *Psychological Review*, *93*(3), 239–257.

Tulving, E. (1972). Episodic and semantic memory 1. *Organization of Memory. London: Academic*, *381*(4), 382–404.

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141–160). New York, NY: Guilford.

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*(4), 377–394.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, *66*(3), 490–501.

Uleman, J. S., Newman, L., & Winter, L. (1992). Can personality traits be inferred automatically? spontaneous inferences require cognitive capacity at encoding. *Consciousness and Cognition*, *1*(1), 77–90.

Uleman, J. S., Winborne, W. C., Winter, L., & Shechter, D. (1986). Personality differences in spontaneous personality inferences at encoding. *Journal of Personality and Social Psychology*, *51*(2), 396–403.

Van Overwalle, F., Drenth, T., & Marsman, G. (1999). Spontaneous trait inferences: Are they

linked to the actor or to the action? *Personality and Social Psychology Bulletin*, *25*(4), 450–462.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition and Emotion*, *19*(2), 175–196.

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.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 385–399.

Whitney, P., Davis, P. A., & Waring, D. A. (1994). Task effects on trait inference: Distinguishing categorization from characterization. *Social Cognition*, *12*(1), 19–35.

Whitney, P., Waring, D. A., & Zingmark, B. (1992). Task effects on the spontaneous activation of trait concepts. *Social Cognition*, *10*(4), 377–396.

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

Wigboldus, D. H., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484.

Wigboldus, D. H., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition*, *22*(3), 292–309.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252.

Winter, L., Uleman, J. S., & Cunniff, C. (1985). How automatic are social judgments? *Jouranl of Personality and Social Psychology*, *49*(4), 904-917.

Wyer, R. S., & Srull, T. K. (1986). Human cognition in its social context. *Psychological Review*, *93*(3), 322–359.

Zárate, M. A., Uleman, J. S., & Voils, C. I. (2001). Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition*, *19*(3), 295-323.

2

A CONNECTIONIST MODEL OF SPONTANEOUS TRAIT INFERENCE AND SPONTANEOUS TRAIT TRANSFERENCE: DO THEY HAVE THE SAME UNDERLYING PROCESSES?

Diana Orghian, Leonel Garcia-Marques, Jim Uleman and Dietmar Heinke

Social Cognition, Vol. 33, No. 1, 2015, pp. 20-66

This Chapter can be regarded as a continuation of the Introduction presented in Chapter 1. The Introduction ends with the discussion regarding the processes underlying STI and STT and, in particular, regarding the differences between the two. We used an auto-associative model to simulate the two phenomena and four of the main empirical differences usually found in the literature. This computational model is a simple demonstration of how, based on associations between behaviors, traits and persons, the main differences between STI and STT can be replicated. In other words, they are being replicated without the need of considering an additional process beyond the associative one. We do however, make an assumption to motivate the difference between STI and STT. We assumed that some stimuli are more salient than other, and as such are paid more attention to. We suggest that the actor of the behavior is a more salient stimulus, and due to top-down differential distribution of attention, it will get more activation. A result of this extra activation is a stronger link that is created between the representation of the actor and the trait activated via the comprehension of the behavior.

2.1 INTRODUCTION

Inferring traits and characteristics about others' personalities is a natural way of knowing each other; a way of organizing the complexity of the social world that allows us to predict others' behaviors and achieve cognitive control over one's environment. In everyday life, we sometimes intentionally form impressions about others' personalities from their behaviors, *e.g.*, in a job interview. But predominantly, impressions occur without any intention and awareness of making such inferences, revealing the remarkable efficiency of this ability (Todorov & Uleman, 2002, 2003, 2004). By definition (Uleman, Newman, & Moskowitz, 1996), a spontaneous trait inference (STI) occurs when a personality trait of an actor (*e.g.*, "honest") is inferred from his/her behavior (*e.g.*, "Johnny told the cashier that he received too much change.") without an explicit intention to form an impression, or to infer a personality trait about the actor. Trait inference seems to be a natural and inherent process in the comprehension of the behavior itself (Winter & Uleman, 1984). But as in all the human cognitive processes, errors can occur. One specific example is when people misattribute the inferred traits to the wrong person, a person who does not enact the behavior in the description but who tells about it. Such errors are called spontaneous trait transference (STT; Carlston, Skowronski, & Sparks, 1995; Skowronski, Carlston, Mae, & Crawford, 1998).

In our social life, besides communication about the self, we are frequently presented with communications in which informants talk about others (*e.g.*, a reporter describing a crime; someone gossiping about a third party). Now imagine that at the first lunchtime with your new colleague Mary, she mentions that a friend of hers, Adam, "never votes in elections". From this description of Adam, you can infer that he is an "apathetic" person. But surprisingly, you may also get a sense that your colleague Mary is "apathetic" too and not even realize that you've formed this impression. STT can occur toward individuals who are merely describing someone else's behavior or who are associated with that behavior due to spatial-temporal contiguity. The fact that the inferred trait can also be transferred to an inanimate object (the case of superstitious banana; Brown & Bassili, 2002) supports the idea that STTs are caused by incidental processes. STIs and STTs are detected with a variety of memory and reaction time measures (Skowronski

et al., 1998; Todorov & Uleman, 2002; Uleman et al., 1996) and are generally independent of awareness of making them.

The STT effect is *per se* interesting because it can have practical and observable consequences in our lives. If an informant comments about something someone else did (gossiping situation), the impression formed about the informant's personality can be affected by the traits implied in behavior he is describing. A second example is reporters who frequently describe particular kinds of events (heroic acts, criminal/aggressive acts, etc.). Also in court cases, the testimony of witnesses may imply negative traits. These traits may rub off on the witnesses and may affect the judge's or jury's impression of the witnesses, consequently influencing the outcome of the trial. The same transference may occur when someone is accidentally seen in a particular scene (*e.g.*, someone passing by a fight on the street is later associated with the violent trait). On a general note, the STT can be seen as a special example of inferences based on second-hand information (*e.g.*, witness statement) and STI as inferences based on first-hand information (*e.g.*, character of the witness). We talk about personality trait transferences in the present paper, but STT may extend to transference of goals, of motivations, of needs or of other inferences made from the informant's message (*e.g.*, Hassin, Aarts, & Ferguson, 2005).

Besides the importance of STT on its own, the discovery of STT also influenced the investigations of STI. For instance, it has focused research on the nature of the link between the person and the trait. It has also raised the question of how STI and STT differ in terms of the underlying mechanisms that produce them. Is the same process responsible for both STI and STT or are there two distinct cognitive processes behind the two phenomena? Of course, typically this question is addressed in an empirical fashion. In contrast, the present paper addresses it in a theoretical fashion, *i.e.*, proposing a connectionist model of the underlying processes. To be more precise, we will propose a model of the experiments typically conducted to address this question. This way we are able to access the validity of the theoretical claim they make. Moreover, and importantly, we will use this model to propose new empirical studies to advance this debate.

The question regarding the processes responsible for STI and STT has been actively debated in the literature, and two views can be clearly distinguished. According to some authors, STI and STT reflect two distinct cognitive processes (*e.g.*, Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007). STI requires an attri-

butional process, whereas STT is based solely on simple associative links. By contrast, Bassili (Bassili, 1976; Bassili & Smith, 1986; Brown & Bassili, 2002) advanced a single process view in which both STI and STT are based on the same simple automatic associative processes. As discussed below, there are clear differences between STI and STT, and these differences have been used to argue that there are two processes. However, an associative account, omitting the attributional one, has not been rule out.

In fact, neither of these "processes" is well specified in this literature. "Associative processes" that link one concept (a person) to another (a trait) with a single bi-directional link seem parsimonious. All this requires is the co-occurrence of a person representation (*e.g.*, a photo) and a trait (inferred from behavior). But this provides no account of how traits are inferred from behaviors in the first place, and ignores the evidence that these inference processes do not result from simple associative links. Inferring people's traits from behavior involves not one but three links: person-trait, behavior-trait, and person-behavior. It is known that the behavior-trait link is not symmetric; people more readily infer traits from behaviors than behaviors from traits (Maass, Colombo, Colombo, & Sherman, 2001). There is also recent evidence that the semantic links between traits and behaviors, as isolated concepts, are not merely associative but causal (Kressel, 2011; Kressel & Uleman, 2010). Much is also known about links between persons and traits, and persons and behaviors, often under the heading of stereotypes (*e.g.*, Schneider, 2005). A person's social category affects the STIs formed from their behaviors (Wigboldus, Dijksterhuis, & Van Knippenberg, 2003). Taken together, treating STT as the result of simple associations does not account for the potentially different nature and roles of actor-behavior or trait-behavior links, or possible interactions among these different types of links. The attributional account of STI is also imperfect and misleading. To begin with, the term "attribution" is an ambiguous one. Attribution can be interpreted as giving explanations or causes to behavior (and these explanations can be related with personality traits or not) or it can be interpreted as making a dispositional inference (that can be explanatory but usually is not) from behaviors (Malle, 2007).

Regardless the clarity of the attribution process, STI actually has shown some of the features of intentional attributions as described by classic attribution theories (Heider, 1958; Jones & Davis, 1965; Kelley, 1967). STIs are sensitive to behaviors' valences, *i.e.*, the attribution made from negative behaviors, which are relatively uncommon and thus diagnostic, tend to be stronger than

attribution made from positive behavior. The possibility of generalization to other traits (halo effect) is another example of how comparable STI and attribution theories can be. However, the catalogue of possible similarities among STI and STT is largely unexplored.

Another important point is STI's unintentional and largely unconscious nature which contrasts with attributions occurring consciously in response to unexpected behaviors (Clary & Tesser, 1983; Hastie, 1984; Kanazawa, 1992; Lau & Russell, 1980; Pyszczynski & Greenberg, 1981; Wong & Weiner, 1981), subjective loss of control (Pittman & Pittman, 1980; Swann, Stephenson, & Pittman, 1981), personal relevance (Berscheid, Graziano, Monson, & Dermer, 1976; Harvey, Town, & Yarkin, 1981), and failure (Diener & Dweck, 1980; Wong & Weiner, 1981). The intentionality argument is not very strong, though, because it has been recently shown that causal attribution can occur in a spontaneous way (for an example see Hassin, Bargh, & Uleman, 2002). But there is another relevant evidence that has to be attended regarding the causality of traits; Kressel and Uleman's (2010) work supports the view that attribution process is not even necessary for traits to function as causes, since traits are considered causes even in isolation.

Thus, both associative and attributional processes are very incomplete in terms of their explanation of the STI and STT phenomena since they only describe the trait-person link.

Nevertheless, the dichotomy between simple associations and attributions has served as a placeholder for explaining the differences between STI and STT. And the dominant view has been the dualistic view which says that STTs are based on simple associations between persons and traits (once traits are inferred by other processes), whereas STIs reflect different, more complex and deeper (although still unintended and largely unconscious) processes, and establish properties of persons rather than mere associations with them.

Therefore, we would like to initiate a deeper discussion of this particular dichotomy between these two processes, but also a broader debate on single versus dual processing dichotomies, which is a recurrent theme in psychological science. For example, the long-standing tradition of dual-process explanations includes explicit versus implicit memory (Schacter, 1987), amodal versus modal representations (Fodor, 1975; Barsalou, 1999, repectively); direct versus indirect routes to action selection (Yoon, Heinke, & Humphreys, 2002), etc. However, there is also a long-standing tradition of using computational models to demonstrate that these dichotomies are not necessarily true. For instance, MINERVA was proposed as a demonstration that a memory

model based on exemplars could account for prototype effects without the need to postulate abstract representations (Hintzman, 1986; Hintzman & Ludlam, 1980). Josewitz, Staddon, and Cerruti (2009) presented a simple associative model (BEM) that does not include metacognitive processes but that can simulate animal behavior previously taken to be diagnostic of metacognition. Seidenberg and McClelland's (1989) connectionist model demonstrated that the pronunciation of words that follow regular pronunciation rules (regular words) and words that don't follow regular rules (irregular words) can be generated with a single process.

So the present paper sees itself in the tradition of these demonstrations. We present a connectionist approach to STI and STT, using a connectionist model named MATIT – Model of Associative Trait Inference and Trait Transference. The aim of this model is to simulate the four main empirical differences between STI and STT with one simple autoassociative connectionist network, based on a single underlying associative process. If our simulations are successful, they suggest that one "process" can produce the main differences found between these two phenomena. Then, the current body of empirical data is not sufficient to support the existence of two different processes.

The application of a connectionist framework to social cognition is not new. There are connectionist models for causal attribution (Read & Montoya, 1999; Van Overwalle, 1998), cognitive dissonance (Shultz & Lepper, 1996; Van Overwalle & Jordens, 2002) and impression formation (Kashima, Woolcock, & Kashima, 2000; Van Overwalle & Labiouse, 2004). In particular, Van Overwalle and Labiouse (2004) used a autoassociative network to investigate phenomena involving primacy and recency in impression formation, the asymmetric impact of ability versus morality behaviors, memory advantages for inconsistencies, assimilation and contrast in priming, and the effect of situational constrains on trait inference. However, it is important to note that computational models are often not falsifiable as they are able to fit any data, even contradictory data (Roberts & Pashler, 2000). A classical example of this problem was highlighted by Wexler's paper (1978) on Anderson's ACT theory (1976) . Wexler (1978) showed that ACT could not only model what it is meant to model (the Sternberg result), but also its opposite. Thus, we follow Roberts and Pashler's recommendation (2000) and, after presenting our model's abilities to mimic existing evidence we will present two predictions from the model and discuss

how particular experiments would have the potential to contradict these predictions, *i.e.*, falsify MATIT.

The present work used a model inspired by Van Overwalle and Labiouse (2004). We believe that our simulation model is cognitively plausible and that it accounts for the most important empirical differences found in the literature between STI and STT. More importantly, by using a relatively simple model to reproduce both STI and STT phenomena, we hope to demonstrate that it is not necessary to postulate two separate processes to account for these phenomena and the differences between them. Rather, their empirical differences may result from the same process, and also from differences in the deployment of attention within the experimental paradigms (an idea developed later on in the paper).

To forestall misunderstandings, below we list all we do and do not mean to show with the implementation of MATIT and this paper in general. We do not intend to a) present a model that describes STI and STT phenomena in their intrinsic complexity; b) explain all the differences between STI and STT; or c) defend a single process view. And what we do intend to show in the paper is that the evidence used to suggest the existence of two processes is easily reproduced by a simple and purely associative model. It is crucial to note that the MATIT model is indeed a very simple model and that, therefore, it can easily go wrong and be disproved. Our point is not that we are able to come up with an associative model that can explain previous results. After all, given some theoretical latitude and/or adhocery, any type of model can simulate (mimic) any pattern of data (J. R. Anderson, 1978; Garcia-Marques & Ferreira, 2011). In that sense, finding a simulation model that simulates a data pattern is like fitting a statistical model. It will be a meaningless achievement unless the model can be falsified by plausible data (*e.g.*, Roberts & Pashler, 2000). As we will demonstrate with MATIT, the advantage of using a simple (baseline) associative model is that even when it fits the data, it can provide clear guidelines for obtaining data that will challenge the model, that is, more diagnostic data. Such a data pattern would be diagnostic in indicating what a critical experimental design would be, that would adequately test the single versus dual process views.

This article is organized as follows: 1) a description of the problem; 2) an overview of MATIT model, qualitatively describing the architecture and how it processes and learns information (the

mathematical details can be found in the section 2.14.1); 3) seven simulations; and 4) a general discussion and conclusions.

2.2 ASSOCIATIVE VERSUS ATTRIBUTIONAL PROCESSES

The two-process view (Carlston & Skowronski, 2005) suggests that both associative and attributional processes may come into play during the spontaneous encoding of the behavioral descriptions and the actors. Attributional processes are elaborative processes activated during the encoding of behaviors and of their actors. They involve deeper mental activity that implicates attributional (causal) knowledge and logic. They produce labelled associations between traits and persons that incorporate retrievable tags that define traits as properties of people (Johnny is honest) or as causes of their behaviors (Johnny returned the wallet with all its money because he is honest; Kressel & Uleman, 2010). Attributional processes are described by classical attributional theories (Hastie, 1984; Jones & Davis, 1965; Kelley, 1967). But something different is said to occur when behaviors are presented with persons who are not the actors. Then, a simple associative process occurs.

The associative process is characterized as relatively shallow and results in generic unlabeled linkages (Carlston & Smith, 1996). It is a consequence of the spatial and temporal contiguity of activated constructs (trait and person). It is insensitive to the information's diagnosticity. Contrary to the attributional processes, where "is a property of" or "is a cause of" or "is an impediment to" links occur, associative links are unlabeled (Johnny is associated with the concept "honest", Carlston & Smith, 1996). Furthermore, the linkages in memory are weaker because they are established through a process that involves little elaboration.

In the single process view (Bassili, 1976; Bassili & Smith, 1986; Brown & Bassili, 2002) STI, like STT, may result from automatic associative links to traits activated during the encoding of behaviors and other stimuli such actors, communicators, bystanders or even inanimate objects that happen to be part of the context at the moment. Note that regardless of the debate about the existence/coexistence of these two processes, it is assumed that they happen during the encoding of the behavior and of the person stimulus, not during the retrieval of this information. The single process view attempts to explain differences between STI and STT through different associative

strengths between the trait and the person. What could be responsible for the different levels of associative strengths in STI and STT? We propose that these differences in associative strength result from different activations of the representations of the presented stimuli (behavior, actor and the communicator).

Note that we use this term activation of representation in the specific sense of the connectionist framework of our model. In this framework, representations are loosely related to neural activities in the brain. There can be many reasons for a representation to become more activated. These include conceptual relevance (*e.g.*, more relevant for the task, see Chapter 3), or activation just because there is more available attention in one case than in others. In fact, MATIT's implementation of differential activations is based on the presumed operations of internal attention (*e.g.*, Chun, Golomb, & Turk-Browne, 2011). Chun and collaborators (2011) distinguish two types of attention, external and internal. External attention deals with perceptual information whereas internal attention operates on internal information such as the contents of working memory, task sets, etc. In MATIT we assume that all the relevant information, *e.g.*, behavioral descriptions, actors, and bystanders, is internally represented, *i.e.*, in working memory, and that we pay more (internal) attention to the actor in the STI condition than to the communicator in the STT condition which, in turn, leads to varying levels of associative strength (for a review of evidence on the link between attention and memory see Naveh-Benjamin, Craik, Perretta, & Tonev, 2000; Craik & Lockhart, 1972). These different levels of associative strength will result in the differences between STI and STT. Note that the influence of internal attention is not a crucial assumption in MATIT.

We chose attention because there are numerous computational models postulating that attention can be responsible for the different levels of activation (*e.g.*, Heinke & Backhaus, 2011; Mavritsaki, Heinke, Allen, Deco, & Humphreys, 2011; Bundesen, Habekost, & Kyllingsbæk, 2005). Also evidence from electrophysiological studies (for one of the first findings of this type see Desimone & Duncan, 1995) and neuroimaging studies (for an example of external and internal attention see Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Lepsien & Nobre, 2006, respectively) point in the same direction. Moreover, the evidence on the relationship between attention and memory as pointed out earlier is very strong (Naveh-Benjamin et al., 2000; Craik & Lockhart, 1972). Furthermore as we discuss below, current evidence on STI and STT

Overview of the Simulations	
	Simulation 1
Finding	STI effect stronger than STT
Method	Higher initial input value actor node than for the irrelevant person node.
	Simulation 2
Finding	Simultaneous presentation of the actor and the irrelevant person reduces STT effect and had no effect on STI.
Method	Increasing the difference between the values for the actor and the irrelevant person nodes.
	Simulation 3
Finding	Lie-detection instruction reduces the STI effect at the level of the STT effect.
Method	Input for behavior node is reduced and input for relevant and irrelevant person node are increased to similar values.
	Simulation 4
Finding	Trait generalization is more likely in the STI than in the STT.
Method	The node representing the critical trait is paired with a consistent trait
	Simulation 5
Demonstration	Replication of Simulation 1 with a wider range of parameters.
	Simulation 6
Dissociation 1	Photo repetition will increase STI and decrease STT, in case dual process view is correct
Result	STI and STT where equally affected by the simulation.
	Simulation 7
Dissociation 2	Presenting each trial twice in the task-specific phase will increase STI and decrease STT effect if dual process is correct.
Result	The manipulation affected STI and STT in the same manner.

Table 1: Overview of the simulations

does not rule out this possibility. Finally, it is generally agreed that attention is a ubiquitous process. Hence, it would be very surprising if attention were not involved in STT/STI.

2.3 EMPIRICAL DIFFERENCES BETWEEN STI AND STT

The idea that STI could be the result of attributional processes is inspired by similarities between characteristics of STI and attributions as described in classical theories (Heider, 1958; Jones & Davis, 1965; Kelley et al., 1972). An example is the well-known negativity effect. In this, because negative behaviors are more uncommon and non-normative, they are more diagnostic and informative than positive behaviors. This makes attributions from negative behaviors stronger (more likely, extreme, and confident; Reeder & Brewer, 1979). Indeed, Carlston and Skowronski (2005) demonstrated that this negativity effects exists for STI but not STT, so they concluded that STI results from attributional knowledge.

Skowronski an collleagues (1998) suggested STT can be described as a pure associative process that results from a series of three steps (see also Mae, Carlston, & Skowronski, 1999). In the

first step, the trait (*e.g.*, "helpful") is activated during behavior comprehension (*e.g.*, "Ben carried the old lady's groceries across the street"). In the second step, the inferred trait is associated with the presented person. Finally, the association made in the previous step implicitly influences the impression of the person with whom the trait was associated. Hence, this process does not reflect trait judgement or attribution, but is merely a result of the simultaneous activation of trait and person, *i.e.*, an associative process. On the other hand, STI is the result of a more elaborate attributional process that accounts for its differences from STT.

Both STI and STT are difficult to control since they occur even when the perceivers are warned of the effects and are told to avoid them (Carlston & Skowronski, 2005) or under cognitive load (Crawford, Skowronski, & Stiff, 2007). We also know that they are equally dependent on concurrent tasks, and working memory (Wells, Skowronski, Crawford, Scherer, & Carlston, 2011).

There are, though, important empirical differences between STI and STT that compelled some investigators to suppose that they involve different processes. The first of these differences is in the magnitude of the effect that is usually greater for STI than STT (*e.g.*, Bassili & Smith, 1986; Goren & Todorov, 2009; Skowronski et al., 1998). For instance, Skowronski and colleagues (1998, Experiment 2) obtained trait ratings of targets two days after participants merely familiarized themselves with targets' photos and descriptions of their own behaviors, or behavioral descriptions of actors (not pictured and opposite-sex) provided by acquaintances (represented in the picture). Effects for STIs ($d = .74$) were about twice those for STTs ($d = .35$). In this paper, we will show that our connectionist model can easily mimic this and others differences between STT and STI through variations in the associative links between faces and traits.

Second, several studies found that if a photo of the actor is presented next to the informant during the encoding of the behavior, the transference effect is reduced or even eliminated (Crawford, Skowronski, & Stiff, 2007; Goren & Todorov, 2009; Todorov & Uleman, 2004).

Third, a concurrent inferential task, such asking participants to detect whether the presented person is lying about the behavioral description (about their own behavior in the case of the actor, or about the other's behavior in the case of the informant), seems to reduce STI's magnitude whereas it has no effect on STT (Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski

et al., 1998). The authors of these studies believe that the lie-detecting task interferes with the attributional process and not with the associative one.

Finally, several studies (Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski et al., 1998) have shown that trait generalization or halo effects are more likely for the actor than for the non-actor. It is said that this happens because in STI, the links between persons and traits are inferential/ attributional and, in accordance with attribution and implicit personality theories, allow generalization from one trait ("smart", an intellectual trait with a positive valence) to other traits ("friendly", a social trait with the same positive valence as "smart").

Table 1 lists the experiments and the corresponding topics each simulation intends to replicate. This set of simulations contrasting STI and STT does not exhaustively include all published studies, but only the most important ones.

Our simulations with MATIT focus on the false recognition paradigm (Todorov & Uleman, 2002, 2003, 2004; Goren & Todorov, 2009) as one of the most common and recent paradigms used to study STI and STT effects. The paradigm consists of two phases. In the study phase the participants are presented with photos of faces along with one of two types of sentences. The first type of sentence includes a trait and behavior, *e.g.*, "Mary is so "helpful" that she carried an elderly lady's groceries across the street". The second type includes only behavior, *e.g.*, "Mary carried an elderly lady's groceries across the street", so that the trait "helpful" is only implied in the sentence. Participants are asked to memorize the pairs of stimuli (photo and sentence).

Subsequently in the test phase, they see a series of face-trait pairs and have to indicate whether the word (trait) previously appeared in the sentence paired with that particular person. For the second type of sentence, "yes" responses constitute false recognitions and indicate spontaneous trait inferences at encoding, because the trait wasn't actually presented but only implied. Participants show more false recognition of traits that were originally implied in the learning phase and then tested with that same face (on "matched" trials – old pairing) than traits presented with faces originally presented with other behaviors (on "mismatched" trials – new pairing). In STI conditions, the person in the photo is said to be the actor of the behavior. In the STT conditions, the person is said to be an informant or communicator, or a bystander, or a photo randomly paired with that sentence.

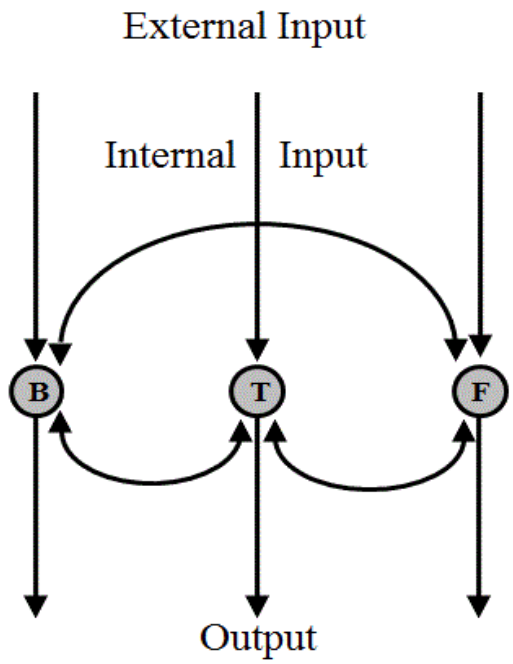

Figure 1: Illustration of an auto-associative recurrent network: The nodes represent the stimuli: B – behavior, T – trait, F – face. The lines represent the connections between these nodes. Each node receives internal input from other nodes which are summed to produce the internal input and also exterior activation.

Each simulation sought to replicate the result of a specific experiment that we describe before describing its implementation in the model. We used the same general model throughout, and kept the simulations as close as possible to the experimental design of the actual studies. But we made some minor simplifications to facilitate modeling and facilitate understanding of the model (*e.g.*, fewer trials and fewer stimuli). The same experimental paradigm was modeled in all the simulations to keep experiments and simulations comparable.

2.4 THE MODEL OF ASSOCIATIVE TRAIT INFERENCE AND TRAIT TRANSFERENCE (MATIT)

The model consists of two parts: the effect of internal attention on the representation of information and the autoassociative network. We will first describe the network and then describe how we model the internal attention. Our network was inspired by the recurrent auto-associative

network that was proposed by McClelland and Rumelhart (1985). Though, we used a simplified version of the network as implemented by Van Overwalle and Labiouse (2004) in their work about person impression formation. As shown in the Figure 1, one of the most important features of the network is that all the nodes in the system are interconnected. The network has two operation modes, a phase where the activation of the nodes is computed, and a second phase where the weights of the connections are updated. In the first phase, the model receives an external input that typically comes from the environment. Because the nodes are interconnected, the activation received from external sources spreads throughout the network. Besides the external input, a node also receives activation from other nodes in the network. A memory trace is created as a consequence of weight changes that are driven by the error between the internal activation generated by the network and the external input received from outside sources. The error-reduction mechanism is based on delta learning algorithm (McClelland & Rumelhart, 1989) that has the function of adjusting the weights of the connections between nodes. When a node receives too much input from other nodes, this means the network is overestimating the external input of that node and the way delta rule acts in this situations is by decreasing the weights of the connections between that node and the other nodes. In case the network underestimates the external input, algorithm's role is to increase the weights of the connection to better approximate the internal to the external input. The error decreases in proportion to a learning rate parameter, which determines how fast the network learns and corrects the discrepancies between the two kind of inputs. After several external input series, the activation in the network becomes better and better in simulating/predicting the external input, and at some point it settles into a stable pattern of activations. For mathematical details see section 2.14.1.

So, the main goal of the network and the delta algorithm is to adjust the input activation to converge on the activation received from the environment, by minimizing the difference (the error). The connection weights are initially set to zero (or random small values). Thus at the beginning of learning, these weights are small and inefficient in predicting the external input. But gradually the accuracy of the network and its ability to represent the external activation pattern increases as more external information is provided and "learned".

Each node in the network represents a construct with psychological meaning. This type of encoding is called localist, as opposed to distributed encoding (Smith & DeCoster, 1999) in which

each concept is represented by a pattern of activation across a group of nodes. Distributed encoding more plausibly represents the organizations and the functioning of neurons in the brain. But localist encoding is useful when a simple demonstration is preferred. There is also some evidence (*e.g.*, Van Overwalle & Labiouse, 2004) that approximately the same results that localist encoding provides can be obtained with distributed representations.

The localist representation in MATIT takes on the following form. Imagine that we present a sentence describing some behavior and two faces (the actor of the behavior and the communicator of the sentence) in the same trial to a participant, for memorization. In computational terms, a way of presenting this specific trial to our algorithm is to present the following input: 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 - where the first 8 digits (1 or 0) represent 8 possible behaviors with one of them activated. The next 8 digits represent actors, and the last 8 digits represent the 8 possible bystanders. Digit one means that that specific node (and the concept it represents) is activated and zero means no-activation is received from the external input for that node. So, the input is a 1×24 matrix of zeros and ones in which the first 8 columns refers to behaviors, the second 8 to actors, and the third set of 8 columns to bystanders. To operationalize the influence of internal attention, these values (activations) are modified by taking into account two different characteristics of attention. First the item's activation should be proportional to the amount of attention that was paid to it. For instance, if an actor is strongly attended, the actor node is highly activated. This operationalization of attentional modulation is inspired by findings in electrophysiological studies (*e.g.*, Desimone & Duncan, 1995) and neuroimaging studies (Lepsien & Nobre, 2006), where the attentional state of observers was found to increase the level of activation.

Second, to mimic capacity limitations we assume that the sum of the activations of the representations is constant at one. Thus, the pattern presented earlier is normalized turning it into 0.5 0 0 0 0 0 0 0 0.25 0 0 0 0 0 0 0 0.25 0 0 0 0 0 0 0. Moreover, if the activation of the actor node is increased (from 0.5 to 0.8, *i.e.*, 0.3 more attention) the activation of other nodes has to be decreased by the same amount (0.3, that is less attention), the final result being: 0.8 0 0 0 0 0 0 0 0.1 0 0 0 0 0 0 0 0.1 0 0 0 0 0 0 0. Similar realizations of attention are found in many computational models of attention (*e.g.*, Heinke & Backhaus, 2011; Mavritsaki et al., 2011; Bundesen et al., 2005). Finally, it is also important to note that as a result of this imple-

mentation, attention also affects the strength of the weights in associative memory. The attention in this case defines how strong is the external input that the model receives. And because on the basis of the weight update algorithm between two nodes is the difference between the external (that depends closely on attention in our case) and the internal input (that depends on the activation that comes from other nodes in the system), the associations in memory will be always affected by the attention (for more details see section 2.14.1). This is consistent with behavioral evidence on the links between attention and memory (Craik & Lockhart, 1972; Naveh-Benjamin et al., 2000), further supporting this implementation of attention.

2.5 COMPUTATIONAL STUDY

For each of the simulated differences between STI and STT, there is more than one behavioral study using various methods, but we only describe and implement one of them here, the false recognition paradigm. This makes the implementation of the MATIT model easier to follow and makes the simulated effects more concrete. Also note that at this point a reader who is not interested in the details of the simulation can skip the following sections and fast forward to the section ”Summary of the simulations” just before the General Discussion.

2.5.1 *General Method of the Simulations*

To implement the false recognition paradigm in the model, the three types of stimuli (faces, behaviors and traits) were represented by three nodes (see Figure 1). The learning was based on the acquisition of patterns of weights (associations) among these nodes (see Table 2 for an example). Note that we did not model responses to trials that contained explicit traits in the behaviors. These trials were necessary for human participants so that they don’t adopt a strategy of simply responding “no” on every trial. Because the model cannot adopt such a strategy (or any strategy, for that matter), these trials were not included in the simulations. Thus, behavior nodes represent sentences that contained no explicit traits.

Each simulation consisted of two learning phases and one test phase. In the learning phases, the model was trained for two different kinds of knowledge, world knowledge (see Table 2) and task-specific knowledge (see Table 3). The world knowledge mimics the fact that spontaneous personality trait inferences rely on general knowledge about people's characteristics and their behaviors (*e.g.* the behavior "shared his/her umbrella with a stranger during the rain" with the trait "friendly"). Learning world knowledge is not part of the false recognition paradigm but is specific to the simulation because human participants already know it, and this knowledge permits the inference of traits from behavior descriptions.

The second learning phase trained the network with the task-specific knowledge (the first step in the false-recognition paradigm), where a specific behavior (*e.g.*, "Anna shared her umbrella with a stranger during the rain") is associated with a specific person (*e.g.*, Anna's face photo). So, the learning of associations between specific traits and specific behaviors corresponds to the world knowledge, and the task-specific knowledge corresponds to the association between behaviors and specific faces.

Knowledge acquisition in the model is determined by the learning rate parameter (ϵ in the Supplementaty Material) and the given input (external input). The input was not always set at the same values in all the simulations because in the original experiments, the instructions, the manipulated variables, and the presumed attention differed. However, the learning parameters were chosen in the first simulation and then the same values were used in all following simulations.

In order to "teach" the model the world knowledge, we had it learn the associations among a series of behavior-trait pairs (see Table 2 for specific values). Each row in Table 2 corresponds to a trial where the values represent the activation of each node. So in each trial, two nodes, a trait and a behavior, were activated simultaneously, which made the model learn that these two are associated. The learning of associations between specific trait-behavior pairs depends on the number of times this trial is given as input to the network (apart from the learning rate). The presentation frequency of each pair captures the way we learn in the real life. An equivalent to the world knowledge learning would be children's learning about how to categorize behavior, when for instance the child observes someone performing a certain behavior and next hears the adult categorizing the observed behavior by naming the correspondent trait (*e.g.*, "friendly"). The more often the individual is exposed to this same situation, the more he/she will think these

two are related (the trait and that type of behavior). Of course this learning process constitutes a great simplification of the way children learn in real life. But we don't intend to explore the complexity of this process in the present paper.

As noted above, the learning in MATIT also depends on the learning rate parameter that governs the speed of learning. The learning of the world knowledge is expected to be slow, since it is acquired over the long term, based on frequencies of exposure rather than explicit propositional learning (*e.g.*, Gawronski & Bodenhausen, 2006).

The second part of the learning phase (see Table 3) is specifically related to the false recognition task described above - memorize pairs of behavioral sentences and faces. The input to the model is a pattern where nodes representing behaviors and nodes representing persons are activated simultaneously. Note that in both learning phases, only one behavior is associated with each trait and only one behavior with each face, and that the behavior in these two learning parts are the same. In this phase, each trial (rows in the Table 3) was presented only once (frequency 1), as in the experimental studies being simulated (participants see each trial once). To make the simulations more realistic, we introduced some noise (equivalent to the variability in behavioral data), and added a random value ranging between 0.0 and 0.1 to the default starting weights of zero.

As shown in the Table 2 and 3, the activations of all nodes in the input pattern add up to 1. An input pattern refers to all nodes in each row, and each row represents a trial. This is a reflection of how MATIT models the operation of attention as explained earlier. We assume here that by asking participants to memorize several simultaneously presented stimuli, the attention (activation) will be divided between the elements to be processed. Due to the capacity limitation of human processing we kept the sum of the activation at one at all conditions. After learning the association between nodes, the test phase (see Table 4) was run in which we turned on some specific nodes (providing an "incomplete patterns") and observe the output (the question marks in the table). The resulting output represents the completion of the patterns, *i.e.*, the activations of other nodes that were encoded in the learning phase but are not presented in the input for test phase. Because all the nodes in the system are interconnected (differing in the weights of their connections), if we give to the model patterns with some nodes activated, it will "recall" associated nodes in the network due to the spread of activation.

For each simulation, the network was run 50 times, simulating 50 participants, and within each learning phase (world knowledge and task-specific knowledge) the trials were randomized for each participant. See section 2.14.1 for a walk-through example.

2.6 SIMULATION 1 – STT VERSUS STI

The first simulation models the first study from Goren and Todorov (2009). In order to investigate STI and STT effects, participants in their experiment were told that some of the sentences describe the behavior of the person presented with the sentence (actor or STI trials in which sentences were presented in blue), and that other behaviors were said to be randomly assigned to the faces (non-actor or STT trials in which sentences were presented in red). A randomly assigned irrelevant face was used rather than a "communicator" because in the communicator case, participants might infer that the communicator shares traits of those he is describing. This non-actor condition eliminates any logical association between the trait and the irrelevant person. Thus any association must reflect only simple associative processing.

The main analyses consisted of comparing the differences in "false recognition" (activation of the behavior nodes in the simulation) on the two pairings, *i.e.*, differences between old and new trials – across each level of the relevance factor (actor and non-actor). We predicted that the difference between old and new trials (the pairing factor) would be greater on actor trials (STI) than non-actor trials (STT), thus simulating a stronger STI than STT effect.

2.6.1 *Method*

In this simulation, all the trials from the world knowledge pattern had a frequency of 8, which means they were given as input to the model 8 times. As shown in Table 2, this input creates the association of 8 different behavior-trait pairs. The activation of all the nodes in the patterns sum to 1.0. Thus in this phase the activation for the trait node was 0.5, and for the behavior node it was 0.5 as well. The learning rate parameter (ϵ) for the world knowledge patterns was 0.1. The selection of this value was based on the assumption that the learning of the world knowledge is

Learning Pattern: Task Specific Knowledge

Behaviors								Traits								Presented Person								Non Presented Person			
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**
0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0

Table 2: Learning Pattern: World Knowledge. In the table each row represents a trial, each block of 8 nodes represent a specific type of stimulus (behavior, trait, presented person and non-presented person). Each block has 8 different nodes representing different stimuli of that type. Zero value means that the stimulus/node is not presented/activated in the trial/input. The activation of a specific node may vary between 0 and 1, with the constraint that the sum of activations in each trial equals 1.

slow (which is why the value is low) and is frequency-based (which is why the frequency is 8). Note that the values for these parameters are used in all the remaining simulations.

In order to simulate the difference between STI and STT, in the second part (the task-specific phase in the model and the first part of the false recognition paradigm), the activations of the person nodes for actor and non-actor conditions were different (see Table 3). To simulate a STI trial – where the actor of the behavior is presented in the photo - we set the activation of the behavioral node to 0.5 and the activation for the actor node to 0.5 as well.

To simulate the STT trial – where the pairing of photo and behavior is arbitrary - the input activation to the behavioral node was the same as in the STI condition (0.5) as there are no reasons to expect difference between the activation of the sentence in STI and STT. However, the activation for the photo-person node was weaker (0.25). There is the actor mentioned in the sentence but not pictured, the behavior in the sentence, and the non-actor photo. So attention is divided among these three elements rather than two, which means less attention is paid to some of them. We set the activation for both the actor and the pictured non-actor at 0.25. As proposed by Brown and Bassili (2002), these attentional differences cause differences in the strength of associations between the person and the trait. In the simulated experiment, this difference in attention is presumably produced by instructions to participants at the beginning of the study (with blue sentences signaling STI trials and red sentences signaling STT conditions).

Learning Pattern: World Knowledge

Behaviors								Traits								Presented Person								Non Presented Person			
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**
0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0.25	0	0	0
0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0.25	0	0
0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0
0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0
0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0.25	0
0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.25	0	0	0	0	0	0.25
0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0
0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0

Table 3: Learning Pattern: Task Specific Knowledge

In Table 3, the STT trials are listed in the 1st, 2nd, 5th and 6th rows (with 0.25 for face input activation for the presented person columns -irrelevant person- and for the non-presented person columns – the relevant one –) and the remaining rows correspond to the STI trials (with 0.5 for face input activation, corresponding to the presented person that is the relevant one in these trials). Thus, we had 4 trials per condition, but half of them are going to be used as controls (new pair trials in the FRP described before) in the test phase. The learning rate (ϵ) for this second learning part was higher than the learning rate for world knowledge because participants are explicitly asked to memorize the information, which makes them expend more effort in the task so that learning is more efficient. Thus the value of the learning rate was set to 0.4.

After the model memorized the association between behaviors and faces, we examined what the model learned and compared the results with those from Goren and Todorov (2009). As in their paper, where participants were asked whether the trait was or was not presented in the sentence with that specific face, requiring them to try to recall the behavioral description, we gave the model an incomplete pattern as input (Table 4). Only faces and traits were activated in this pattern. Then, we observed the model's response, *i.e.*, its output in the activation of the behavior nodes.

As noted above, half the trials were new pairs (the last 4 rows in the Table 4, 2 for STI and 2 for STT), where the traits were presented with mismatched faces. Thus, as in Goren and Todorov (2009), the overall design is a 2 (Pairing: old versus new pairing) $\times$ 2 (Relevance: actor versus non-actor) within-subjects ANOVA.

Simulation 1: Test pattern

Behaviors								Traits								Presented Person								Non Presented Person			
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**
?	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0
0	?	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0
0	0	?	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0
0	0	0	?	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0
0	0	0	0	?	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0
0	0	0	0	0	?	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0
0	0	0	0	0	0	?	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0.5	0	0	0	0
0	0	0	0	0	0	0	?	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0.5	0	0	0	0	0

Table 4: Simulation 1: Test pattern

2.6.2 *Simulation Results and Discussion*

The dependent variable in all the simulations is the final activation of the behavior node. There was a main effect of Relevance (actor versus non-actor), $F(1,192) = 55.48$, $p < 0.001$, $\eta^2 = 0.09$, and a main effect of Pairing (old versus new pairs), $F(1,192) = 247.48$, $p < 0.001$, $\eta^2 = 0.50$, found in this simulation. There was also a significant Relevance $\times$ Pairing interaction, $F(1,192) = 33.45$, $p < 0.001$, $\eta^2 = 0.09$ (see Figure 2). There was a STI effect because the mean activation for the old trials was significantly greater ($M = 0.16$, $SD = 0.02$) than the new trials ($M = 0.11$, $SD = 0.02$), $t(49) = 15.62$, $p < 0.001$, $d = 3.14$, 95% CI $[0.05, 0.06]$. The same effect occurred for the mean difference in the non-actor condition, with old trials ($M = 0.13$, $SD = 0.02$) showing more activation than new trials ($M = 0.11$, $SD = 0.02$), $t(49) = 5.40$, $p < 0.001$, $d = 1.17$, 95% CI $[0.01, 0.03]$. To test whether there was a significant difference between STI and STT effects, as there is in the empirical study, we calculated the differences between the old and new pairings for the actor and for the non-actor conditions. This difference was larger for actor ($M = 0.05$, $SD = 0.02$) than for non-actor ($M = 0.02$, $SD = 0.03$), $t(49) = 5.78$, $p < 0.001$, $d = 1.26$, 95% CI $[0.02, 0.04]$. The pattern of behavioral activation is shown in Figure 2.

To understand the results, it is important to realize that the behavioral activation in the old pairing is the outcome of two different associations, the trait - behavior associations and the behavior – face associations. The trait – behavior associations do not vary between STI and STT conditions since they were learned in the world knowledge phase which is identical for STI

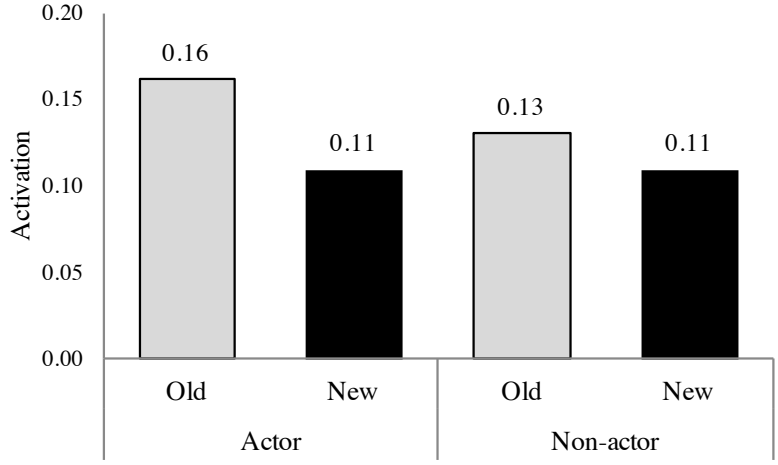

Figure 2: Simulation 1: Mean activation of the behavior node.

and STT. This means that if we only presented the trait in the test phase (without the face), the activation of the behavioral node would be similar in STI and STT conditions. The second type of association, behavior – face, was created in the task-specific phase where two kind of trials were used: a behavior – actor pairing that results in a strong weight between these two nodes, and a behavior – non-actor pairing where the link is weaker due to the assumed distribution of attention mentioned above. This means that in the test phase, links to the faces are responsible for the different results in actor and non-actor conditions. In the STI condition, the actors are a better cue for the retrieval of the behavior than the non-actors in the STT condition. In the STI condition, the traits and especially the faces activated the behavioral node more than in the STT condition, where the trait cues work in a similar way but the face cue is less effective, so the spread of activation to the behavioral node is less. To sum up, the results for the old parings were obtained because 1) learning world knowledge set up behavior-trait links, 2) learning the task-specific knowledge set up person – behavior links, and 3) these links were stronger for the actor than for the non-actor. For the new pairings, the activation of behaviors is the smallest (black bars in Figure 2) because these faces were not linked to the behaviors at all.

This simulation shows that this autoassociative model is able to simulate both STI and STT effects, as well as the typical magnitude difference between the two. It also suggests that is not necessary to consider an additional process (Gorenand Todorov, 2009; Skowronski et al., 1998) to explain this specific difference because differential activation of the person nodes (with

Learning Pattern: Task-Specific Knowledge

Behaviors								Traits								Presented Person								Non Presented Person							
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0
0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0
0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0.1	0	0	0	0	0
0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0.1	0	0	0	0
0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0.1	0	0	0
0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0.1	0	0
0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0.1	0
0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0.1

Table 5: Simulation 2: Task-Specific Knowledge Pattern

stronger activation for actor than non-actor nodes) was sufficient to simulate the experimental results.

2.7 SIMULATION 2 – RELEVANT AND IRRELEVANT TARGETS SIMULTANEOUSLY PRESENTED

Todorov and Uleman (2004) used a different kind of manipulation in order to reduce the STT effect in the false recognition paradigm. In their version, both a relevant and an irrelevant target face were presented in the same trial with the behavior. The presence of relevant faces diminished the STT effect. Todorov and Uleman argued that the presence of the actor leads to a deactivation of the associative process and hence to a failure of STT (Todorov & Uleman, 2004; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Leonards, 2008).

The design of this simulation was a 2 (Faces presentation: simultaneous presentation, *i.e.*, actor and non-actor, versus standard, *i.e.*, only actor or only non-actor) $\times$ 2 (Pairing: old versus new pairs) $\times$ 2 (Relevance: actor versus non-actor) ANOVA, with the first factor between-subjects (simulations) and the rest within-Ss.

2.7.1 *Method*

The world knowledge was the same as in Simulation 1. Table 5 lists the modified task-specific knowledge patterns. As in Simulation 1, actor faces were presented with a higher value (0.4 in

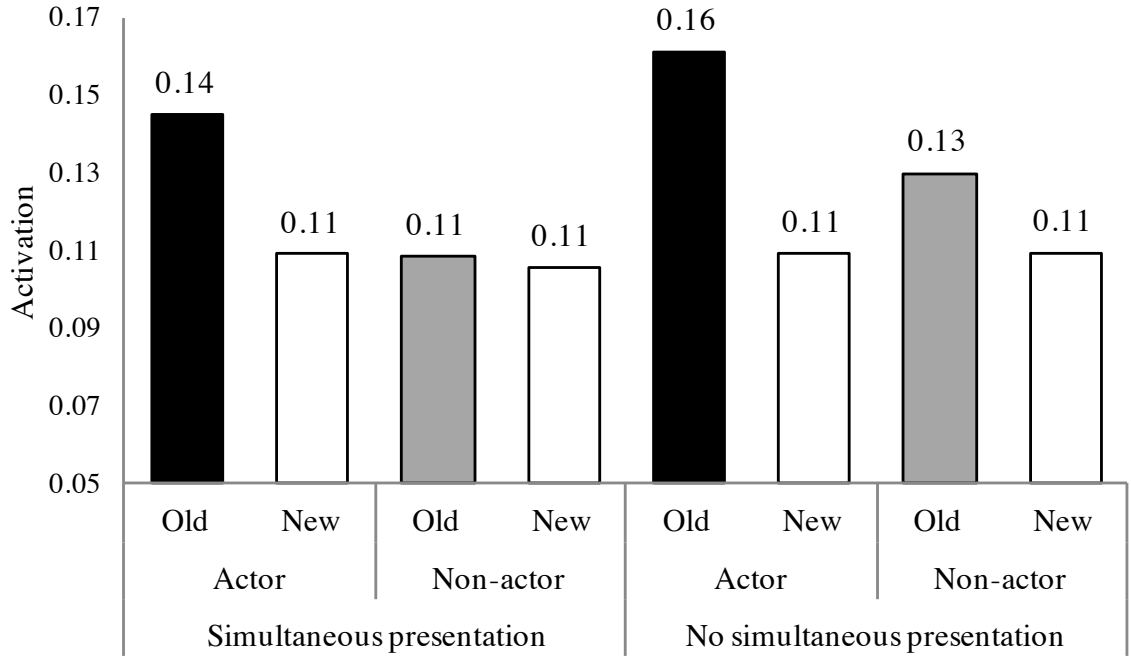

Figure 3: Simulation 2: the results of Simultaneous Presentation versus No Simultaneous Presentation simulation; the dependent variable is the behavioral activation difference between old and new pairing conditions.

this case) than non-actor faces (0.1). However, and importantly, these two levels of attention were entered in the same rows/input reflecting the simultaneous presentation of the two faces in Goren and Todorov's experiment. Here, 0.5 units of attention were divided between the two photos, and since the relevance of the actor is higher in this context (compared to a non-relevant person presented alone), more attention is paid to the actor than to the non-actor. The remaining 0.5 was assigned to the behavior. This implementation of Goren and Todorov's experiment illustrates how our idea of divided attention can easily be generalized to a different experimental design. In the test pattern, as in all other simulations, only one face (the actor's or non-actor's) and one trait was activated, in the incomplete pattern to be completed by spreading activation.

2.7.2 *Simulation Results and Discussion*

We found a main effects of face Presentation, $F(1,388) = 29.16$, $p < 0.001$, $\eta^2 = 0.04$, of Pairing, $F(1,388) = 302.87$, $p < 0.001$, $\eta^2 = 0.30$, and of Relevance, $F(1,388) =$

169.39, $p < 0.001$, $\eta^2 = 0.12$. The was no significant Presentation $\times$ Relevance interaction, $F(1,388) = 1.46$, $p = 0.23$, $\eta^2 = 0.00$, but importantly there was a significant Presentation $\times$ Pairing interaction effect, $F(1,388) = 27.94$, $p < 0.001$, $\eta^2 = 0.03$ and also a significant Relevance $\times$ Pairing interaction, $F(1,388) = 74.88$, $p < 0.001$, $\eta^2 = 0.10$. There was no 3-way interaction, $F(1,388) = 0.05$, $p = 0.82$, $\eta^2 = 0.00$. But note that this is not the effect we are looking for, because we are interested in comparing very specific cells of this design (specifically the simultaneous presentation condition). When faces were presented alone, as in a standard STI/ STT study, both effects emerged. The difference in behavior activation between old and new trials was significant for the actor $t(49) = 15.62$, $p < 0.001$, $d = 3.14$, 95% CI $[0.05, 0.06]$), and was smaller but still significant for non-actor $t(49) = 5.40$, $p < 0.001$, $d = 1.17$, 95% CI $[0.01, 0.03]$). STI persisted when targets were presented simultaneously, with a new versus old trials mean difference of 0.04 ($SD = 0.02$), $t(49) = 11.38$, $p < 0.001$, $d = 2.09$, 95% CI $[0.03, 0.04]$. But consistent with past empirical findings, STT disappeared just like in Experiment 1 conducted by Crawford and collaborator (2007). The difference in activation between old ($M = 0.11$, $SD = 0.02$) and new trials ($M = 0.11$, $SD = 0.02$) for the non-actor showed no significant STT effect, $t(49) = 0.71$, $p = 0.48$, $d = 0.14$, 95% CI $[0.00, 0.01]$, in the simultaneous face presentation condition. Thus, the simultaneous presentation of the relevant and irrelevant targets affected STT but not STI (Figure 3). In the model this occurs because of the different values in the learning input (different levels of attention), without any assumption about differences in the processes as posited by some authors (Goren & Todorov, 2009; Todorov & Uleman, 2004; Crawford, Skowronski, & Stiff, 2007; Crawford et al., 2008).

2.8 SIMULATION 3 – LIE-DETECTION TASK

Crawford and collaborators, in 2007, conducted a study where the savings-in-relearning paradigm was combined with a lie-detection task. As in previous savings-in-relearning experiments, participants were first exposed to photos paired with descriptions of behaviors.

However, rather than memorizing or familiarizing themselves with the stimuli, the authors asked participants to decide whether the person in each photo was lying. Relevance was manipulated with self-descriptive relevant actors versus other-informant irrelevant persons. After the

lie-detection task, the same photos were presented again but this time paired with a single word. The task here was to memorize the word-photo pairs. In some of these trials, photos previously seen were paired with traits implied by behaviors presented in the initial phase with that picture –relearning trials; others had novel photo – trait pairs. Finally, in the last part, participants had to recall the words/traits (from the memorization task) that were cued with actor, informant, or novel photos. Recall performance showed the saving effect, *i.e.*, photo-word pairs that repeated photo-trait pairs from the behavioral presentation task were learned better than novel pairs.

The lie-detection manipulation diminished the recall performance in the self-descriptive condition (actor) leading to similar sized STI and STT effects. Crawford and colleagues (2007) concluded from these results that the attributional process that normally causes strong actor-trait linkages (in the initial familiarization or memory tasks) is disrupted by the lie-detection instruction, leaving only association processes common to STI and STT. Despite this explanation, it remains unclear exactly what happens when we ask a participant to detect whether the person in the photo is lying. Here we assume that rather than disrupting attributional processes, more attention is paid to the photos under lie-detection, and that the amount of attention is similar in STI and STT conditions. The rationale behind these assumptions is that by asking participants to detect whether the person is lying, most of their attention focuses on the face in the photo, regardless the relevance (whether the person is communicating her own behavior or someone else's behavior). This assumption is supported by the lie-detection literature which is replete with evidence that people believe that they can detect lying by carefully attending to the actor's appearance and facial behavior – averted gaze, speech fluency, etc. – even though most evidence contradicts this (*e.g.*, Bond & DePaulo, 2006). The behavioral nodes receive less activation (because they are not relevant in this context anymore), but the same amount in STI and STT, because people are not even asked to memorize the behaviors (the instruction is to detect liars).

The design of this simulation is a 2 (Instruction: lie-detection versus memorization) $\times$ 2 (Pairing: old versus new pairs) $\times$ 2 (Relevance: actor versus non-actor) ANOVA, the first factor being between-subject and the rest within-subject.

Learning Pattern: Task-Specific Knowledge

Behaviors								Traits								Presented Person								Non Presented Person							
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.85	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0	0
0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.85	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0
0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.85	0	0	0	0	0	0	0	0.05	0	0	0
0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.85	0	0	0	0	0	0	0	0.05	0	0
0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.9	0	0	0	0	0	0	0	0

Table 6: Simulation 3: Task-Specific Knowledge Pattern

2.8.1 *Method*

To simulate this result, the learning rate, the simulated number of subjects in each instruction condition, and the world knowledge inputs were kept the same as in the first simulation. The main difference was in the activation values for both actor (STI) and non-actor (STT) faces, reflecting our assumptions about the distribution of attention in the lie-detection condition (see Table 6). The activation for the behavior was only 0.1 in this simulation because it was not so important in this specific task where the participant's focus is mainly on the person in the picture. The actor's face was set to 0.9 and non-actor's face was set to a similar value (0.85), while the remaining 0.05 was reserved for the actor mentioned in the STT sentences. The nodes for the actor and non-actor faces were both highly activated because our hypothesis is that participants strongly focused on the photos to detect whether or not the persons were lying.

2.8.2 *Simulation Results and Discussion*

The ANOVA on behavior activation showed a main effect for Pairing type, $F(1,388) = 193.40$, $p < 0.001$, $\eta^2 = 0.33$, for Relevance, $F(1,388) = 30.55$, $p < 0.001$, $\eta^2 = 0.04$, and for Instruction (lie-detection versus memorization), $F(1,388) = 12.70$, $p < 0.001$, $\eta^2 = 0.02$. We also found a Pairing $\times$ Relevance interaction, $F(1,388) = 16.09$, $p > 0.001$, $\eta^2 = 0.02$, a Relevance $\times$ Instruction interaction, $F(1,388) = 12.93$, $p < 0.001$, $\eta^2 = 0.02$, a Pairing $\times$ Instruction interaction, $F(1,388) = 26.01$, $p < 0.001$, $\eta^2 = 0.04$, and a Pairing $\times$ Relevance $\times$ Instruction interaction, $F(1,388) = 22.26$, $p < 0.001$, $\eta^2 = 0.03$. Breaking this down by

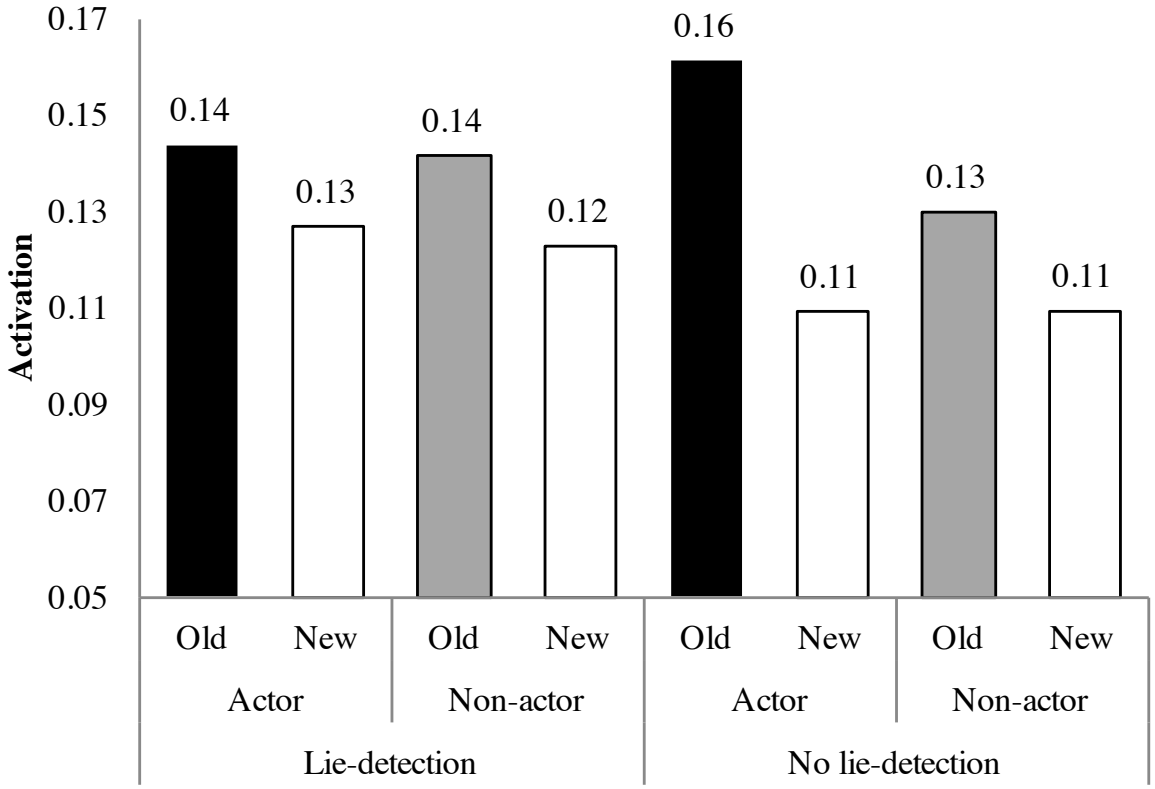

Figure 4: Simulation 3: the results of Lie-Detection versus No Lie-Detection simulation; the dependent variable is the activation difference between the old and new pairing conditions.

instruction, the difference between old and new pairs under lie-detection instructions showed an STI effect ($M = 0.02$, $SD = 0.03$) but not any larger than the STT effect ($M = 0.02$, $SD = 0.03$), $t(49) = -0.31$, $p = 0.76$, $d = -0.06$, 95% CI $[-0.01, 0.01]$ (see Figure 4). STI was greater under memorization than lie detection, $t(49) = 6.35$, $p < 0.001$, $d = 1.34$, 95% CI $[0.02, 0.05]$, whereas STT was not, $t(49) = 0.40$, $p = 0.69$, $d = 0.07$, 95% CI $[-0.01, 0.01]$. These results demonstrate that our assumptions about attention allocation and our corresponding implementation of the lie-detection task in the model, successfully replicated the experimental data. Changing the relative activation of faces and behaviors made it possible to replicate the behavioral data. With the lie-detection instruction, both types of faces receive roughly the same amount of attention which produced similar effects for STI and STT.

2.9 SIMULATION 4 – GENERALIZATION EFFECT

Additional support for the two-process view comes from evidence that the behavior-trait inferences generalize to other traits in STI conditions, whereas no such generalization is found in STT conditions (Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski et al., 1998). This finding is often interpreted as evidence that attributional processes entail implicit theories of personality and thus generalize to other trait dimensions.

Experimentally, the halo effect is usually investigated using a trait rating task (*e.g.*, Carlston & Skowronski, 2005). The first part of the experiment is similar to the first part of the false recognition paradigm. In the second part of the study, participants are asked to rate how much of a specific trait each person possess. Three different types of traits are used: a critical trait that was implied by the behavior in the first part of the study (*e.g.*, "helpful"), a trait consistent with the critical trait's evaluative valence (*e.g.*, "smart"), and a trait inconsistent with the critical trait's valence (*e.g.*, "rude"). In the STI condition, the ratings for the congruent traits were significantly higher than in STT condition, where they were not above chance. The facts that the transference effect was specific to traits implied by the informants' descriptions, and that the impression of the actors was influenced by the implied traits' valence activating non-implied traits with the same valence, was taken as evidence of attributional processes in STI (Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski et al., 1998).

From the simulations so far, it may not be immediately obvious how MATIT could mimic these findings. The success of the simulations is based on the idea that the distribution of attention affects the strength of the weights in the auto-associative memory. However, it is difficult to see how the halo effect could be based on such a relationship. Indeed, simulating the halo effect is based on the auto-associative memory only, not on attention allocations. Thus as part of the MATIT's world knowledge, valence-consistent traits can be associated with each other just as traits can be linked with trait-implying behaviors. In other words, the generalization effect can be realized through MATIT's auto-associative memory. Such an explanation of the halo effect

Learning Pattern: World Knowledge

Behaviors								Implied Traits								Valence-consistent Traits								Presented Person								Non Presented Person			
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**
0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0

Table 7: Simulation 4: World Knowledge Pattern

would be consistent with classic models of implicit personality theory that assume inter-trait relations (*e.g.*, C. A. Anderson, 1995; Schneider, 1973; Carlston & Skowronski, 2005).

The essential design of this study is a 2 (Pairing: old versus new pairs) $\times$ 2 (Relevance: actor versus non-actor) $\times$ 2 (Trait: implied versus valence-consistent) ANOVA, all within-Ss.

2.9.1 *Method*

In this simulation as in the last, we focused on the important conditions and designed a slightly simplified version of the original experimental procedure. We only used implied traits and valence-consistent traits, since the valence-inconsistent traits are not crucial for our demonstration. Besides, the valence-inconsistent trait results did not vary with the actor /non-actor manipulation (Carlston & Skowronski, 2005).

In order to simulate implied and valence-consistent traits in world knowledge, we used two different frequencies for the trait-behavior pairs. For the implied traits (see the first 8 rows in the Table 7) the world knowledge patterns were exactly the same as in simulation 1 and the frequency was also the same (8). For the valence-consistent traits, the pattern was trained with frequency of 2, where the implied traits were associated with the valence-consistent traits, equivalent in real life to fewer observations of the implied traits co-occurring with valence-consistent ones (see from 9th to 16th row in Table 7). The learning rate for world knowledge was 0.1.

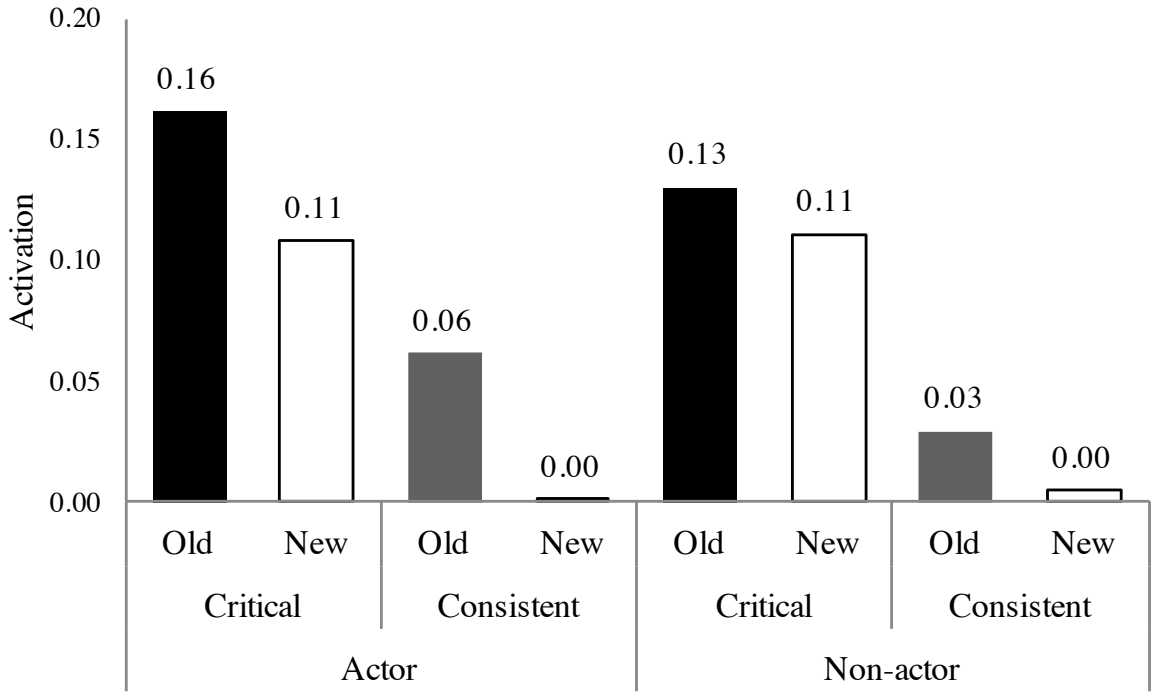

Figure 5: Simulation 4: Mean activation of the behavior node.

In the task-specific learning, the pattern was similar to simulation 1, where the activation of the actor node (0.5) is larger than the activation of the non-actor node (0.25 plus 0.25 for the actor in the sentence). The learning rate in the task specific knowledge was 0.4. The test pattern was similar to simulation 1 as well; where pairs of faces and traits (some of the traits were "implied", *i.e.*, were directly associated to behaviors and other traits were "valence-consistent", *i.e.*, were directly associated to "implied" traits and not to behaviors) were presented, and the activation of behavioral nodes was analyzed.

2.9.2 *Simulation Results and Discussion*

Figure 5 shows the mean activations of the behavior nodes. In a repeated measures 2 (Pairing: old versus new pairings) $\times$ 2 (Relevance: actor versus non-actor) $\times$ 2 (Trait: implied versus valence-consistent) ANOVA, we found a main effect of Pairing, $F(1,384) = 246.34$, $p < 0.001$, $\eta^2 = 0.10$, a main effect of Relevance, $F(1,384) = 30.48$, $p < 0.001$, $\eta^2 = 0.01$, and a main effect of Trait, $F(1,384) = 2065.16$, $p < 0.001$, $\eta^2 = 0.72$. This last effect shows that the

activation for implied traits ($M = 0.13$, $SD = 0.03$) exceeded that for valence-consistent ones ($M = 0.02$, $SD = 0.03$).

Of the interaction effects, only the Pairing $\times$ Relevance interaction was significant, $F(1,384) = 81.67$, $p < 0.001$, $\eta^2 = 0.02$, replicating the simulation 1 results. There was no significant 3-way interaction, $F(1,384) = 0.05$, $p = 0.83$, $\eta^2 = 0.00$, but again it is not this interaction we are looking for. The Pairing $\times$ Relevance interaction occurred because differences between old and new trials were greater for the actor ($M = 0.05$) than for the non-actor ($M = 0.02$), with $t(49) = 4.92$, $p < 0.001$, $d = 1.02$, 95% CI $[0.02, 0.05]$ for the case of the implied traits. This was equally true for valence-consistent traits with $t(49) = 6.29$, $p < 0.001$, $d = 1.10$, 95% CI $[0.02, 0.05]$, showing that the difference between old and new trials for the actor, M = 0.06, was greater than the difference for non-actor, $M = 0.02$, *i.e.*, halo effect for the actor was stronger than the halo effect for the non-actor. Thus, by looking at the valence-consistent traits we can conclude that there is a halo effect for STIs, $t(49) = 12.65$, $p < 0.001$, $d = 2.66$, 95% CI $[0.05, 0.07]$ and a smaller halo effect for STT, $t(49) = 5.30$, $p < 0.001$, $d = 1.02$, 95% CI $[0.02, 0.03]$.

The behavioral activations in Figure 5 can be understood as resulting from: 1) behavior-trait links in world knowledge; 2) behavior-face links from the learning trials in which 3) actor faces are linked more strongly than non-actor faces, as in simulation 1; and 4) indirect behavior links with valence-consistent traits through their links with implied traits. The pairs of faces and traits presented at test determine which of these links is activated and hence the net levels of behavioral activation shown in Figure 5.

2.10 SIMULATION 5: THE ROBUSTNESS OF THE MODEL

In this section we examine the robustness of the model in terms of its parameters. The robustness of the model can be tested by varying the parameters and testing whether the different parameter settings reproduce the empirical results. Since the parameters of the first simulation form the basis for all subsequent simulations, we used this as initial parameters. We also focused on the two crucial parameters, differences of attention in the STT condition and the STI condition, and the learning rate. The former is crucial because if it is very small there would be no a

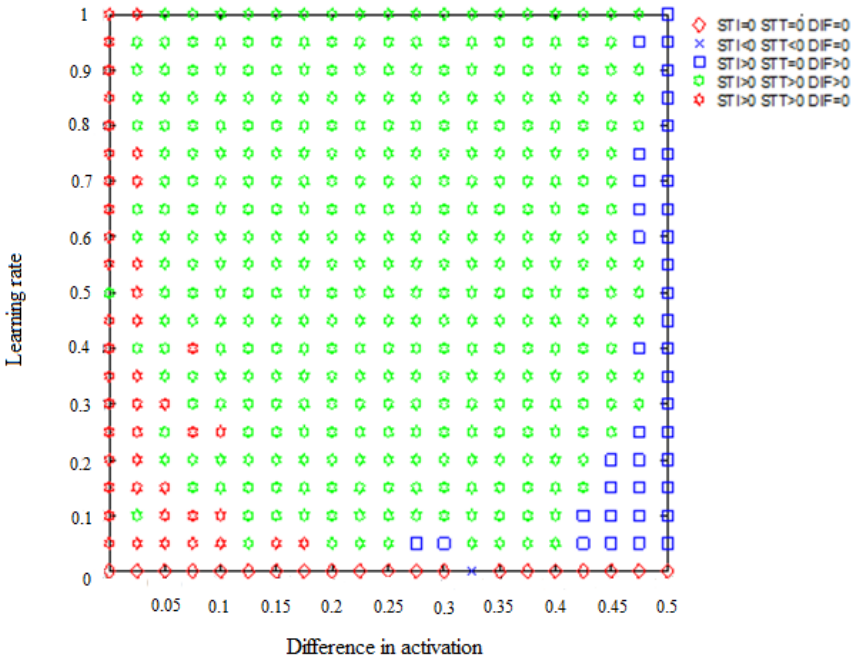

Figure 6: Simulation 5: The graph illustrates the robustness of the model in terms of the two parameters, the actor minus non-actor difference in activation (attention) on the x axis and learning rate (y-axis).The green pentagons indicate the parameter settings for which the simulation replicated the results from Simulation 1 (STI > 0, STT > 0, STI > STT). The diamond markers indicate that MATIT did not produce a significant effect. For the red pentagons, MATIT produced STT and STI effect, but there was no difference between them (STI > 0, STT > 0, DIFF = 0). The squares indicate an STI effect but no STT effect.

difference between STI and STT (red pentagon markers in Figure 6). On the other hand, if too large, *i.e.*, much more attention is paid to the actor than to the non-actor, the STT effect is not observed anymore (square markers in the Figure 6). Furthermore, if the learning rate is too small no learning would occur. Figure 6 shows the results.

The attentional difference ranged from 0 to 0.5 (with a 0.05 interval). Note that the remaining 0.5 is assigned to the behavior node. The learning rate ranged from 0 to 1 (with a 0.1 interval). Thus we run the simulation 230 times using the method described in simulation 1 section. The green markers in Figure 6 indicate that MATIT can replicate the results of Experiment 1 across a broad range of parameter settings, *i.e.* our model is very robust. Similar explorations of robustness could be done for the other results modeled here, but would take us well beyond the scope of this paper.

2.11 SIMULATION 6: FALSIFICATION - DOUBLE DISSOCIATION 1

This simulation is the first of two simulations which examine how MATIT might be falsified. The previous simulation results suggest that MATIT could be falsified by a double dissociation (*e.g.*, Dunn & Kirsner, 2003) as all simulated effects constitute single dissociations. In the context of the STT/STI-effects, a double dissociation would occur if one effect (*e.g.* STI) increases while the other effect (*e.g.* STT) decreases (*i.e.* shows a worse performance than baseline). Our parameter scan did not show such an effect. That is, any combination of the distribution of attention and learning rate could not simulate the experimental result of a double dissociation. (Even though this result is based on the particular method of simulation 1, and simulation 4 showed it is possible to model other methods in MATIT, we focus on this simulation method for now.) When considering the question of falsification, it is also important that the falsification is achieved by plausible data (*e.g.*, Roberts & Pashler, 2000). Hence the question, is a double dissociation a potential outcome of an experiment?

Interestingly, a recent study by Carlston and Skowronski (2005) appeared to find such an effect. Carlston and Skowronski asked participants to familiarise themselves with pairs of behavior descriptions and photos (some were photos of actors and others of communicators) in a savings-in-relearning paradigm. In the following phase of the experiment, they were shown a photo and a trait and had to rate the extent to which they thought the person in the photo possessed the trait. But before the rating task, the authors asked half of the participants to recall whether the target of the informant's description was the self or the other. They observed that this interposed recall task increased the extremity of ratings made of the actors relative to a control (thus, a higher STI) and reduced the extremity of inferences made about communicators. The STTs were even lower than the control condition, thus totally eliminating them. Hence an effort to recall details of the original descriptions seemed to lead to a double dissociation.

However, this apparent double dissociation with STI and STT is merely an effect on recognition performance, not an effect on the processes at encoding that MATIT is designed to model. Carlston and Skowronski's manipulation was not originally meant to differentiate STIs and STTs and thus was carried out during the test phase. This is important because the attribution versus associative debate focuses primarily on encoding processes, and no manipulation was conducted

to affect encoding in this experiment (see the General Discussion for more on this point). Rating the extent to which the person in the photo possessed the trait is also not the best dependent variable, as it is an explicit task about the formed impression. Our focus and the MATIT model concern how participants make inferences in a spontaneous fashion at encoding.

One way to adapt this manipulation to our goals may be to consider an experiment using the false recognition paradigm, but making participants memorize the photos before they memorize the photo-sentence pairs. Thus, they would first be presented repeatedly with the whole set of photos (one by one), with each as either an actor or an informant about the behavioral descriptions that they will see next. After they memorize who has which role, they would memorize the sentence-photo pairs in the usual way. And finally they would do the recognition test, indicating whether the trait was or was not presented in the sentence. We expect this manipulation to negatively affect the level of false recognitions in STT because the participants would know that it wasn't presented as an actor, creating in this way a biased response towards no-answers ("no the word was not presented with this person"). For STI the same knowledge will work in the opposite direction, knowing that the person in the picture was the actor of the behavior will bias the response towards more false alarms.

Two outcomes seem possible, a double dissociation or a reduced difference between STI and STT. First, better memory of the material may increase the STI effect since participants would know very well that that person actually performed the behavior, whereas STT will decrease since they would know that the implied trait is not related to that person which was only the informant. On the other hand, if the difference between STIs and STTs is simply a matter of the amount of attention paid to the photo, then this procedure should guarantee equivalent attention in both cases, and the difference between STI and STT should disappear. Because we do not have the behavioral evidence from such a study, the question here is whether or not MATIT produces a double dissociation for this particular design.

The simulation is a 2 (Faces presentation: multiple presentations versus standard single presentation) $\times$ 2 (Pairing: old versus new pairs) $\times$ 2 (Relevance: actor versus non-actor) ANOVA, with the first factor between-Ss and the rest within-Ss.

Learning Pattern: World Knowledge

Behaviors								Implied Traits								Valence-consistent Traits								Presented Person								Non Presented Person			
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**1**	**2**	**3**	**4**
0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0

Table 8: Simulation 6: World Knowledge Pattern

2.11.1 *Method*

For this simulation, world knowledge is similar to simulation 1, but there are more input patterns where the photos are repeatedly presented alone (8 times for the actor photo and 8 times the communicator photo; see Table 8). Thus the frequency for the learning phase remained at 8, and the learning rate was 0.1. The task-specific learning was exactly the same as in the Simulation 1, in that actor faces were presented with a higher value (0.5) than non-actor faces (0.25).

2.11.2 *Simulation Results and Discussion*

We found a main effect of Face Presentation, $F(1,388) = 42.47$, $p < 0.001$, $\eta^2 = 0.04$; of Pairing, $F(1,388) = 254.39$, $p < 0.001$, $\eta^2 = 0.19$; and of Relevance, $F(1,388) = 27.71$, $p < 0.001$, $\eta^2 = 0.02$. The was no significant Presentation $\times$ Relevance interaction, $F(1,388) = 2.15$, $p = 0.15$, $\eta^2 = 0.002$, no significant Presentation $\times$ Pairing interaction, $F(1,388) = 0.02$, $p = 0.89$, $\eta^2 = 0.00$, and also no 3-way interaction, $F(1,388) = 0.69$, $p = 0.41$, $\eta^2 = 0.00$. The only significant interaction was Relevance $\times$ Pairing, $F(1,388) = 26.91$, $p < 0.001$, $\eta^2 = 0.03$. In terms of the effects we are looking for, as can be seen in Figure 7, and can be concluded from the lack of interactions with the Presentation factor, the main manipulation didn't affect either STI or STT. There was an STI effect, as in the standard simulation (higher

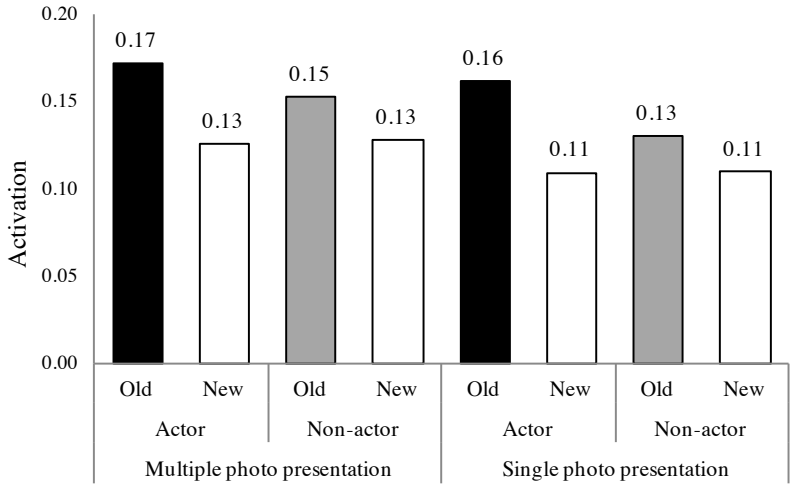

Figure 7: Simulation 6: Mean activation of the behavior node.

false recognition in old trials than in new ones), $t(49) = 6.74$, $p < 0.001$, $d = 1.38$, 95% CI $[0.03, 0.06]$; an STT effect, $t(49) = 4.49$, $p < 0.001$, $d = 1.38$, 95% CI $[0.03, 0.06]$; and a stronger STI than STT effect, $t(49) = 2.31$, $p = 0.03$, $d = 0.50$, 95% CI $[0.003, 0.04]$. Comparing single (standard simulation) and multiple presentations, the STI effect (old minus new trials) in single presentations was not different from the STI effect in multiple presentations, $t(49) = 0.65$, $p = 0.52$, $d = -0.14$, 95% CI $[-0.01, 0.02]$, and the same was true for the STT, $t(49) = -0.63$, $p = 0.53$, $d = 0.12$, 95% CI $[-0.02, 0.01]$.

These results show that the model cannot produce the elimination of the STT effect and an increase in the STI effect with this multiple photo presentation manipulation. This means that if this double dissociation were obtained experimentally, it would falsify our model. We will turn to this point in the General discussion.

2.12 SIMULATION 7: FALSIFICATION - DOUBLE DISSOCIATION 2

This second demonstration has the same aim as the previous one, to examine a plausible double dissociation that, as it turns out, the model cannot simulate. Imagine a hypothetical experiment where we present the same trial twice in the task-specific phase, *i.e.*, present each behavior-photo pair twice. This should differently affect STI and STT if attributions are important. The attribu-

tional view of STI assumes that the processing of the actor and his/her trait-related behavior is deeper and more elaborated than the associative processing of the informant, actor, and behavior in STT. So it is plausible that STI would decrease with an additional exposure to the actor, because participants have the opportunity to better memorize the photo and the behavior and thus to better recall them in the test phase (especially the behavior and the presence/absence of the trait). The better they recall the behavior, the fewer false recognitions they will show. In the STT condition, however, because the processing is more shallow, the actual recollection of the sentence should improve less with the double presentation of the informant photo – behavior pair. The double presentation may nevertheless strengthen the association between the trait and the informant's face, leading to a stronger STT effect.

2.12.1 *Method*

For this simulation, world knowledge is the same as in the simulation 1. The task-specific learning is similar as well. The only difference is the frequency of each pattern of input; two instead of one, because the sentence – photo pairs are presented twice (Table 3). All the rest of the parameters were the same.

2.12.2 *Simulation Results and Discussion*

The ANOVA on behavior activation showed a main effect of Relevance, $F(1,388) = 164.60$, $p < 0.001$, $\eta^2 = 0.07$; of Pairing, $F(1,388) = 1262.97$, $p < 0.001$, $\eta^2 = 0.58$; and no main effect of Repetition, $F(1,388) = 2.09$, $p = 0.15$, $\eta^2 = 0.00$. The was a significant Repetition $\times$ Relevance interaction, $F(1,388) = 7.88$, $p = 0.01$, $\eta^2 = 0.004$; a significant Repetition $\times$ Pairing interaction, $F(1,388) = 143.94$, $p < 0.001$, $\eta^2 = 0.08$; a significant Relevance $\times$ Pairing interaction, $F(1,388) = 121.59$, $p < 0.001$, $\eta^2 = 0.08$; and also a three-way interaction $F(1,388) = 7.83$, $p = 0.01$, $\eta^2 = 0.01$. As one can see in Figure 8, with repetition of the task-specific learning, there is a strong STI effect, $t(49) = 24.24$, $p < 0.001$, $d = 5.31$, 95% CI $[0.10, 0.12]$; a strong STT effect, $t(49) = 16.70$, $p < 0.001$, $d = 3.20$, 95% CI $[0.05, 0.06]$; and

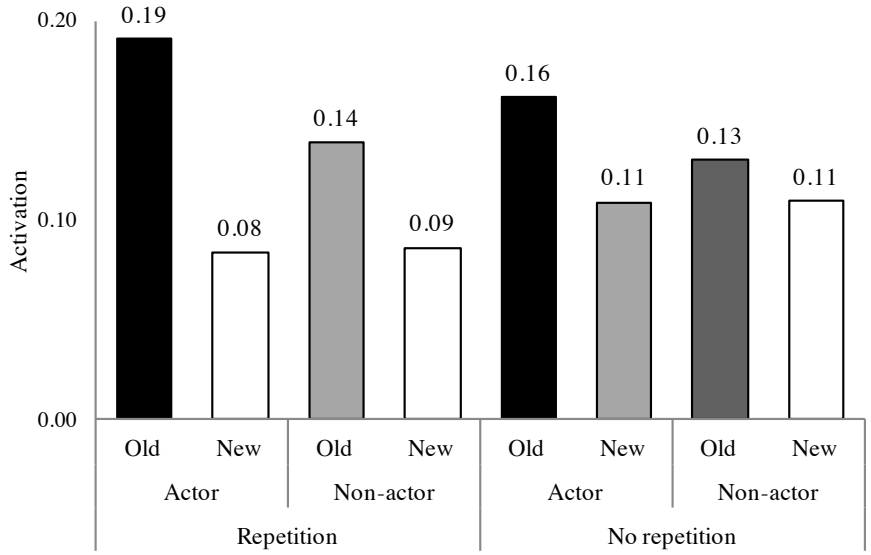

Figure 8: Simulation 7: Mean activation of the behavior node.

a stronger STI than STT effect, $t(49) = 9.48$, $p < 0.001$, $d = 2.00$, 95% CI $[0.04, 0.07]$. So although the STT increased compared with the standard effect under no repetition, ($t(49) = 6.38$, $p < 0.001$, $d = 1.30$, 95% CI $[0.02, 0.4]$), so did the STI effect ($t(49) = 9.51$, $p < 0.001$, $d = 1.98$, 95% CI $[0.04, 0.07]$). Thus once again the model doesn't produce the decrease in STI and increase in STT as is suggested by the attributional vs. associative view.

The first four simulations were developed with an aim of showing what data the model can reproduce (the single dissociations). The two last simulations were conducted to deal with the "overfitting" problem (Roberts & Pashler, 2000), pointing out what data the model would not be able to reproduce (the double dissociations). This is important because every theory or model should provide a way to corroborate or refute itself. Besides illustrating how the model is falsifiable, there is another advantage in thinking about what the model cannot simulate. These two simulations suggest the kind of experiments we should design to seek double dissociations and thus compelling evidence for the existence of two processes or systems (Shallice, 1988). Single dissociations never rule out the possibility of mere quantitative rather than process differences, related in our case to differences in the amount of activation or cognitive resources (*e.g.*, attention, working memory) applied to each task.

2.13 GENERAL DISCUSSION

Current theorizing on social inferences explains the findings on spontaneous trait inference (STI) and spontaneous trait inference (STT) with a dual processing approach, *i.e.*, attributional and associative processes. The current paper presented a computational model (MATIT) framed as a single process approach, the associative process. With this model we sought to demonstrate that the attributional explanation usually offered for the differences between STI and STT is not required by the existing data, and is not even the most parsimonious one. We do not yet have good evidence for dual process claims for STI and STT.

Using simple computer simulation models as gatekeepers that limit premature dual-process conclusions is not new. They have been used in the past to discredit other more complex dual-process models. MINERVA (Hintzman & Ludlam, 1980; Hintzman, 1986), that was created to prove that abstract representations weren't necessary to explain the prototype effect, and BEM (Jozefowiez et al., 2009), that was introduced to disprove the need of metacognition in animal behavior, are examples of computer simulations used for this purpose. However, the current paper goes beyond this common confirmatory approach and discusses which kind of data would disconfirm MATIT, thereby suggesting experimental designs which would support the dual-process view. We will discuss this further at the end of this general discussion. But first we focus the confirmatory part of our work and the theoretical characteristics of MATIT.

Four empirical findings were considered, each of them illustrating a different aspect of the differences between STI and STT. Our hypothesis was that these behavioral data could be simulated by a connectionist model, based on a simple associative learning rule.

The attributional account interprets the difference between STI and STT as being the result of a special linkage which only exists in STI. This link is said to be a "labelled link" created between the actor and the trait, which labels the trait as a property of the actor. In contrast, when the person is not the one who enacts the behavior, a simple association takes place, resulting in a connection only based on space-temporal contiguity. This qualitative difference between STI and STT is said by some authors (*e.g.*, Skowronski et al., 1998; Mae et al., 1999) to account for the larger effects in STI than STT. By contrast, in our associative model, this magnitude difference was ascribed to higher activation values to the actor than to the non-actor in the learning phase.

These levels of activation instantiate the presumed levels of attention paid to stimulus persons as a result of the instructions. These in turn influence the weight of the connections between traits and persons, and subsequently produced the difference in magnitude in STI and STT in the first simulation.

Simulation 2 showed that our model can also account for a result where STT is eliminated or reduced. Presenting the actor at the same time as the irrelevant person during encoding the behavioral description eliminates STT and the link between the trait and the non-actor person (Goren & Todorov, 2009; Todorov & Uleman, 2004; Crawford, Skowronski, & Stiff, 2007). This result per se is interesting as it shows that this misattribution can be prevented by the presence of the relevant target. In our simulation, the external input consisted of three nodes simultaneously activated, the behavior, the relevant person (with high activation value) and the irrelevant person (with lower activation value). The results of this implementation went in the same direction as the behavioral data, showing again that is not necessarily the presence of an attributional process that disrupts the STT effect, but rather differential deployment of node activation and presumably attention.

The third simulation replicates a study where participants are not asked to memorize or familiarize themselves with the material, but rather attempt to judge whether the person in the photo is lying about the behavioral information in the sentence – information that could be about him/herself or about another person. This lie-detection task affects STI but not STT. This has been interpreted as uniquely affecting attributional processes (Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski et al., 1998). But this result was easily simulated by our model. During the encoding phase of the lie-detection condition, we increased the activation values of actor and non-actor to near the same maximum values, to model increased attention to faces in order to detect evidence of lying, and reduced the activation of behaviors to instantiate this presumed shift of attention. This manipulation reduced the activation of the behavior node more for STI than for STT.

The fourth simulation aimed at replicating the halo effect that is more likely to occur in STI than in STT (Skowronski et al., 1998; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007; Wells et al., 2011). This result is said to be attributional because it allows the creation of valence-congruent impressions and generalization to other traits (Carlston

& Skowronski, 2005). To simulate the halo effect, we used two types of traits in the world knowledge provided to the model, an implied trait and a valence-consistent trait with associative links with the implied trait. In the test phase, the activation of the behavioral node (the output of the model) was observed in STI and in STT conditions when the consistent trait and the face were presented. Consistent with the experimental data described in the literature, the behavior activation was superior for STI, showing its higher sensitivity to halo effects relative to STT. Crucially, our model explains these findings as a result of the interplay between different associative strengths in world knowledge of how traits relate to behaviors, and task-specific knowledge arising from differential attention to actor and others.

Our simulations were based on variations in the activation values in the learning patterns related to actor and non-actor target. This difference in the activation of actor and non-actor produces differences in the weight of connections (strength of associations) between behaviors and faces, and consequently between faces and traits. Later in the test phase, when we presented only a face and a trait, the activation of the behavioral node depended on connections among nodes and their weights, and these weights usually benefitted STI more than STT.

While we assumed that the (internal) attention paid to the stimuli is the plausible cause of the differences in activation between STI and STT, this is an assumption that needs to be explored. There are two studies (Crawford et al., 2008) that measured visual attention by recording participants' response times to directional probes in various parts of the display (Experiment 1) or eye movements during encoding (Experiment 2). Two photos were presented simultaneously with one behavioral description. In some conditions, an actor describes his/her own behavior to a bystander; in other conditions, an informant describes the behavior of a target. The results offered no support for the role of "external attention" in the visual modality (Chun et al., 2011) in producing differences between STI and STT. However this study does not rule out the influence of internal attention.

In another study, Skowronski and collaborators (1998) examined whether the STT effect was smaller simply because participants did not pay as much attention to irrelevant behaviors. Behaviors' relevance on each trial in study 2 was not signaled until participants read it, guaranteeing equal attention under STI and STT conditions. However, the STT effect was not affected by this manipulation, ruling out external attention as a plausible explanation. But their findings do

not rule out the influence of internal attention, *i.e.* participant could have stored the displayed information in working memory and only once they knew the relevance of the information was their internal attention directed accordingly. Nevertheless the present paper does not provide behavioral support for the internal attention hypothesis. It only suggest it as a plausible cause for the difference in activation of the actor node and the non-actor node. In fact, it is conceivable that other psychological processes (*e.g.* task setting) can initiate a similar modulation of the activation suggested in our four simulations. For instance, relevance of faces may lead to higher activation of representations whereas irrelevance of faces may lead to lower activation of their representations.

As we stressed through this paper, MATIT is a very simple model. For example, MATIT does not address the complex processes by which people construct verbal descriptions of behavior from observations, and infer trait concepts from those behaviors or descriptions. "Telling the cashier that he received too much change" is but one way to describe an observed behavior. It might also be described as "telling the cashier that she had made a mistake" or "commenting on the contentiousness of those making minimum wage" or "making small talk with the cashier". Encoding a behavior into verbal form, and extracting a summarizing trait concept (or gist or goal or style description) involves complex processes and choices (*e.g.*, Semin & Fiedler, 1992). MATIT does not address these.

Finally we return to theme of falsifying MATIT through exploring plausible data and outcomes that it cannot simulate. In simulations 6 and 7, we demonstrated that the MATIT model would be falsified by double dissociations. We also argued that a double dissociation is a plausible experimental outcome by drawing on the dual-process attribution vs. association account. Thus, one might conclude that a falsification of MATIT by finding a behavioral double dissociation would imply confirmation of the dual-process account. It would not, in part because MATIT can be extended in a sensible way. The auto-associative memory in MATIT has only limited abilities to model complex relationships due to the single layered structure and the linear function governing node activation. This is why it cannot model a double dissociation. Hence, a structure along the lines of a multilayer perceptron/model would be a natural extension, and could model a double dissociation. In fact, multilayer perceptrons are well-known for modeling double dissociations in single process frameworks (*e.g.*, Seidenberg & McClelland, 1989).

But remember that our goal here was not to show that a sophisticated associative model could account for any conceivable data that might be used to support dual-process claims. Our goal was, instead, to show that a very simple associative model that does not posit dual processes could account for major differences between STI and STT that have been interpreted as supporting a dual process account, and to describe some of the consequences of such a model.

2.14 SUPPLEMENTARY MATERIAL

2.14.1 *Details of the Auto-associative Model*

Each node in the model represents a construct with psychologically interpretable meaning. The activation of these nodes leaves a "memory trace" behind that results from changes in the weights of the connections between nodes, *i.e.*, changes in the strength of these connections that are responsible for the learning gains of the model (McClelland & Rumelhart, 1985). These weighted connections between units store the information required to complete familiar (learned) patterns.

During the recall phase, the network receives activation from the exterior – external input. But nodes also receive input from the other nodes in the network – internal input. Considering two nodes, i and j, the input from j to i is i_{ij} and is:

$$i_{ij} = a_j w_{ij} \tag{1}$$

where a_j is the activation of the node j, and w_{ij} is the weight that defines the influence of node j on node i. The internal input is the sum of the activation coming from all the other nodes on the network:

$$int_i = \sum(a_j w_{ij}) \tag{2}$$

The sum of the external and internal inputs is the net input of the node:

$$net_i = ext_i + int_i \tag{3}$$

where ext_i is the external input and int_i is the internal input.

The memory trace is created so as to better anticipate and characterize the future external input. This memory trace is constructed from the discrepancy between the internal input of the network from the last updating cycle and the external input. Mathematically, weights between nodes are adjusted by the delta rule algorithm (McClelland & Rumelhart, 1989):

$$\Delta w_{ij} = \epsilon(ext_i - int_i)a_i \tag{4}$$

where w_{ij} is the connection's weight from j to i, ϵ is the learning rate that defines the learning speed of the model and a_j is the activation of the node j. In a model with enough learning trials and reasonable learning rate, the w_{ij} will tend to zero as the model reaches a stable state. In such a state, the model can anticipate efficiently the input received from the external environment.

Thus, the learning in the model is accomplished through the computation of errors and the updating of the weights between nodes so as to minimize this error.

A linear version of the auto-associative network was applied the current work, and differently from McClelland and Rumelhart (1985), we used only one internal updating cycle rule proposed by Van Overwalle and Labiouse (2004) which allows for faster and simpler simulations. For those unfamiliar with such models, a walk-through example may be helpful, keyed to the false recognition paradigm of Todorov and Uleman (2002, 2003, 2004). Subjects view a series of stimuli, each containing a photo of a person's face and a sentence (on critical trials) describing a trait-implying behavior. They view these stimuli in this first phase in preparation for "a memory test" in the second phase of the paradigm. Some of the sentences contain the trait explicitly, but these merely set up the subsequent false recognition test in which subjects have to remember whether or not the trait was explicitly in the sentence. In the second phase, each memory test item re-presents a person photo paired with a trait term, and subjects must judge whether the trait was explicitly in the sentence they saw earlier. On critical trials, the trait was merely implicit so the correct answer is "No". False recognition errors ("Yes", with appropriate controls) measure the extent to which subjects spontaneously (*i.e.*, unintentionally and unconsciously) inferred traits during the first phase.

Therefore in the model above, there are three concepts (or open circle nodes): B, the trait-implying behavior; T, the trait implied; and F, the person's face. They can each receive external

input, and are connected through bi-directional links to each other. So every node is potentially connected to every other node as well as to the input and output signals. The connections are "potential" because links can vary in conductivity from 0 (not connected) to 1 (connected with complete conductivity or no resistance). A "weight" describes the conductivity of each link, and varies with each trial according to the delta learning algorithm. The algorithm adjusts the weights in the network so that the resultant activation of the nodes on trial n+1 matches more closely their activation on trial n. Activation is introduced into the system through the "external input."

For simplicity's sake, considered only the effects of presenting photos with trait-implying (not trait-explicit) behaviors. Links among nodes in the model are initially set to zero, so that no activation is transmitted from node to node. Trait inference depends on "teaching" the model the world knowledge that particular behaviors are associated with particular traits, and that traits can therefore be "inferred" from these behaviors. Subjects enter the study with this world knowledge, but it must be imparted to the model via the input matrix in Table 2. This shows 8 behaviors, 8 traits, and 8 faces in the model (so the actual running model is more complex than the simple figure). The model "learns" slowly and imperfectly, so that links among nodes are never completely conductive with weights of 1. In simulation 1 (above), world knowledge was imparted by presenting the matrices of Table 2 as input 8 times. In the first presentation, behaviors 1 through 8 are activated by external input, each along with its corresponding trait. (Faces were not activated because world knowledge is about behavior-trait implications, not knowledge about who did what). Imagining that the initial weight between nodes is zero, in the first trial, the delta learning algorithm adjusted the zero weights between pairs of behaviors and traits slightly from zero to a value of .0025. (The change in weight, Δw, is given by $\epsilon(ext - int)a$, where $\epsilon = .01$, $ext = 0.5$, $int = 0$, and $a = 0.5$. Note that although the links are potentially asymmetric, in that activation from node i to j need not be the same as activation from j to i, we've dropped this feature here for simplicity and because it does not affect these simulations.)

After this first input of world knowledge, nodes receive both external activation (ext) from a repetition of the Table 2's paired behaviors and traits (7 more times), and internal activation (int) from other nodes; so calculating Δw becomes slightly more complex. After the second input, $\Delta w = .01 \times (0.5 - .0025) \times 0.5 = .00248$, raising w between behavior and trait to .0050. Δw decreases with each new input, so that after 8 inputs, $w \approx .020$, the conductivity of the link

connecting pairs of behaviors and traits. Thus the pairs of behaviors and traits become linked in the model through their simultaneous activation and the delta learning rule, which moves the model's internal activation values toward the external activation values that it has experienced.

Now that the model "knows" as much as the subjects do, it can participate in the first (study) phase of the false recognition paradigm, learning behavior-face pairs. This occurs through input matrices like Table 2 but with different values, as in Table 3 for simulation 1. Note that all the values for the traits in the matrix are zero, because traits are never presented, only face-behavior pairs. This produces some activation of trait nodes because world knowledge links behaviors and traits. It also associates faces and traits because faces and trait-implying behaviors occur simultaneously. The input matrices are presented once, just as the stimuli are presented to subjects once. Then the test phase follows, in which subjects are asked whether particular traits appeared in behavioral sentences with particular photos. This test is simulated by presenting the model with face-trait pairs and seeing how much they activate the corresponding behavior nodes (because the question is whether or not the trait appeared in the behavior). That is, behavior node activation is read off, and serves as the dependent variable in the false recognition paradigm.

2.15 REFERENCES

Anderson, C. A. (1995). Implicit theories in broad perspective. *Psychological Inquiry*, 286–290.

Anderson, J. R. (1976). *Language, memory, and thought.* Lawrence Erlbaum.

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*(4), 249–277.

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(04), 637–660.

Bassili, J. N. (1976). Temporal and spatial contingencies in the perception of social events. *Journal of Personality and Social Psychology*, *33*(6), 680–685.

Bassili, J. N., & Smith, M. C. (1986). On the spontaneity of trait attribution: Converging evidence for the role of cognitive strategy. *Journal of Personality and Social Psychology*, *50*(2), 239–245.

Berscheid, E., Graziano, W., Monson, T., & Dermer, M. (1976). Outcome dependency: At-

tention, attribution, and attraction. *Journal of Personality and Social Psychology*, *34*(5), 978–989.

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234.

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.

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: bridging cognition and neurophysiology. *Psychological Review*, *112*(2), 291–328.

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.

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: Ii. on the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, *69*(3), 420–436.

Carlston, D. E., & Smith, E. R. (1996). Principles of mental representation. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic prin- ciples* (pp. 184–210). New York, NY: Guilford.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.

Clary, E. G., & Tesser, A. (1983). Reactions to unexpected events the naive scientist and interpretive activity. *Personality and Social Psychology Bulletin*, *9*(4), 609–620.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*(3), 466–472.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Leonards, U. (2008). Seeing, but not thinking: Limiting the spread of spontaneous trait transference ii. *Journal of Experimental Social Psychology*, *44*(3), 840–847.

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.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Diener, C. I., & Dweck, C. S. (1980). An analysis of learned helplessness: Ii. the processing of success. *Journal of Personality and Social Psychology*, *39*(5), 940–952.

Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex*, *39*(1), 1–7.

Fodor, J. A. (1975). *The language of thought* (Vol. 5). Harvard University Press.

Garcia-Marques, L., & Ferreira, M. B. (2011). Friends and foes of theory construction in psychological science vague dichotomies, unified theories of cognition, and the new experimentalism. *Perspectives on Psychological Science*, *6*(2), 192–201.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, *27*(2), 222–248.

Harvey, J. H., Town, J. P., & Yarkin, K. L. (1981). How fundamental is" the fundamental attribution error"? *Journal of Personality and Social Psychology*, *40*(2), 346–349.

Hassin, R. R., Aarts, H., & Ferguson, M. J. (2005). Automatic goal inferences. *Journal of Experimental Social Psychology*, *41*(2), 129–140.

Hassin, R. R., Bargh, J. A., & Uleman, J. S. (2002). Spontaneous causal inferences. *Journal of Experimental Social Psychology*, *38*(5), 515–522.

Hastie, R. (1984). Causes and effects of causal attribution. *Journal of Personality and Social Psychology*, *46*(1), 44–56.

Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley.

Heinke, D., & Backhaus, A. (2011). Modelling visual search with the selective attention for identification model (vs-saim): a novel explanation for visual search asymmetries. *Cognitive Computation*, *3*(1), 185–205.

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, *93*(4), 411–428.

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory and Cognition*, *8*(4), 378–382.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. *Advances in Experimental Social Psychology*, *2*, 219–266.

Jozefowiez, J., Staddon, J., & Cerutti, D. (2009). Metacognition in animals: How do we know that they know. *Comparative Cognition and Behavior Reviews*, *4*, 29–39.

Kanazawa, S. (1992). Outcome or expectancy? antecedent of spontaneous causal attribution. *Personality and Social Psychology Bulletin*, *18*(6), 659–668.

Kashima, Y., Woolcock, J., & Kashima, E. S. (2000). Group impressions as dynamic configurations: The tensor product model of group impression formation and change. *Psychological Review*, *107*(4), 914–942.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–761.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15). Lincoln.

Kelley, H. H., et al. (1972). *Causal schemata and the attribution process*. General Learning Press Morristown, NJ.

Kressel, L. M. (2011). *The functional meaning of traits and spontaneous trait inferences.* (Unpublished doctoral dissertation). New York University.

Kressel, L. M., & Uleman, J. S. (2010). Personality traits function as causal concepts. *Journal of Experimental Social Psychology*, *46*(1), 213–216.

Lau, R. R., & Russell, D. (1980). Attributions in the sports pages. *Journal of Personality and Social Psychology*, *39*(1), 29–38.

Lepsien, J., & Nobre, A. C. (2006). Cognitive control of attention in the human brain: Insights from orienting attention to mental representations. *Brain Research*, *1105*(1), 20–31.

Maass, A., Colombo, A., Colombo, A., & Sherman, S. J. (2001). Inferring traits from behaviors versus behaviors from traits: The induction–deduction asymmetry. *Journal of Personality and Social Psychology*, *81*(3), 391–404.

Mae, L., Carlston, D. E., & Skowronski, J. J. (1999). Spontaneous trait transference to familiar communications: Is a little knowledge a dangerous thing? *Journal of Personality and Social Psychology*, *77*(2), 233–246.

Malle, B. (2007). Attributions as behavior explanations: Toward a new theory. In D. Chadee & J. Hunterm (Eds.), *Current themes and perspectives in social psychology.* (pp. 3–26). St. Augustine, Trindade: SOCS, The University of the West Indies.

Mavritsaki, E., Heinke, D., Allen, H., Deco, G., & Humphreys, G. W. (2011). Bridging the gap between physiology and behavior: evidence from the ssots model of human visual attention. *Psychological Review*, *118*(1), 3.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*(2), 159–188.

McClelland, J. L., & Rumelhart, D. E. (1989). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. MIT press.

Naveh-Benjamin, M., Craik, F. I., Perretta, J. G., & Tonev, S. T. (2000). The effects of divided attention on encoding and retrieval processes: The resiliency of retrieval processes. *Quarterly Journal of Experimental Psychology*, *53*(3), 609–625.

Pittman, T. S., & Pittman, N. L. (1980). Deprivation of control and the attribution process. *Journal of Personality and Social Psychology*, *39*(3), 377–389.

Pyszczynski, T. A., & Greenberg, J. (1981). Role of disconfirmed expectancies in the instigation of attributional processing. *Journal of Personality and Social Psychology*, *40*(1), 31–38.

Read, S. J., & Montoya, J. A. (1999). An autoassociative model of causal reasoning and causal learning: Reply to van overwalle's (1998) critique of read and marcus-newhall (1993). *Journal of Personality and Social Psychology*, *76*(5), 728–742.

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review*, *86*(1), 61–79.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? a comment on theory testing. *Psychological Review*, *107*(2), 358–367.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 501–518.

Schneider, D. J. (1973). Implicit personality theory: A review. *Psychological Bulletin*, *79*(5), 294–309.

Schneider, D. J. (2005). *The psychology of stereotyping*. Guilford Press.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Semin, G. R., & Fiedler, K. (1992). *Language, interaction and social cognition.* London: Sage Publications, Inc.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge University Press.

Shultz, T. R., & Lepper, M. R. (1996). Cognitive dissonance reduction as constraint satisfaction. *Psychological Review*, *103*(2), 219–240.

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.

Smith, E. R., & DeCoster, J. (1999). Associative and rule based processing. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 323–336). New York, NY: Guilford.

Swann, W. B., Stephenson, B., & Pittman, T. S. (1981). Curiosity and control: On the determinants of the search for social knowledge. *Journal of Personality and Social Psychology*, *40*(4), 635–642.

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.

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.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology*, *28*, 211–279.

Van Overwalle, F. (1998). Causal explanation as constraint satisfaction: A critique and a feed-

forward connectionist alternative. *Journal of Personality and Social Psychology*, *74*(2), 312–328.

Van Overwalle, F., & Jordens, K. (2002). An adaptive connectionist model of cognitive dissonance. *Personality and Social Psychology Review*, *6*(3), 204–231.

Van Overwalle, F., & Labiouse, C. (2004). A recurrent connectionist model of person impression formation. *Personality and Social Psychology Review*, *8*(1), 28–61.

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.

Wexler, K. (1978). A review of john r. anderson's language, memory, and thought. *Cognition*, *6*(4), 327–351.

Wigboldus, D. H., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252.

Wong, P. T., & Weiner, B. (1981). When people ask" why" questions, and the heuristics of attributional search. *Journal of Personality and Social Psychology*, *40*(4), 650–663.

Yoon, E. Y., Heinke, D., & Humphreys, G. W. (2002). Modelling direct perceptual constraints on action selection: The naming and action model (nam). *Visual Cognition*, *9*(4-5), 615–661.

3

THE INVOLVEMENT OF ATTENTION IN SPONTANEOUS TRAIT INFERENCES AND SPONTANEOUS TRAIT TRANSFERENCES

Diana Orghian, Leonel Garcia-Marques and Dietmar Heinke

In preparation

In Chapter 2 we suggested that actors are paid more attention to than people that are irrelevant to the behavior. In the model, these differences in attention motivated the differences in associations between the traits and the persons. In the experiments presented in this Chapter we try to test this attentional assumption, that is, we examined the influence of attention on STI and STT trials. In experiment 1 we employed a modified spatial cueing paradigm where participants had to detect a probe appearing immediately after the face disappeared. In experiments 2 and 3 we monitored participants' eye movements while they were memorizing faces and behaviors. In these three experiments we found that more attention was engaged by the actor's face than it was by the irrelevant person. Although, we did not find differences between STI and STT effects (they present the same magnitude), we did find that attention is engaged more strongly in STI than in STT. In the third experiment we also found a correlation between the STI effect and the number of times the participants fixated on the actor's face, and also a correlation with the number of transitions made by participant's eyes between the actor's face and the behavioral description. These results highlight an important and usually understudied factor in spontaneous trait inferences and they also point out the need to control for attention in future experiments.

3.1 INTRODUCTION

Let us imagine that your new colleague, Ben, told you he recently ran a marathon. In another scenario, imagine that Ben told you that his friend, Tom, recently ran a marathon. How much attention do we pay to Ben in the first scenario in comparison to the second? In the first case the attended person, Ben, is the actor of the described behavior, whereas in the second case Ben is a communicator of Tom's behavior. The question that we would like to answer in this Chapter is whether the amount of attention paid to a person describing a particular behavior (*e.g.*, Ben's description about running a marathon) varies depending on the person's relevance to that behavior (*e.g.*, whether Ben or Tom ran the marathon). The answer to this question might seem trivial, and, intuitively, one would say that relevance affects attention. As we discuss below, however, the few experimental studies that have approached this question tend to conclude the opposite, *i.e.*, that relevance does not affect attention.

Why is this important? The level of attention to the provided information can affect the way that information is processed, memorized and organized (for a recent review see Chun, Golomb, & Turk-Browne, 2011). Thus, the question we would like to answer is whether attention can also influence the way traits about others' personality are inferred. People often infer personality traits from others behaviors in an attempt to organize or categorize the social world. This phenomenon of organizing social targets in terms of personality traits (*e.g.*, friendly, rude) can occur spontaneously. For instance, when talking to Ben we are not necessarily trying to form an impression about him, but we may end up creating one unintentionally. Such spontaneous trait inference (STI) arises when we categorize a person based on his behavior without any awareness or intention. Thus, a personality trait is derived from the behavior and associated with the actor of that behavior (*e.g.*, Uleman, Newman, & Moskowitz, 1996). For example, when Ben runs a marathon, others may perceive him as athletic.

In contrast, sometimes a personality trait is assigned to the wrong person (*i.e.*, someone that is not the actor of the behavior from which the trait was inferred). This person can be present while processing that behavior, for instance, the communicator of the behavior, a simple bystander (Skowronski, Carlston, Mae, & Crawford, 1998) or even an inanimate object such a banana (Brown & Bassili, 2002). Thus, a personality trait may be transferred to irrelevant stim-

uli (Carlston & Skowronski, 2005), a phenomenon called spontaneous trait transference (STT). In our example, Ben might be perceived as athletic when he described how Tom ran a marathon. While research to date agrees that STI allows us to organize our social world, and that STT is an error arising as a by-product of our cognitive functioning, the mechanisms underlying STI and STT are disputed (*e.g.*, Orghian, Garcia-Marques, Uleman, & Heinke, 2015). The goal of the present research is to examine whether attention is differently involved in STI and STT since such a difference could contribute to some of the empirical differences between STI and STT. Specifically, we hypothesize that, when people receive information about someone's behavior, more attention will be paid to the person relevant to that behavior (*i.e.*, the actor) than to irrelevant people (*i.e.*, a communicator, bystander, informant, or a randomly paired face). To test this hypothesis, we conducted three experiments with two different methods. This way we were able to produce converging evidence for our hypothesis. In the first experiment we applied a modified version of the spatial cueing paradigm. In the second experiment we combined a standard experimental design used in STI and STT studies (the false recognition paradigm) with a commonly accepted method in measuring attention (the eye movements recording).

Examining this hypothesis can be a step forward in answering one of the main questions that is currently debated in the STI and STT literature - the question concerning the processes responsible for these two phenomena, and, specifically the process that underlies the linkage between the personality trait and the person. There are important empirical differences between STI and STT that are motivating the debate. For example, the magnitude of the STI effect is usually greater than the STT effect (*e.g.*, Bassili & Smith, 1986; Goren & Todorov, 2009; Skowronski et al., 1998). Furthermore, STI and STT pose different sensitivities to negative behaviors (Carlston & Skowronski, 2005), with more trait-inferences being made from negative behaviors in STI than in STT. The phenomena are also associated with a different size of the halo effect (Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski et al., 1998), with more trait-generalization in STI than in STT. These differences have compelled some researchers to postulate that STI and STT involve different processes (*e.g.*, Carlston & Skowronski, 2005). Specifically, STI is assumed to involve an attributional process (*i.e.*, the trait is seen as a property of the actor), whereas STT is based solely on simple associative links (*i.e.*, based on temporal and spatial contiguity). By contrast, Bassili and colleagues (Bassili & Smith, 1986;

Brown & Bassili, 2002) advanced a single process view in which both STI and STT are based on the same simple, automatic associative links between the trait and the person, with stronger links in STI. Despite the differences between the dual and single process view, both accept that there are differences between STI and STT. We argue that there is one more difference between STI and STT: an attentional one.

Attention is central to information processing (Chun et al., 2011). A key property of visual attention involves the selection of relevant perceptual stimuli among competing alternatives (Chun et al., 2011). Visual attention may be facilitated by social cues, such as faces (*e.g.*, Hershler & Hochstein, 2005) and animations (*e.g.*, New, Cosmides, & Tooby, 2007); conversely, visual attention is also directed by intention (*e.g.*, Serences et al., 2005). Further, the degree of visual attention paid to perceptual information determines the depth to which the information is processed (*e.g.*, Lamme, 2003). Perceptual information receiving high levels of visual attention may receive great memory storage as a consequence of strong encoding (Craik & Lockhart, 1972). In support of this idea, emotionally charged stimuli have been found to facilitate visual attention, and improve memory of those stimuli as a consequence of stronger encoding (*e.g.*, Öhman & Mineka, 2001). Along similar lines, Orghian and colleagues (2015) postulated that visual attention might contribute to the differing STI and STT effects. They predicted that the proportion of visual attention paid to relevant faces (*i.e.*, actors) was greater than the proportion paid to irrelevant faces (*i.e.*, people randomly paired with behaviors). This difference in attention is leading to stronger encoding, stronger linkages in memory, and consequently to stronger STI than STT effects. In the present paper we aim to validate this hypothesis.

To be clear, we do not argue that attention is the only mechanism underlying the differing processes of STI and STT. We believe that there is something substantially different between these two phenomena, and discovering what this something is is currently one of the main challenges in the field. We believe that the provided information (*e.g.*, the behavioral description and the corresponding inferred trait) modifies the representation of the actor, but does not lead to significant modifications in the representation of an irrelevant person. In other words, we believe that the way in which this information is organized or integrated into existing knowledge should vary depending on the relevance of the person. However, we propose that attention may be a contributing factor to these substantial, quantitative differences between STI and STT.

As far as we know in the spontaneous trait inference literature, the only study to investigate the role of visual attention in STI and STT in detail was conducted by Crawford, Skowronski, Stiff and Leonards (2008). Critically these authors did not obtain attentional differences between STI and STT. Similarly to the present paper, the authors adopted a spatial cueing paradigm and an eye-tracking device as methods of measuring visual attention. However, the aim of Crawford and colleagues' work was to ascertain that the different processes underlying STI and STT were related to the presence of attribution as opposed to attentional differences, as explained further below. The STI and STT effects are typically operationalized by presenting one photograph of a face alongside a trait-implying behavioral description (see experiment 2 or 3 for details), whereas the design adopted by Crawford and colleagues involved the presentation of two photographs alongside one trait-implying behavioral description. The authors presented two photographs in order to replicate a previous finding in which the STT effect was eliminated by presenting an actor's face and an informant's face simultaneously (Crawford, Skowronski, & Stiff, 2007). With this design, the authors aimed to demonstrate that the elimination of the STT effect was due to an attribution process (triggered by the presence of the actor) that disrupted the association between the trait and the irrelevant person. In the first experiment, Crawford and colleagues (2008) measured visual attention by using probe detection latencies (*e.g.*, MacLeod, Mathews, & Tata, 1986). In each trial, following a short presentation of the two faces, participants were presented with an arrow pointing either upwards or downwards. The arrow appeared in either the actor's location or in the informant's location, and the participants indicated the orientation of the arrow with a key press. Facilitated detection was assumed to signify that there was more visual attention being paid to the location of the arrow. The findings showed that reaction time did not differ between the two locations, suggesting that the actor and the irrelevant person received a comparable proportion of visual attention. In the second experiment the authors, recorded eye-movements with an eye-tracking device and found that participants were not fixating the actor's face more than the informant's face. Overall, the authors concluded that visual attention is not an important factor in STI and STT. Conversely, we argue that these findings do not rule out the possibility that attention is an important factor. Our goal, in fact, is to further explore, with a much simpler design, the possibility that attention is a considerable factor underlying the differing processes of STI and STT.

Crawford and collaboratos' (2008) data suggest that the disruption of the STT effect, previously found (Crawford, Skowronski, & Stiff, 2007), may not have been due to attention, even though this conclusion is based on a null effect. And importantly, this finding does not necessarily indicate that the distribution of attention is comparable in STI and STT. For instance, one can speculate that the simultaneous presentation of two faces may have led to an additional attentional effort to focus on the stimuli in an attempt to compensate for the simultaneous execution of attributional and associative processes. The test would be much clearer if attention was assessed separately in two conditions. Thus, in our research, we present only one person per trial. Moreover, as we argue later, the authors' probe detection task did not control for confounding influences from spatial Inhibition of Return (IOR) effects (see the Procedure section of experiment 1 for details).

Goren and Todorov (2009) also tried to control for the possible influence of attention in STI and STT effects – as opposed to evaluate the role of attention - by using the False Recognition Paradigm (FRP). The FRP operationalizes the typical situation of STI and STT by presenting a photograph of either an actor or an irrelevant person alongside a trait-implying behavioral description (*e.g.*, Ben ran a marathon; see experiment 2 for a detailed description). The authors aimed to demonstrate that the smaller effect of STT may not be due to a reduction in attention paid to irrelevant faces. Thus, they presented the relevance information (*i.e.*, whether the picture was of the actor or of an irrelevant person) at the end of each trial. Critically, this meant that participants processed the face and the behavior before knowing the relevance of the person in the photograph. The authors argued that, with this method, the proportion of attention paid to the stimuli during encoding may be similar in both the relevant and irrelevant conditions.

We maintain, however, that Goren and Todorov's (2009) data is also not sufficient to rule out an explanation based on attentional differences between STI and STT. We argue that participants may have adopted an "information processing postponing strategy" in which the main processing of information is delayed until all information was provided. The same strategy has been found to occur in the intentional forgetting literature where the cue to forget or remember a word is presented only after its disappearance (Basden, Basden, & Gargano, 1993). At this point, attention, albeit internal attention, could have still been differently engaged in the STI and STT. Internal attention is assumed to operate on internal information, for instance information that

is stored in working or long-term memory (Chun et al., 2011). Nevertheless, in this paper we focus on the contribution of external attention to STT and STI. In experiment 1, we employed a modified spatial cueing paradigm to investigate visual attention. In experiment 2 and 3, we employed the FRP to elicit STI and STT effects and, in order to measure visual attention, we monitored participants' eye-movements.

3.2 EXPERIMENT 1

The paradigm we apply in this experiment is based on the spatial cueing paradigm initially developed by Posner and colleagues (*e.g.*, Posner, Snyder, & Davidson, 1980). They presented three boxes on the screen, one in the center, one on the left side and one on the right side of the central box. In each trial, one of the peripheral boxes flickered briefly, followed by the presentation of the target that had to be detected as quickly as possible. The target (a #) was present either in the flickered box (*i.e.*, the valid condition) or in the box opposite to the flickered box (*i.e.*, the invalid condition). Participants were asked to press a key as quickly as possible once they detected the target. Posner and colleagues (1980) found that participants responded faster in the valid condition than in the invalid condition. There is a general agreement in the literature that the finding indicates that the cue (flickering) captures participants' attention whereas the slower RTs in the invalid condition is seen as evidence that participants first had to disengage their attention and re-orient it to the target (however, see Heinke & Humphreys, 2003, for an argument against the existence of such a disengagement mechanism).

Fox, Russo, Bowles, and Dutton (2001) applied this procedure to examine the influence of threatening stimuli on attention. They used threat words and angry faces as cues and found that these stimuli tend to increase response time ("dwell time") in the invalid condition compared to neutral stimuli. Response times in the valid condition were not affected. They interpreted these findings as evidence that emotional stimuli tend to "hold" attention rather than capture attention. It is also worth noting that Crawford and colleagues (2008) followed the same logic as Fox and collaborators (2001). Crawford et al., however, asked participants to perform an identification task (arrow up or arrow down) rather than a detection task. In our experiment we followed

Crawford and colleagues' (2008) procedure as it ensures that participants really focus on the task.

In each trial, we presented a photograph together with a behavioral description. The photograph could be of the actor performing the behavior (*i.e.*, the STI condition), or it could be of someone non-related to the behavior (*i.e.*, the STT condition). Participants were instructed to memorize the sentence and the photograph with which it is paired. The photograph is cueing the target (*i.e.*, an arrow); thus, as soon as the face disappeared, participants had to indicate the direction of the arrow. The arrow either appeared on the same side of the screen as the face (*i.e.*, the valid condition), or on the opposite side of the face (*i.e.*, the invalid condition).

Our hypotheses is fourfold. First, we expect a facilitated identification of the arrow in the valid condition because the face has already captured attention to that location. In addition, if the actor captures in fact more attention than the irrelevant person, in the valid trials we expect more facilitation in the STI condition than in the STT condition. On the other hand, we predict slower identification of the arrow in the invalid condition because attention is captured on the side of the screen opposite to the arrow. Finally, in the invalid trials we also predict slower identification of the arrow in the STI condition than in the STT condition. This is because we expect actor's faces to hold more the participant's attention than irrelevant faces. Note, however, that the instruction is to pay attention to and remember all the presented stimuli for a memory test later on in the experiment. In this experiment we measure the attention paid to the photographs and not the inference of the traits from behavioral descriptions.

3.2.1 *Method*

Participants

Sixty-five undergraduate students participated, receiving course credits. The sample size was defined by the number of show-ups in two weeks and the data were only analyzed when the reported sample was complete.

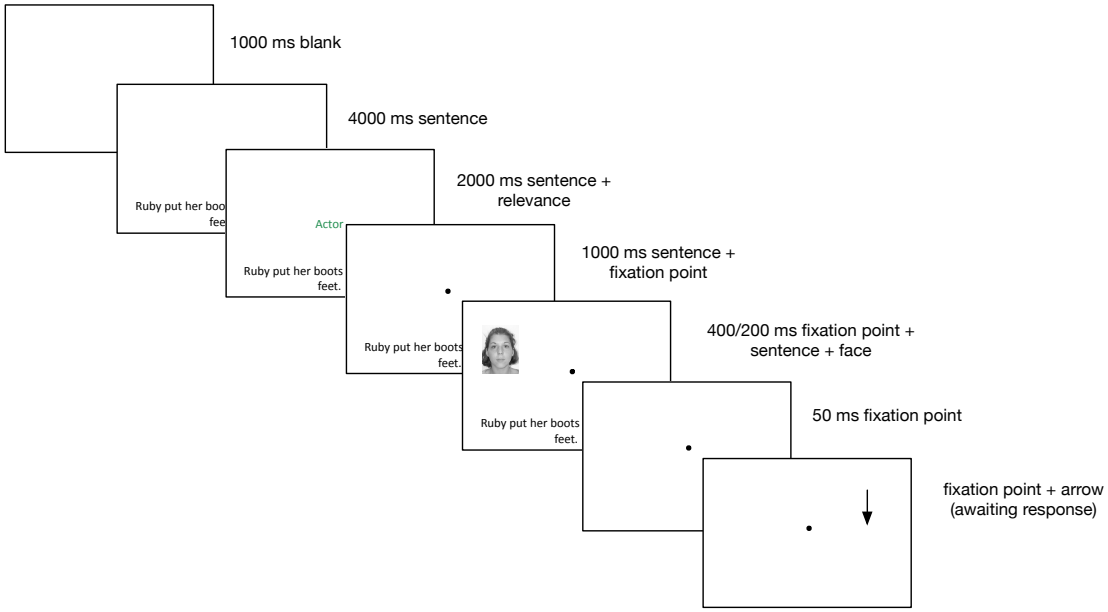

Figure 9: An example of an experimental trial, where the arrow is preceded by an invalid cue.

Material

60 trait-implying sentences and 60 photographs of individual faces (13 cm $\times$ 17cm) were used.

3.2.2 *Procedure*

The experiment had a 2 Face relevance (relevant versus irrelevant) $\times$ 2 Cue (valid versus invalid) $\times$ 2 Face presentation time (200 ms versus 400 ms) design, with the third factor being the only between-subjects factor. Time taken to react to the arrow was the dependent variable.

Participants were told that they would complete a social memory task, and were instructed to read and sign a consent form. First, participants completed a learning phase that started with a block of eight practice trials followed by a block of 60 experimental trials. Each trial in the learning phase started with 1000 ms blank screen followed by 4000 ms presentation of the sentence that participants were instructed to memorize for a test later on (see Figure 9). Next, before the photograph was presented, an on-screen prompt informed participants of the relationship between the behavioral sentence and the face in the soon-to-appear photograph. Participants were told that, when the word "Actor" appeared on the screen, the person in the photograph was the

person described in the corresponding sentence (*i.e.*, the relevant or STI condition). Additionally, participants were also told that, when the word "Random" appeared on the screen, the person in the photograph was randomly paired with the description (*i.e.*, the irrelevant or STT condition). The "Actor" or "Random" prompt, together with the sentence, remained on screen for 1000 ms. Following the disappearance of the prompt, a fixation dot appeared in the centre of the screen and remained there for 1000 ms. The fixation dot informed the participants that the photograph was about to appear.

Participants were informed about the short duration of the face (200 ms in one condition and 400 ms in the second condition). When the Stimulus Onset Asynchrony (SOA) is less than about 300 ms, findings show that reaction times are faster in the valid condition than in the invalid condition (Posner et al., 1980). However, when the SOA is greater than 500 ms, the effect is reversed and the reaction times are slower in the valid trials than in the invalid trials (inhibition of return - IOR; Posner & Cohen, 1984). To ensure that the interval for IOR is avoided, we included a condition where the SOA was less than 300 ms - 250 ms. A second condition with a 450 ms SOA was also included to ensure that there was enough time to process the face. In Fox and colleagues' (2001) studies, the stimuli were threatening faces whereas in our case they are neutral and thus less salient faces. Because of that, larger presentation time might be necessary in order to detect differences between STI and STT.

Following the 200 or 400 ms display of the face, the face and the sentence disappeared from the screen and a fixation dot appeared and remained on screen for 50 ms. Next, an arrow was displayed. The arrow pointed either up or down, and appeared on either the left side or the right side of the screen. Critically, in half of the trials the arrow appeared on the same side as the face (*i.e.*, the valid condition) whereas in the second half it appeared on the opposite side of the face (*i.e.*, the invalid condition). As soon as the participant pressed the response key (up or down arrow on the keyboard) a new trial started. Participants were instructed to indicate the direction of the arrow as quickly and accurately as possible.

Finally, participants completed two test phases. One regarding the faces and one regarding the sentences. The instruction was to indicate whether the photo/sentence were presented in the learning phase. 40 sentences and 40 photographs were used. Half of the sentences and photographs were old (*i.e.*, from the learning phase) and half were new (in the test concerning

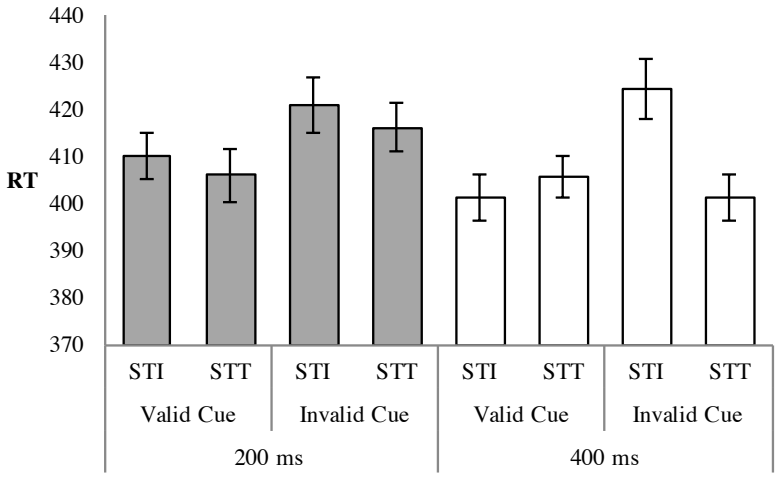

Figure 10: The average reaction time in function of relevance of the person in the photograph, the type of cue and the presentation time of the face in experiment 1. The bars on the graph are standard errors on the mean.

the sentences, the new sentences were modified versions of the old ones). Participants used the keyboard to answer (the S key for "yes" and the L key for "no"). The purpose of the test phase was to ensure that participants attended the material. We did not expect a good memory for faces because of the short presentation time.

3.2.3 *Results and Discussion*

The data from the test phase was used to verify if the participants paid attention to the sentences. It is crucial to analyze this data since, participants were not asked to do anything during the presentation of the sentences (except to memorize the stimuli). Six participants were eliminated because they had 40% or more errors in the memory test for sentences. The average error rate for the arrow identification task was 8% and all the errors were omitted from the following analysis. We also omitted data from a participant because he presented more that 40% errors in the arrow task. Finally, we eliminated all the responses that took more than 3 standard deviations or less than 200 ms (this constituted 1.7% of the answers).

Next, a 2 Face (relevant vs irrelevant) $\times$ 2 Cue (valid vs invalid) $\times$ 2 Face presentation time (200 ms vs 400 ms) Analysis of Variance (ANOVA) was conducted, with the third factor being the

only between-subjects factor. The dependent variable was the reaction time (RT). The analysis revealed a main effect of cue, $F(1,56) = 13.42$, $p = .001$, $\eta^2 = 0.03$, a main effect of relevance, $F(1,56) = 14.63$, $p < .001$, $\eta^2 = 0.01$, an interaction between the cue and the relevance, $F(1,56) = 7.90$, $p = .007$, $\eta^2 = 0.01$, and a three-way interaction, $F(1,56) = 6.83$, $p = .01$, $\eta^2 = 0.01$. No other effects were significant.

To disentangle the three-way interaction, we conducted two simple effect analyses, one for the 200 ms condition and one for the 400 ms condition. For the 200 ms condition we found a marginal main effect of relevance, $F(1,27) = 3.68$, $p = .07$, $\eta^2 = 0.02$, and a main effect of cue, $F(1,27) = 4.83$, $p = .04$, $\eta^2 = 0.10$, however, no interaction, $F(1,27) = .03$, $p = .87$, $\eta^2 = 0.00$ was detected. When the face was presented for 400 ms, a main effect of cue was found, $F(1,27) = 9.35$, $p = .005$, $\eta^2 = 0.17$, an interaction between relevance and cue, $F(1,27) = 7.78$, $p = .009$, $\eta^2 = 0.02$, but no main effect of relevance was found, $F(1,27) = .44$, $p = .51$, $\eta^2 = 0.00$.

Critically for our hypothesis (see Figure 10), with a 200 ms presentation of the photo no difference was observed between STI trials ($M = 410$; $SD = 143$) and STT trials ($M = 406$; $SD = 159$) when the photo was a valid cue, $t(27) = 1.02$, $p = .32$, $d = .14$. We also observed no difference between STI trials ($M = 421$; $SD = 169$) and STT trials ($M = 416$; $SD = 148$) when the photo was a invalid cue, $t(27) = 1.26$, $p = .22$, $d = .17$. When the presentation time of the face was 400 ms, there were no differences between the reaction time to STI trials ($M = 401$; $SD = 138$) in comparison to STT trials in the valid cue condition ($M = 405$, $SD = 122$), $t(27) = -1.34$, $p = .19$, $d = -.16$. However, when the invalid trials were compared, the RT for the STI condition was larger ($M = 424$; $SD = 185$) that the RT for the STT trials ($M = 401$; $SD = 138$), $t(27) = 3.87$, $p = .001$, $d = .75$.

Overall, we found that participants were faster to detect the target when it was presented at a location previously occupied by a face in comparison with an empty location (*i.e.*, cueing effect). Moreover, in support of our attentional hypothesis, when the photo is presented during 400 ms, the cueing effect was larger for STI than for STT. Interestingly, this increase is due to longer reaction times in the invalid condition rather than the valid condition. In other words, the relevance of the faces leads to holding attention rather than capturing attention. Furthermore, and seemingly contradictory with our attentional hypothesis, we failed to find an influence of

relevance for the 200 ms presentation time. We will discuss this apparent inconsistency in the discussion section.

3.3 EXPERIMENT 2

The second experiment explored the attention hypothesis with a different paradigm. Eye-tracking may be a better measure of visual attention because it provides a continuous, online measure of visual attention during the encoding of information (*e.g.*, Hermans, Vansteenwegen, & Eelen, 1999; Anderson, Heinke, & Humphreys, 2013) whereas, in the spatial cueing paradigm, attention is measured at the end of the presentation time of the cue. The main purpose of this experiment was to investigate whether the proportion of visual attention paid to faces is larger in the STI condition than in the STT condition by recording eye movements during the learning phase of the FRP (see the description of the paradigm below). Furthermore, this experiment represents the first attempt to replicate the differential effect between STI and STT found by Goren and Todorov (2009) with the FRP. This is important because their attempt to replicate the effect in their paper wasn't very convincing, since the difference between STI and STT did not reach significance (see their experiment 2).

The main advantage of the FRP is that it explores the link between the person and the trait in an elegant way (Todorov & Uleman, 2002, 2003, 2004). The first stage of the FRP involves a learning phase in which participants are told to memorize pairs of faces and sentences; and each sentence describes a unique behavior implying of a personality trait (*e.g.*, "Mary helped an elderly person cross a busy road" which implies "helpfulness"). Importantly, participants are informed that the face corresponds either to the actor of the corresponding behavioral description (*i.e.*, the relevant or STI condition), or to randomly paired stimuli (*i.e.*, the irrelevant or STT condition). Thus, the FRP operationalizes the typical situations of STI and STT. Importantly, the random pairing in the STT condition excludes any possibility of inferring a relationship between an actor and the irrelevant person.

The second stage of the FRP involves a test phase in which participants are presented with pairs of faces and personality traits (*e.g.*, helpful) and is instructed to indicate whether the word (trait) appeared previously in the sentence paired with that particular face. Believing that the

trait was part of the sentence (when in fact is wasn't) constitutes an error, and indicates that the trait was inferred from the sentence and is misrecognized as being presented in the sentence (*i.e.*, false recognition). In one half of the trials, the pairs of faces and traits correspond to the pairs of faces and behavioral sentences learned previously (for example Mary's face paired with "helpful" *i.e.*, matched face-trait pairs). These matched face-trait pairs corresponded to either relevant faces from the learning phase (*i.e.*, the relevant-match condition) or to irrelevant faces (*i.e.*, the irrelevant-match condition). In the remaining half of the trials, participants are presented with pairs of faces and traits that do not correspond to the pairs of faces and behavioral sentences learned before. For example, when another person's face, from the learning phase, is paired with the trait "helpful" instead of Mary's face (*i.e.*, mismatched face-trait pairs).

In order to verify the linkage created between person and trait, a comparison is conducted between the proportion of errors in match and mismatch conditions. If the trait is simply inferred from the behavior without being associated with the person, then the rate of errors in the match and mismatch conditions should be comparable. If, on the other hand, a link is created between the person and the trait during the learning phase, then there should be a greater proportion of false recognitions in the match condition than the mismatch condition. Indeed, past research has found a greater proportion of false recognitions in the match condition than in the mismatch condition, and, in particular, this difference (match minus mismatch) is greater in STI that it is in STT.

These findings suggest a stronger linkage between the face and the trait in STI than in STT (Goren & Todorov, 2009). In the present experiment, we recorded eye - movements with an eye-tracking device during the learning phase. This provided a continuous indicator of visual attention during the encoding of STI and STT trials. Our predictions were fourfold. The first two predictions concern the replication of Goren and Todorov's (2009) findings. We predicted that there would be a greater proportion of false recognitions (*i.e.*, error responses) in the STI in comparison to the STT and also that the match minus mismatch difference will be larger in the STI condition than in the STT. The following two predictions were derived from our attentional hypothesis. We predicted that there would be a greater proportion of visual fixations on the relevant face in comparison to the irrelevant face in the learning phase. Finally, we predicted that

there would be no difference in the proportion of visual attention paid to sentences in the STI condition in comparison to the STT condition.

3.3.1 *Method*

Participants

Thirty-two undergraduate students participated in this study, receiving course credits or 6 pounds. Data of 9 participants were excluded from the statistical analysis of the eye-movements due to invalid data, leaving a final sample of 23. All 32 participants were included in the analysis of the data in the false recognition task. The selection of the sample size was determined based on the sample size reported by Goren and Todorov in their studies. Again the data were analyzed only when the collection ended.

Material

60 sentences were used, all of them being behavioral descriptions that implied personality traits. 20 descriptions were filler sentences, whereby the traits were actually included in these sentences (*e.g.*, "Thrifty Adam only buys shoes and clothes on sales"). In the remaining 40 descriptions the trait was only implied in the sentence (*e.g.*, "Johnny told the cashier that he got too much change" implied the trait "honest"). We are following the 2/3 proportion of trials with absent targets/traits (trials where the correct response would be "no") as suggested by McKoon and Ratcliff (1986). In addition, 60 photographs of individual faces with neutral expressions (13 cm $\times$ 17 cm) were used.

3.3.2 *Procedure*

The experiment had a 2 Relevance (relevant versus irrelevant) $\times$ 2 Pairing (match versus mismatch) within-subjects design. False recognition rate (*i.e.*, error responses) and the proportion of visual fixations on the face and behaviors were the main dependent variables.

Participants were taken individually to the laboratory containing a computer and a head-mounted eye-tracker. Participants were told that they would complete a memory task, and were asked to read and sign the consent form. The eye-tracking device was placed on the participant's head in order to track the movements of the pupil in the left eye. Participants then completed a nine-point grid calibration. The eye-tracking device used was an Eye-Link II; eye-movements were recorded at a rate of 1 kHz, and with an average accuracy of 0.5 degrees.

Once the eye-tracking device was placed on the participant's head and calibrated, participants began the task on the computer. First, they completed a learning phase, during which a short block of four practice trials was completed before starting a longer block of 60 experimental trials (lasting a total of 8 minutes). Participants were presented with single pairs of faces and behavioral sentences (*i.e.*, face-behavioral pairs) and were instructed to memorize each pair for a non-specified memory test later on. Photographs were presented above the corresponding sentence. Face-behavioral pairs were presented one-at-a-time for 6 seconds, with a 2 second interval between trials. Importantly, participants were informed that the person in the photograph was the actor of the corresponding behavior when the photograph was presented with a blue frame (*i.e.*, the relevant condition), and that the person and the behavior were randomly paired by the computer when the photograph was presented with a red frame (*i.e.*, the irrelevant condition). Participants were instructed to memorize the stimuli regardless of their relevance. Finally, the presentation of the pairs of faces and sentences were randomized for each participant.

After completing the learning phase, participants were instructed to take a 5 minutes break in which they completed a Sudoku puzzle (*i.e.*, a distraction task). After the distraction task, participants completed the test phase in which they were presented with single pairs of faces and traits (*i.e.*, face-trait pairs). The faces were all presented in the learning phase, and the traits were implied or presented in the sentences seen in the learning phase (*e.g.*, "honest"). Participants were instructed to press the "C" key if they believed that the trait was part of the sentence presented in the learning phase with that photo, and the "M" key if they did not. In addition, participants were instructed to respond as quickly and accurately as possible. Face-trait pairs were presented one-at-a-time, and remained on screen until participants responded. Importantly, photographs were presented without frames this time and the presentation was randomized for each participant.

Participants were presented with 60 face-trait pairs, involving four different types of pairs. The first group consisted of 10 trials corresponding to pairs of a relevant face and the trait implied in the sentence paired with that photo in the learning phase. Answering "yes" to these trials constituted an error response, since the trait is falsely recognized as being present in the sentence when it was just inferred and linked to the actor. In the second group, 10 trials corresponding to pairs of irrelevant faces and traits implied in the sentence from the learning phase were presented. As mentioned before, these 20 pairs (10 relevant plus 10 irrelevant) are called match trials because the trait presented in the test phase was implied in the sentence presented in the learning phase with that person. The opposite was true for other 20 trials in which participants were presented with photos paired with traits that were implied in sentences that in the learning phase accompanied different faces (mismatch trials). Thus, in the relevant-mismatch condition we paired the faces presented in the STI trials in the learning phase with traits implied in sentences presented in the STT trials in the learning phase. In the irrelevant-mismatch condition the faces from the STT condition were paired with traits implied in sentences from the STI condition. Thus, we avoided the face being presented with its correspondent trait. The two kinds of trials (match and mismatch) were compared in order to verify whether a link was created between the person and the trait. The link is expected to be present in the match condition and absent in the mismatch condition as the mismatch consists of a completely new pairing. Participants were presented with additional 20 face-trait pairs, which corresponded to the fillers trials from the learning phase (the traits in these trials were actually part of the sentences). Answering "yes" to these trials represents a correct response. Importantly, the four types of face-trait pairs were presented in a randomized order.

3.3.3 *Results and Discussion*

False recognition rate

The filler trials in which the person in the photo is the actor were used to clean the data. Note that for these trials the person in the photo is the actor and the trait is actually part of the sentence, so both the inference and the actual recall will lead to the same “yes” response if the participant

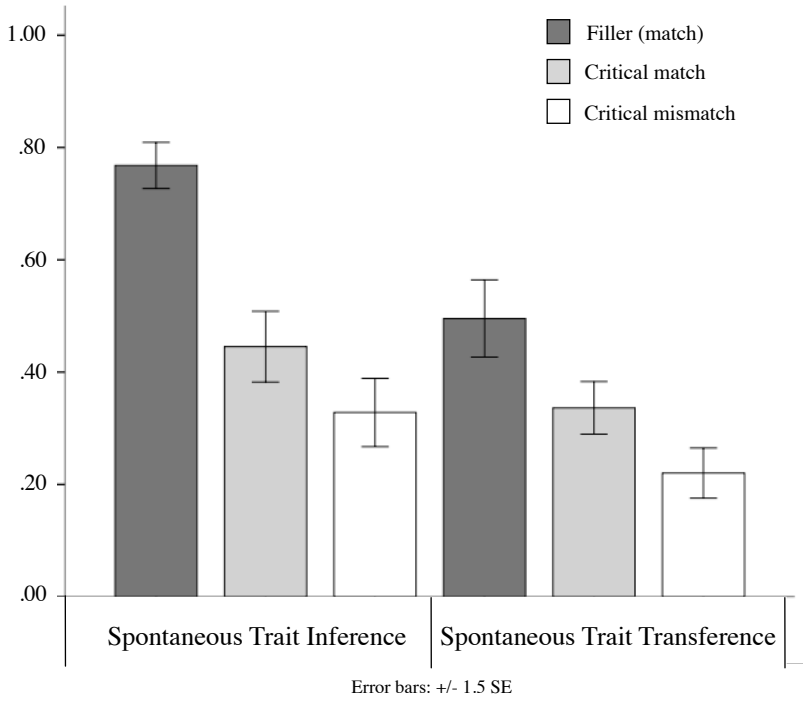

Figure 11: The proportion of "yes" responses in function of type of trial, relevance and type of pairing in experiment 2.

is motivated to process the material. Thus, 10 participants were eliminated based on the criteria of having 50% or less correct responses in these trials. Also note, that except for data cleaning, these trials were not important for the analysis, only the trials corresponding to the trait-implying sentences are.

In order to test the prediction that there would be a larger rate of false recognitions in the relevant condition in comparison to the irrelevant condition, a 2 Relevance (relevant vs irrelevant) $\times$ 2 Pairing (match vs mismatch) repeated-measures ANOVA was conducted, with the rate of false recognition as the dependent variable. A main effect of face relevance was observed, $F(1,21) = 14.42$, $p = .001$, $\eta^2 = .41$, meaning there are more false recognitions in the relevant condition than in the irrelevant one, and a main effect of pairing was detected, $F(1,21) = 11.30$, $p = .002$, $\eta^2 = .36$, meaning people not only infer the trait from the sentence, they also link it to the person presented. The interaction between face relevance and the pairing was not significant, $F < 1$.

As figure 11 shows the mean error responses in the relevant-matched condition ($M = .46$, $SD = .20$) were relatively greater than in the relevant-mismatched condition ($M = .33$; $SD =$

.19), $F(1,21) = 6.46$, $p = .019$, $\eta^2 = .24$. Similarly, the mean error responses in the irrelevant-matched condition ($M = .34$; $SD = .15$) were also greater than in the irrelevant-mismatched condition ($M = .22$; $SD = .14$), $F(1,21) = 7.85$, $p = .011$, $\eta^2 = .27$. These results show that both STI and STT effects are observed, but their effect sizes do not vary from each other. Thus, we did not replicate the past finding showing stronger STI than STT effects (Goren & Todorov, 2009).

Fixations on the faces and text

In this analysis we included only the participants with valid eye-tracking recordings. 11 participants were removed because the eye tracker was not well-calibrated and, consequently, we failed to monitor their eye movements accurately. Thus, only the data of 21 participants from the original sample were used in the analysis of eye-movements. To analyze visual fixations, the screen was split up into two areas - the "face" area and the "text" area. Next, the number of fixations that fell into each area was counted, and those constitute our dependent variables in this analysis.

In order to test the prediction that there would be more fixations on the relevant faces than on the irrelevant faces, we compared the proportion of fixations on the relevant faces in the learning phase with the proportion on the irrelevant faces in the learning phase. A repeated measures ANOVA was conducted, where the two within-subject factors were the relevance (actor or irrelevant person) and the area were the fixations fell (text vs face). A main interaction was found between the relevance and the area, $F(1,20) = 3.99$, $p = .059$, $\eta^2 = .17$. In support of our prediction, there were more fixations on the actor's face ($M = 10.82$; $SD = 2.81$) than on the irrelevant person's face ($M = 10.01$; $SD = 3.30$), $F(1,20) = 6.39$, $p = .020$, $\eta^2 = .24$. While the difference between the number of fixations on the text in the STI condition ($M = 9.91$; $SD = 4.06$) did not reach significance when compared with the text fixations in STT condition ($M = 10.49$; $SD = 4.26$), $F(1,20) = 1.94$, $p = .179$, $\eta^2 = .09$. This result corroborates our hypothesis that there is an attentional difference between the processing of the actor and the irrelevant person.

In summary, we observed the expected larger number of fixations on the actor's face in comparison to the irrelevant face. We also observed a similar number of fixations on the behavioral

sentence in the relevant and irrelevant conditions. However, we did not replicate the stronger face-trait linkage in the STI condition than in the STT condition. The differences are restricted to the overall proportion of false recognitions that was higher in STI than in STT.

3.4 EXPERIMENT 3

The purpose of the third experiment was to replicate the attentional result from the second experiment, while aiming at replicating the difference between STI effect and STT effect usually found in the literature (*e.g.*, Goren & Todorov, 2009). In this experiment we used a more realistic manipulation of the relevance, instead of telling participants the truth about the pairing of the sentences and the photos in the STT condition (that it is random), we told them that the person in the photo was the communicator of the behavior that someone else enacted.

In this study we tried to collect more data, so a correlation analysis can be performed in order to understand the relationship between the amount of attention paid to the material and the STI and STT effects.

3.4.1 *Method*

Participants

42 participants took part in this study, 14 were males and their average age was 24.57 years old. They were students at Birmingham University. Each of them got a compensation of 4 pounds. In this experiment our goal was to collect more participants than in the previous experiment. What determined the stopping point in the collection was the end of the semester.

Material

The same material as the one used in the second experiment was used here. The only exception is that the relevance manipulation was also evident in the sentence. For the relevant condition, where the person in the photo was the actor, besides the color of the frame, the sentences were written in the first person (*e.g.*, "I ran a marathon"). While for the STT trials, the sentences were

written in the third person ("*e.g.*, "He ran the marathon") and the participants were instructed that the person in the photo was providing the information that sometimes was about themselves and other times was about a third party (*i.e.*, the person in the photo is a communicator).

3.4.2 *Procedure*

The experimental design of this study is: 2 Relevance (relevant versus irrelevant) $\times$ 2 Trial type (experimental/critical versus fillers) $\times$ 2 Pairing (match versus mismatch) design, with all the factors being within-subjects. The rate of false recognitions and the number of fixations are the dependent variables.

The eye-tracking device used to measure the number of fixations was Eye-Link 1000, with 1000 Hz rate and 0.5 degrees average accuracy. We monitored the eye-movements during the whole experiment, but only the fixations in the learning phase were analyzed.

In the learning phase participants were presented with behavioral descriptions and photos. In 20 of the trials the trait was part of the sentences (filler trials) and in the remaining 40 trials the trait was implied in the sentence. Half of the trials corresponded to STI trials and the other half to STT trials. After memorizing the trials, participants performed a short distraction task (spot the difference puzzle). In the test phase, just as in the second experiment, participants had to indicate if a word presented with a photo, was or not part of the sentence learned together with that photo. The types of trials are similar to experiment 2, except for the fact that the pairing manipulation was also applied, this time to the filler trials, such that we ended up with 10 match and 10 mismatch fillers (5 regarding the actor and 5 regarding the communicator).

3.4.3 *Results and Discussion*

10 participants were eliminated from the sample because their accuracy for the filler trials (relevant match condition) was 50% or less.

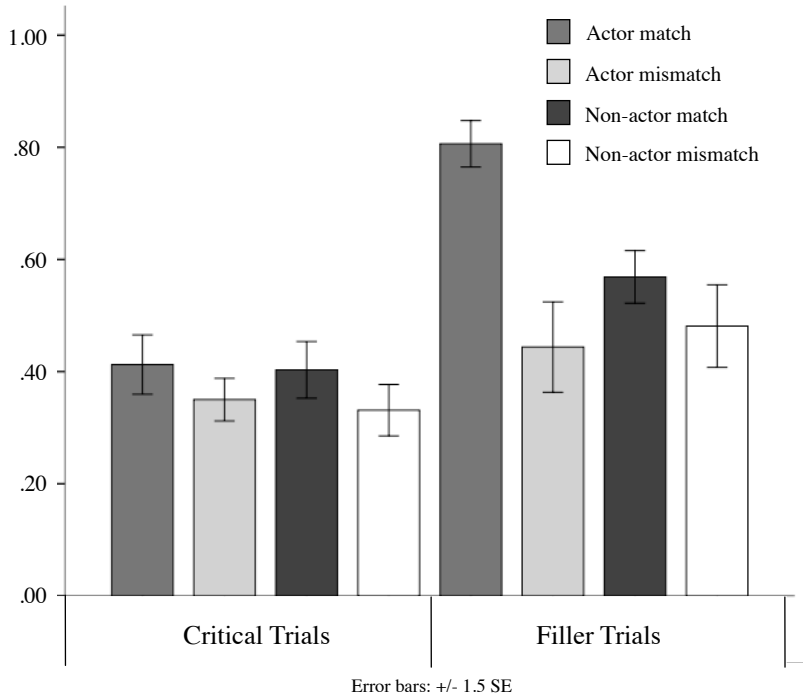

Figure 12: Proportion of false recognitions in function of type of trial (critical, fillers), relevance (STI, STT), and pairing (match, mismatch) in experiment 3.

False recognition rate

A separate analysis was conducted for the filler and the critical trials.

For the critical trials, a main effect of pairing was found, $F(1,31) = 4.42$, $p = .044$, $\eta^2 = .13$, with larger false recognition rate for the match trials ($M = .41$; $SD = .15$) than for the mismatch ones ($M = .34$; $SD = .13$). However, again, and as shown in figure 12, no interaction was detected between the relevance and the pairing. When the same analysis is performed for the filler trials, a main effect of relevance is detected, $F(1,31) = 7.83$, $p = .009$, $\eta^2 = .20$, a main effect of pairing, $F(1,31) = 24.14$, $p < .001$, $\eta^2 = .40$, and an interaction, $F(1,31) = 12.22$, $p = .001$, $\eta^2 = .28$. This interaction is detected due to the larger difference between the match ($M = .81$; $SD = .17$) and the mismatch conditions ($M = .44$; $SD = .30$) for the relevant person, $F(1,31) = 37.51$, $p < .001$, $\eta^2 = .55$, than the difference between match ($M = .57$; $SD = .18$) and mismatch conditions ($M = .48$; $SD = .28$) for the irrelevant person, $F(1,31) = 2.02$, $p = .165$, $\eta^2 = .06$.

Fixations on the faces and text

Next, we conducted an ANOVA, with the dependent variable being the proportion of fixations in the critical trials and the independent variables being the area where the fixations fell and the relevance. A main effect of area was found, $F(1,30) = 7.49$, $p = .010$, $\eta^2 = .20$, as well as an interaction between the factors, $F(1,30) = 19.17$, $p < .001$, $\eta^2 = .29$. As expected, and replicating the results found in the previous experiment, more fixations were observed on the face of the actor ($M = 12.82$; $SD = 4.00$) than on the irrelevant person's face ($M = 11.98$; $SD = 3.71$), $F(1,30) = 13.23$, $p = .001$, $\eta^2 = .31$. The opposite pattern was found for the fixations on the text, $F(1,30) = 21.05$, $p < .001$, $\eta^2 = .41$, with more fixations detected on the sentence in the STT trials ($M = 10.50$; $SD = 4.41$) than in the STI trials ($M = 9.46$; $SD = 3.88$). While the actors' faces are more attended to than irrelevant people's faces, the behavioral description received more fixations when the person in the photo was irrelevant, which might be merely a consequence of less attendance to the face. Note however that, again, this attentional differences were not translated into differences in the STI and STT effects.

Correlations among response rate and the number of fixations

Next, we present the correlations between the fixations and the STI and the STT effects. For both, the STI and the STT conditions, we calculated the difference between match and mismatch trials. Next the correlations between this difference and the amount of fixations on the face and on the text were calculated. Additionally, the number of transitions between the face and the text were calculated as a proxy to the linkage between the person and the sentence and the information implied in it. In the STI condition, we found a strong positive correlation between the amount of fixations on the face and the STI effect (the difference between the match and the mismatch), $r(31) = .46$, $p = .009$, while the fixations on the text were not correlated with the STI effect, $r(31) = -.06$, $p = .767$. The number of transitions between the face and the text also positively correlated with the STI effect, $r(27) = .42$, $p = .028$. There was no correlation between the number of fixations on the face and the STT effect, $r(31) = .01$, $p = .974$, no correlation between the fixations in the text and the effect, $r(31) = -.05$, $p = .778$, and also no correlation between the transitions and the effect, $r(31) = .09$, $p = .648$. These results suggest

that the attention might have a role in the STI effect, such that more attention means a stronger STI effect. The way the participants monitor the link between the trait and the face might be via attention allocation. Whereas in the STT condition there is no relationship between the amount of attention paid to the face and the effect.

3.5 GENERAL DISCUSSION

The main goal of the present research was to show that there are differences between STI and STT that can be attributed to attention. In particular we wanted to show that the relevance of the person presented with the behavioral description influences the amount of attention paid to that person. In the first experiment we applied a spatial cueing paradigm. The sentence and the information prompting the relevance of the face (*i.e.*, "Actor" or "Random") were given at the beginning of the trial, and, afterwards, the faces were briefly added at either sides of the screen towards the end of the trial. Following this, participants had to detect an arrow and indicate whether it pointed upwards or downwards. The arrow was either spatially aligned with the face (*i.e.*, the valid condition) or on the opposite side of the screen (*i.e.*, the invalid condition). This part of the experiment was analogous to the spatial cueing paradigm with the faces functioning as cues to a target.

The main result of the first experiment was a slower response to identify the target when the face was flagged up as an actor compared to the cases when the face represented an irrelevant person. However, and importantly, this differential effect occurred only in the invalid condition and not in the valid condition. This effect is commonly interpreted as meaning the cue is not capturing more attention, but holding more the attention (*i.e.*, it is more difficult to disengage the attention from the stimulus; *e.g.*, Fox et al., 2001).

The evidence from experiment 1 is also consistent with the findings in experiment 2 and 3. In the second study we used the False Recognition Paradigm, whereby participants were presented with pairs of photographs and behavioral descriptions. One half of the photographs were said to be depicting the actor of the behavior presented, and the other half of the photographs were said to be of people randomly paired with the behavior by the computer. While the participants were

learning this information their eye-movements were monitored with an eye tracker. We found that people fixate more on the actor than on the irrelevant person.

In the third experiment we replicated the data from the second experiment, but this time in the STT condition the irrelevant person was said to be the communicator of someone else's behavior, a more plausible and realistic manipulation than the random pairings. The reason for this modification was to avoid the disregard of the irrelevant person just because it was associated with the "Random" label. In this study we replicated the previous finding, that is, more attention was paid to the actor's face than to the communicator's face. Moreover, we found a positive correlation between the number of fixations and the STI effect and no such correlation was found for the STT effect.

Hence, the three experiments indicate that engagement of attention is stronger in the STI condition than in the STT condition (*i.e.*, attention dwells more on the face of the actor than on the face of an irrelevant person).

Interestingly, an interaction between cueing (*i.e.*, the reaction times difference between valid and invalid condition) and relevance was found only when the photo was presented for 400 ms and not when it was presented for 200 ms. On the face of it, this seems to contradict our findings and to weaken the support for our attentional hypothesis. Nevertheless, this differential effect is highly consistent with well-established findings in the spatial cueing literature and, in fact, can give further credence to our attentional hypothesis. Typically, spatial cueing experiments utilize two types of cues: a peripheral and a central cue. In both designs, the response target appears peripherally. In experiments with the peripheral cue the target can be either aligned with the cue (*i.e.*, the valid condition) or appear on the opposite side of the cue (*i.e.*, the invalid condition). The central cue is typically an arrow placed at the centre of the screen which can point either to the left or to the right. Participants are informed that the direction of the arrow indicates where the target is most likely to appear (even though there is typically no relationship between arrow orientation and target location). For both types of cues, participants show faster reaction times in the valid condition compared to the invalid condition. However, the size of the cueing effect depends on the time interval between the appearance of the cue and the target (*i.e.*, Stimulus Onset Asynchronicity; SOA). The peripheral cue is most effective at around 100 ms. In contrast, the effect of the central cue is small at 100 ms but then builds up and peaks at around

300 ms (*e.g.*, Cheal & Lyon, 1991; Nakayama & Mackeben, 1989). Shifts of attention due to the central cue are typically considered to be under "voluntary", "conscious" or "top-down" control whereas the peripheral cue is thought to cause involuntary shifts of attention. Hence, our results are highly consistent with these findings. The fact that we failed to find a cueing effect at short SOA is likely due to the fact that the instruction about the relevance of the faces directs attention in a similar vein as the central cue directs attention. In the context of our experiment, the faces - irrespective of whether they were relevant or irrelevant - constitute a peripheral cue (*i.e.*, perceptual stimulation at either side of the screen). Thus, we might expect a cueing effect at short SOA, and possibly at long SOA. This result allows us to verify whether participants direct their attention to faces at all. Furthermore, the relevance of the face was introduced to participants by written instructions. This is similar to the way in which the meaning of the central cue is established. In both cases simple perceptual stimulation (as with the face or the peripheral cue) is not enough to influence attention. Instead participants need to translate instructions into the manipulation of attention (the case of the relevance of the face is the central conjecture of this paper). Hence, we think it is plausible that relevance is modulating the cueing effect at later SOA rather than at early SOA (given our attention hypothesis); thus, this adds further credence to our hypothesis that the relevance of faces affects attention.

Furthermore, we should stress that the present results do not suggest that the influence of attention is the sole reason for the difference between STI and STT. In other words, the results don't rule out the dual process explanation or the single process explanation. However, our results suggest that future experiments need to make an effort to keep participants' attention balanced between the STI and the STT.

Our results, however, are not consistent with past literature that finds no difference in the way in which the actor and the irrelevant person are attended to (Crawford et al., 2008). However, there are some major differences between our methodology and Crawford and colleagues' methodology. First, Crawford and colleagues' presentation of two faces alongside behavioral descriptions was remarkably more complex than the traditional presentation of a one face per behavioral description. In their study, in half of the trials, participants were told that an actor presented in one photograph was describing his own behavior to a bystander presented in the other photograph (the actor is the speaker, *e.g.*, "I ran a marathon"). Participants may associate the trait, for ex-

ample “athletic”, to the actor (*i.e.*, a STI effect) or to the bystander (*i.e.*, a STT effect). In the other half of the trials, participants were told that an informant presented in one photograph was describing the behavior of a target presented in the other photograph (the speaker is not the actor, *e.g.*, “He ran a marathon”). Participants may associate the trait to the informant (*i.e.*, a STT effect) or to the target (*i.e.*, a STI effect). By presenting a STI and STT condition simultaneously, Crawford and collaborators compared the proportion of visual attention allocated to each condition. They predicted that, if visual attention was involved in the stronger linkage between face and trait in STI than in STT, then a greater proportion of visual attention would be paid to actors and targets than to informants and bystanders. Findings showed that the average proportion of visual fixations on the actor was greater than on the bystander, indicating that the actor received a greater proportion of visual attention than the non-relevant person. In contrast, the proportion of visual fixations on the informant and target was comparable. The limitation of Crawford and colleagues' design is that, in each trial, participants had to first identify which of the two people was speaking, and next, whether the speaker was referring to himself or to another person (*i.e.*, an actor describing himself or an informant describing the target). We decided, in our study to eliminate the speaker confound. In addition, Crawford and colleagues claim that the inclusion of a STI condition (*e.g.*, a target) alongside a STT condition (*e.g.*, an informant) may reduce or eliminate the STT effect; therefore, findings may not reflect a true comparison between STI and STT. In this Chapter we explored the attention paid to the face when only one person is presented. Moreover, the probe detection latencies employed by Crawford and collaboratos may reflect a snapshot of visual attention after the encoding of information, as opposed to during (*e.g.*, Koster, Crombez, Van Damme, Verschuere, & De Houwer, 2004), especially because the presentation time of the cues (*i.e.*, faces) was 12 s (in contrast with our 200 and 400 ms). This is much longer than the time recommended by the inhibition of return literature. Moreover, these results are based on null effects. Overall, Crawford and colleagues' (2008) findings in the eye-tracking experiment may have been weakened by the complexity of the paradigm. The findings in the probe detection experiment may be affected by the complexity of the paradigm, but also by the drawbacks of the probe detection task combined with the presentation time of the cue.

Another purpose of our second and third experiments was to replicate the finding that the effect of STI is stronger than STT by using the false recognition paradigm (Goren & Todorov,

2009). However, we failed to replicate this difference. A possible reason is that the size of the interaction effect in this paradigm is small, meaning that the paradigm is not sensitive enough to detect the difference between STI and STT. Furthermore, interactions tend to require larger power than main-effects, which might be the most plausible reason for our lack of replication. However, there could be at least one more reason for this result. In our experiment, we altered the color of the photograph frame in order to indicate the relevance of the face. In contrast, Goren and Todorov indicated the relevance of the face by writing the corresponding sentences in different colors. Why should this make such a difference? We think it is conceivable that, in Goren and Todorov's design, participants' attention to sentences could have been moderated by the relevance of the sentences. Consequently, participants may pay less attention to the irrelevant sentences and would infer fewer traits from the behaviors. Hence, the Goren and Todorov's design may not have manipulated the linkage between traits and faces but, rather, the inference of the trait itself (via attention to the sentence). The reduction in the inference of the trait itself (resulting from manipulating the relevance of the sentence) might be leading to stronger effects (stronger STI – STT differences) than the manipulation of relevance of the faces. There are less traits to be linked with faces in the first case in the STT condition. In other words, our manipulation of relevance can be seen as weaker than Goren and Todorov's. However, and importantly, our design can be considered a more veridical implementation of the original research question, as the inference mechanism (*i.e.*, the comprehension of the sentence) is not affected by the instruction. Instead, only the linkage between trait and faces is being manipulated.

Together, these three experiments suggest a very basic cognitive difference between STI and STT. Comparing a situation in which an actor is communicating its own behavior with a situation in which a person is talking about someone's behavior, we pay more attention to the first person (*i.e.*, the actor) than to the second (*i.e.*, the informant). Thus, we recommend that this attentional disparity is taken into account in future studies that aim to explore the differences between STI and STT.

3.6 REFERENCES

Anderson, G. M., Heinke, D., & Humphreys, G. W. (2013). Top-down guidance of eye move-

ments in conjunction search. *Vision Research*, *79*, 36–46.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 603–616.

Bassili, J. N., & Smith, M. C. (1986). On the spontaneity of trait attribution: Converging evidence for the role of cognitive strategy. *Journal of Personality and Social Psychology*, *50*(2), 239–245.

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.

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.

Cheal, M., & Lyon, D. R. (1991). Central and peripheral precuing of forced-choice discrimination. *The Quarterly Journal of Experimental Psychology*, *43*(4), 859–880.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*(3), 466–472.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Leonards, U. (2008). Seeing, but not thinking: Limiting the spread of spontaneous trait transference ii. *Journal of Experimental Social Psychology*, *44*(3), 840–847.

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.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General*, *130*(4), 681–700.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associa-

tions with faces. *Social Cognition*, *27*(2), 222–248.

Heinke, D., & Humphreys, G. W. (2003). Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (saim). *Psychological Review*, *110*(1), 29–87.

Hermans, D., Vansteenwegen, D., & Eelen, P. (1999). Eye movement registration as a continuous index of attention deployment: Data from a group of spider anxious students. *Cognition and Emotion*, *13*(4), 419–434.

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research*, *45*(13), 1707–1724.

Koster, E. H., Crombez, G., Van Damme, S., Verschuere, B., & De Houwer, J. (2004). Does imminent threat capture and hold attention? *Emotion*, *4*(3), 312–317.

Lamme, V. A. (2003). Why visual attention and awareness are different. *Trends in Cognitive Sciences*, *7*(1), 12–18.

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, *95*(1), 15–20.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, memory, and cognition*, *12*(1), 82–91.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research*, *29*(11), 1631–1647.

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*(42), 16598–16603.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522.

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition*, *33*(1), 20-66.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. *Attention and performance: Control of language processes*, *32*, 531–556.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals.

Journal of Experimental Psychology: General, *109*(2), 160–174.

Serences, J. T., Shomstein, S., Leber, A. B., Golay, X., Egeth, H. E., & Yantis, S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. *Psychological Science*, *16*(2), 114–122.

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.

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.

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.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology*, *28*, 211–279.

4

INFORMATION PROCESSING IN SPONTANEOUS TRAIT INFERENCE AND TRANSFERENCE

Diana Orghian, Tânia Ramos, and Leonel Garcia-Marques

In preparation

In this Chapter we continue investigating the difference between STI and STT. More precisely, we test the hypothesis that the difference between STI and STT is not due to different processing of the behavioral description. In other words, we test the idea that the trait is inferred from the behavioral description at the same extent in both STI and STT. Additionally, and in agreement with the attentional hypothesis presented in the previous Chapters, we suggest that what varies in these two conditions is the way the person (the actor and the irrelevant person) is processed and the link that is being created between the person and the inferred trait.

In the first experiment, the false recognition task was employed in order to replicate a stronger STI effect than STT effect, a difference frequently reported in the literature. In particular, we show that this stronger effect concerns the link that is created between the trait inferred and the representation of the actor. After validating our experimental setting as capable of detecting this difference in the link, we used a similar setting to test participants' memory of the sentences and of the faces. To test the memory for the sentences we used the forced choice recognition paradigm and to test the memory for the faces we used a recognition task. We demonstrate that the trait is equally inferred from the behavioral sentence in both STI and STT, whereas participants show a better memory for the face of the actor than for the face irrelevant to the behavior.

4.1 INTRODUCTION

Imagine that the first time you meet your new neighbor he makes a joke about a disabled person. If you are trying to deliberately create an impression about him, you will very likely think that he is a cruel person. However, even if you are not explicitly trying to form an impression, you will still probably activate the trait "cruel". For many years, researchers working on person perception believed that single observations like this would not be enough to trigger trait inferences, especially without the intention to do so (Srull & Wyer, 1979). But research on spontaneous trait inferences (STI) has now established that people do infer traits without intention and without awareness (Winter & Uleman, 1984), that is, spontaneously (for a recent review see Uleman, Rim, Adil Saribay, & Kressel, 2012).

This kind of inferences has important social purposes, one of which is to help us distinguish, between individuals that might be threatening in some way (and that we should be careful about), from individuals that are friendly or helpful (who we should approach). However, we not only quickly infer traits from observed behavioral information, we also share and communicate that information to others. Indeed, a great amount of our communication consists of us talking about other people, others talking about us, or others talking to us about third parties. Thus, besides inferring the trait cruel about your neighbor, you may also share that information with others. One of the consequences of sharing this kind of behavioral information is that it allows others to form their own impression about the actor being described. So, besides our great ability to extract personality characteristics about others directly from their observed behaviors, we also form impressions about others indirectly, based on what we are told about them.

Now imagine that after your neighbor makes the cruel joke, later on that day, you meet a friend at a pub and tell him about the event. Your friend will probably infer that your neighbor is cruel, just as you did. However, according to recent research, the inferred trait can also become associated with the communicator of the behavior – which would be you in this case. This surprising phenomenon has been called spontaneous trait transference - STT (Skowronski, Carlston, Mae, & Crawford, 1998), and is said to occur when an inferred trait is erroneously transferred to an irrelevant person. It has been shown that spontaneously inferred traits can be transferred not only to communicators, but also to faces that are said to be randomly paired with the behaviors (Goren

& Todorov, 2009; Skowronski et al., 1998, experiment 3), and even to inanimate objects (Brown & Bassili, 2002) that happen to be present in the same context as the behavior being described.

The STT effect had an enormous impact in the field, lighting up the discussion about the nature of the processes underlying STI and STT. A natural question that immediately arises is: how are STI and STT different? There is no simple answer to this question. Nevertheless, STT has been consensually described as a consequence of associative processes that link the implied trait to the irrelevant stimulus, as a result of their spatial and temporal co-occurrence (Carlston & Skowronski, 2005; Goren & Todorov, 2009; Orghian, Garcia-Marques, Uleman, & Heinke, 2015). In regard to STI, some authors suggested that attributional thinking is the responsible process (*e.g.*, Skowronski et al., 1998; Crawford, Skowronski, & Stiff, 2007). However, the use of attributional processes to explain STI is not very consensual (Bassili & Smith, 1986; Orghian et al., 2015), mainly because of the effortful, logical and causal legacy that the concept of "attribution" brings with it from classical attributional theories (Jones & Davis, 1965; Kelley, 1967). Using such a complex process as attribution doesn't seem adequate to describe what is believed to be so efficient and spontaneous. The debate about the nature of STI and STT has been stimulated by empirical differences that have been reported in the literature between the two effects. One of the most cited differences in the magnitude of the effect: STI effect is usually stronger than STT effect (*e.g.*, Goren & Todorov, 2009; Brown & Bassili, 2002; Skowronski et al., 1998). Also, if people are instructed to perform a concurrent inferential task, such to detect whether the person is lying (about their own behavior or about someone else's behavior), the magnitude of STI is reduced, while no such reduction is detected in STT (Crawford, Skowronski, Stiff, & Scherer, 2007). Another important difference is the fact that STT can be reduced or eliminated if the actor is presented at the same time next to the non-actor (*e.g.*, Goren & Todorov, 2009; Todorov & Uleman, 2004). Moreover, there is a negativity effect (more inference from negative behaviors than positive) for the STI, while no such negativity is detected in STT (Carlston & Skowronski, 2005).

All these evidence suggests that STI and STT are different, however it does not explain exactly what this difference consists of.

In the current paper our goal is to explore the way the different elements (*i.e.*, the behavior, the trait, and the face) are processed in a typical STI or STT experiment. In order to consider this

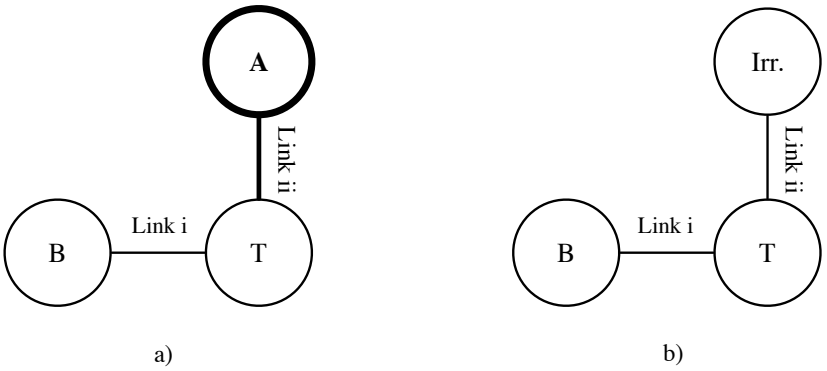

Figure 13: Representation of the two sub processes underlying the occurrence of STI and STT. Link i: activation of the trait from the behavior and Link ii: the connection between the activated trait and the person. Thicker lines connecting the nodes stand for stronger associations between the concepts represented by the nodes. Circles with thicker edges stand for strongly activated nodes.

issue, we should first look at how a typical STI and STT experiment is carried out. Usually in these studies, photos of faces of people are presented together with behavioral descriptions, and participants are instructed to familiarize themselves with the material or to memorize it for a later unspecified test. We specifically want to explore how the *face* and the *behavioral description* are differently processed under STI and STT conditions.

Before considering different processes, we took one step back and decided to investigate how the different elements, the behavior, the person and the trait, are processed in function of the relevance of the actor. We would like to suggest that the behavior is encoded in a similar way in both STI and STT conditions. However, we think the face, because more relevant and salient, will be processed more extensively under STI than under STT conditions. We believe these differences are important and can bring us a step forward towards the understanding of the specific processes underling the two phenomena.

The occurrence of STI and STT can be regarded as a consequence of two sub-processes: first, the trait is activated upon reading the behavioral sentence (due to the link i in figure 13) and, second, the activated trait establishes a link in memory with the actor (link ii in figure 13). STI and STT can differ either in the initial trait activation process, or in the establishment of the subsequent trait-actor link, or they can differ in both processes. Skowronski and colleagues (1998) defended in their model that the activation of the trait from the behavior is an initial stage

common to both effects (STI and STT). In other words, the same categorization of the behavior in terms of traits (*e.g.*, cruel behavior) takes place in both STI and STT. What is different is the link subsequently established between the person and the trait.

However, this idea that the behavioral descriptions are equally processed was never fully tested. STI measures typically focus on the link ii (person-trait), but not on eventual differences in link i (behavior-trait). In both false recognition and in savings in relearning tasks, the two most reliable measures of STI and STT, link ii is tested after subjects learn the association between the person and the behavioral description. In the false recognition paradigm, pairs of faces and traits are presented during the test phase and the underlying assumption is that participants will falsely indicate that the trait was part of the sentence previously presented with the actor, due to the specific trait-actor link that was encoded in memory. In saving in relearning paradigm, the face of the actor is presented as a cue for the recall of the trait, allowing researchers to access the trait-actor link. This is more a virtue than a limitation, since these paradigms were developed precisely to prove that the inferred trait is not merely a description of the behavior, and that a specific link between the actor and the trait is established in memory. Previous studies using the false recognition and the savings tasks have shown that STI are stronger than STT (*e.g.*, Skowronski et al., 1998; Goren & Todorov, 2009). This difference indicates that, comparing with irrelevant individuals, the actors are more likely to be bounded to the activated traits. But whether STI and STT also differ in the initial trait activation process has not yet been tested.

Our first goal is to directly test the idea that STI and STT involve the exactly same initial trait inference process, but differ in the binding between the face and the trait. Specifically, according to the model proposed by Skowronski and colleagues (1998), during behavior encoding, the trait is equally activated, in both STI as in STT conditions. It is further proposed that the difference between STI and STT resides in the binding between the trait and the photo. If the photo belongs to the actor of the behavior, then the activated trait will establish a strong connection with the trait implied by the behavior. However, if the photo belongs to an unrelated individual then the associative link between the implied trait and the face will be weaker (see figure 13).

Moreover, we suggest that what motivates the differences in the binding is the difference in the way the photo is processed due to the difference in relevance. In particular, the actor is more

relevant and salient in the context of its own behavior than the irrelevant person that did not enact the behavior.

What we suggest in the present paper is to separately access the two sub-processes involved in STI and STT, by using two different measures, one measurement that tests the actor-trait link (*i.e.*, the typical false recognition task) and another measurement that accesses the behavior-trait link (a forced recognition task). In our first experiment we used the false recognition task and expected to find stronger STI effect than STT effect. After replicating the difference in magnitude between STI and STT and confirming that our experimental setting was sensitive to this difference we conducted a second experiment. The experiment 2 was conducted using exactly the same materials and the same learning phase. In the test phase the participants performed a forced recognition task, in which two versions of the initial sentence were presented: one corresponded exactly to the sentence presented during encoding, for example, a sentence with the trait just being implied in it (*e.g.*, He won the science quiz) and a second sentence that includes the implied trait in it ("He is *smart* and won the science quiz") or vice versa. Participants had to choose the version that was presented before. If the same amount of processing was engaged by the sentence in the two conditions, then participants were expected to have a similar performance in both STI and STT conditions. In other words, if the model proposed by Skowronski and colleagues (Skowronski et al., 1998; Crawford, Skowronski, & Stiff, 2007) is correct, the difference between STI and STT should have nothing to do with the processing of the sentence or the inference of the trait from the sentence, and thus, no differences in the performance in the forced recognition task should be found between STI and STT conditions.

In our second experiment, we also included a face recognition task with the goal of providing complementary evidence that the difference between STI and STT resides in the strength of the link between the trait and the face and specifically in the processing of the face itself. According to our view, the stronger binding between the face and the trait in the STI condition is triggered by the relevance of the actor that makes it more salient in the context of the behavioral description. If that is the case, then a better memory for the actor's faces than for the irrelevant faces is expected. Thus, in addition to the stronger trait - person link, we believe that the activation of the representation the actor is also stronger than the representation of the irrelevant person.

4.2 EXPERIMENT 1

In this experiment we wanted to replicate one of the results previously found in the literature, a larger STI than STT effect (Goren & Todorov, 2009) using the false recognition paradigm (Todorov & Uleman, 2002, 2003, 2004). In this study, participants started with a learning phase where they were instructed to memorize pairs of behavioral descriptions and photos of people. In half of the trials, the photo belonged to the actor of the behavior being described, and in the other half, the photos belonged to an irrelevant actor (said to be randomly paired by the computer). In the test phase, pairs of photos and traits were presented and the participants' task was to indicate whether the trait was presented in the sentence that was paired in the learning phase with that photo. In this experiment we also wanted to explore the processing of the behavior and the face when the trait is explicitly provided as opposed to inferred. STI and STT are very difficult to avoid (*e.g.*, Carlston & Skowronski, 2005), including the extraction of the trait from the behavior, because the participants are not usually aware of its occurrence (exactly because they are spontaneous), but what would happen is the trait would be explicitly mentioned in the sentence? Will the irrelevance of the trait be more noticeable and will the participant, because of that, try to avoid linking it to the person presented? Going back to our example, would my friend still transfer the trait to me had I not only described my neighbor's behavior but also communicated my impression of him ("He made a joke about a disables person, how can he be so cruel?")?

4.2.1 *Method*

Participants

Ninety participants took part in the experiment, from which 16 were males. The average age of the sample was 20.26 years old. Because we had troubles in replicating the difference in magnitude between STI and STT with the false recognition paradigm in previous studies (previous Chapter), we decided to collect more participants than the samples sizes usually reported in the literature. However, the stopping moment was defined by another experiment that was conducted

in the same session that had a requirement of 90 participants. In all the experiments reported in this Chapter, the data were only analyzed after the reported sample sizer was complete.

Material

We used 52 trait-implying sentences, previously pre-tested to imply a trait, and 52 black and white photos of people with neutral expressions. The trait-implying sentences were chosen in such a way that the trait implied in half of the sentences were antonyms (or opposite in meaning) of the trait implied in the second half of the sentences.

4.2.2 *Procedure*

The experimental design of this study is: 2 Relevance (actor versus irrelevant person) $\times$ 2 Trial type (experimental/critical versus filler) $\times$ 2 Pairing (match versus mismatch), all the factors being within-subject. The rate of false recognitions is the dependent variable.

The task was presented as a memory task, where participants were instructed to memorize sentences and photos of people. Instead of having typical STT trials, in which the irrelevant person is a friend or a communicator, we told participants that some of the photos were randomly paired with the sentences (the trials marked with the red frames around the photos). Regarding the STI trials, the participants were told that the person described in the sentence is the same as the one in the photo (the trials marked with the blue frame). Using random pairing in the STT trials is a way of avoiding any logical reason for linking the trait to the irrelevant person. If the person communicating the behavior is a friend of the actor, the participant might think that because friends are similar in their personalities the communicator might share the trait of the actor. The same can happen even when the communicator is not a friend, the participant might think that if the communicator chose to communicate the behavior that is because somehow he/she agrees, or disagrees, etc. This type of random paring was already used in the past (*e.g.*, Todorov & Uleman, 2002; Goren & Todorov, 2009).

The learning phase started with 4 practice trials. Each trial started with a 500 ms fixation dot followed by the sentence and the photo that remained on the screen for 7000 ms. Finally, the stimuli were followed by 100 ms blank screen. As mentioned before, in half the trials, the photo

came with a red frame (irrelevant person trials) and the second half with a blue frame (actor trials), however the order of relevant and irrelevant trials was randomized for each participant. In one third of each of the two types of trials (actor and irrelevant), the sentence contained the implied trait ("He is so lazy that spent the whole day watching series on the TV"), these are filler trials. In the remaining two thirds of the trials the trait was only implied in the sentence ("He spent the whole day watching series on the TV"). Participants weren't given any specific information about the type of memory test they were going to perform later on. After memorizing the 48 pairs of sentences and photos (24 actor trials and 24 irrelevant trials) participant did a short distractor task (solving an anagram puzzle).

Next, they performed a memory test where the same photos from the learning phase were presented together with a word. The word was always a trait and the task was to indicate whether the word was presented in the sentence seen alongside that photo in the learning phase. If the trait was inferred from the sentence then, in the test, it will be more difficult for the participant to say whether the trait was actually in the sentence or whether he only thought/inferred it, and thus a false recognition may take place. In the test phase there were two different conditions, one where the pairing between the photo and trait was in agreement with the trial presented during the learning phase, that is, the trait was implied in the sentence presented with that person – match trials, and trials where the photo was paired with a trait that was implied in a sentence that was presented with some other person – mismatch trials. Having these two conditions answers a frequent question in trait inference studies: whether the trait is actually an attribute of the actor or just a categorization of the behavior. We do not only want to know if the trait is inferred from the sentence but also if the trait is linked to person. Calculating the difference between match and mismatch gives us exactly that, the amount of the false recognition that is due to the link to the person, since the rest (the familiarity with the faces and the traits) is kept constant.

The mismatch trials were created is such a way that the meaning of the trait presented at test was as opposite as possible (taking into account the available pool of material) to the meaning of the trait implied in the sentence presented with that person. This is an important aspect because we know that research has been showing a halo effect in trait inference, especially strong in STI (Skowronski et al., 1998; Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007). This means that if a certain trait is inferred, the

effect generalizes to other traits of similar valence. So, even if we present a different trait in the mismatched trial, if the trait is similar in valence to the inferred trait then the mismatch will work in similar way as would the match condition. This is specially problematic because the halo effect is not constant in both phenomena. Choosing antonyms eliminates any possibility of having overlapping meanings between the trait inferred and the trait given at the test. Moreover, Todorov and Uleman (2002) used antonyms in their studies, and found that the false recognition for the antonyms is even lower than for unrelated control traits, meaning it can favor the detection of the effect. This is in agreement with Gross, Fischer and Miller's findings (1989) that have demonstrated that adjectives are organized as opposites and as such, it makes sense to assume that if a trait is positively associated with an actor, the antonym will be negatively associated with that actor.

The test phase also started with 4 practice trials. Each trial started with a fixation dot that remained on the screen for 500 ms. Next, the photo and the trait appeared and remained there until the participant gave an answer. To give the answer subjects used the "S" key to say "Yes" and the "L" key to say "No" and the instruction was to answer as quickly and accurately as possible. After their response a black screen was presented for 100 ms before the next trial started.

4.2.3 *Results and Discussion*

We started the statistical analysis with a repeated measure ANOVA where both the experimental trials and the fillers were included. The dependent variable is the rate of "yes" responses. The independent factors are: the type of trial (experimental versus fillers), the relevance (actor versus irrelevant person) and the pairing (match versus mismatch). The p-value presented in the following analysis are two-tailed except for the cases where the result is a replication. Since a significant three-way interaction was observed, $F(1,89) = 22.96$, $p < .001$, $\eta_p^2 = .21$, two separate ANOVAs were conducted for the experimental and for the filler trials.

The repeated measures ANOVA conducted for the experimental trials, revealed a main effect of pairing, $F(1,89) = 62.93$, $p < .001$, $\eta_p^2 = .41$, with a larger rate of false recognitions in the match condition ($M = .40$, $SD = .19$) that in the mismatch condition ($M = .26$, $SD = .15$).

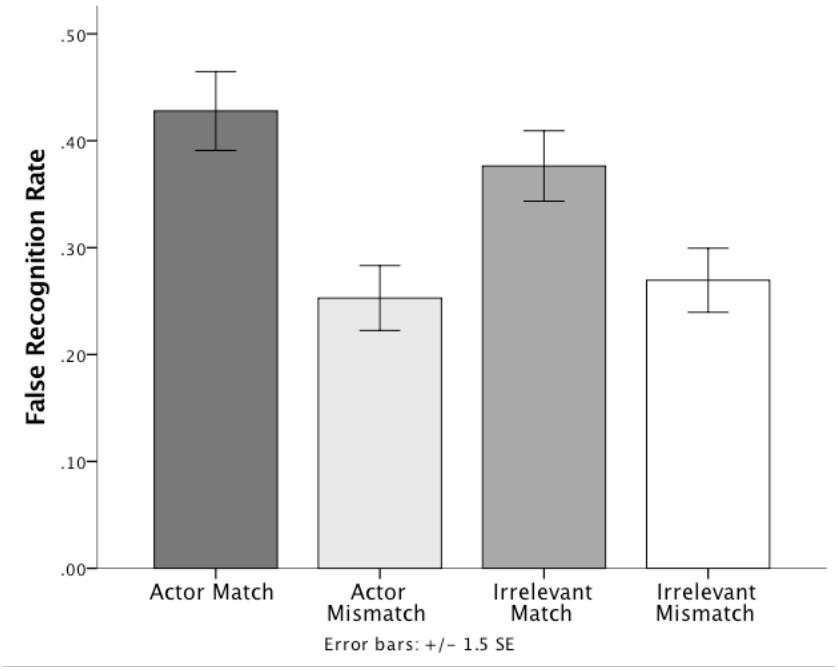

Figure 14: False recognition rate in function of relevance and pairing for the experimental trials in experiment 1.

This result means that participants are not only inferring the trait from the sentences, they are also linking the traits to the presented persons. And, as expected and replicating the results found by Goren and Todorov (2009), a significant interaction is detected between the relevance and the pairing, $F(1,89) = 3.79$, $p = .028$, $\eta_p^2 = .04$. The STI effect, that is detected by comparing the match and the mismatch condition for the actor, is significant, $F(1,89) = 50.38$, $p < .001$, $\eta_p^2 = .36$, and the same is true for the STT effect, $F(1,89) = 18.03$, $p < .001$, $\eta_p^2 = .17$. The interaction is suggesting that the effect is stronger in the STI ($M(match - mismatch) = .18$, $SD = .24$) than in the STT ($M(match - mismatch) = .11$, $SD = .24$). A graphical illustration of the result is presented in figure 14.

Also note that what drives the interaction between relevance and pairing is mainly the difference between match condition for the actor and match for the irrelevant person, $F(1,89) = 4.34$, $p = .04$, $\eta_p^2 = .05$, with larger false recognition for the actor (M = .43, SD = .23) than for the irrelevant person ($M = .38$, $SD = .21$). There is no difference between the two mismatch conditions, $F < 1$.

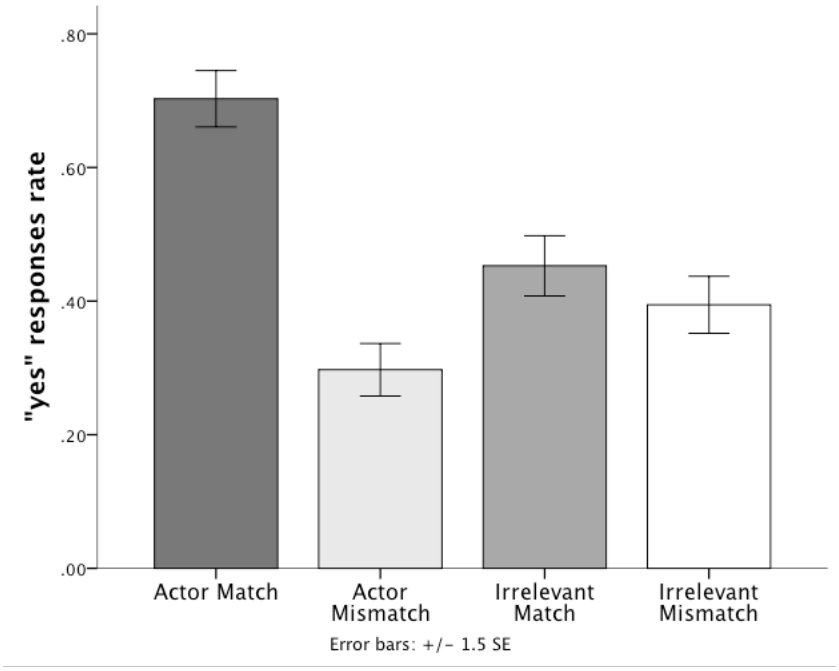

Figure 15: The rate of "yes" responses in function of relevance and pairing for the filler trials in experiment 1.

A similar analysis was performed for the filler trials. The dependent variable in the analysis is the proportion of “yes” responses and independent variable is the relevance (actor versus irrelevant person) and the pairing (match versus mismatch). These trials are not about inference because the trait was provided in the sentence, but we still can compare the match with the mismatch condition since it will inform us about the way the provided trait is being linked to the relevant and irrelevant person. We don’t have a hypothesis here, since this analysis in not usually reported in the literature. Also note that responding “yes” in the match condition is a correct answer, but the same is not true for the mismatch, where the correct response is “no” since the trait was presented in the sentence but not with that person.

As it can be observed in figure 15, a main effect of relevance, $F(1,89) = 7.84$, $p = .006$, $\eta_p^2 = .08$, and a main effect of pairing, $F(1,89) = 70.15$, $p < .001$, $\eta_p^2 = .44$, were verified. And more importantly, a very significant interaction was found between the relevance and the pairing, $F(1,89) = 47.49$, $p < .001$, $\eta_p^2 = .35$. The link between the relevant person, the actor, and the trait is strong since the difference between match and mismatch is significant ($M(match - mismatch) = .41$, $SD = .36$), $F(1,89) = 116.36$, $p < .001$, $\eta_p^2 = .57$, while the

same is not true for the irrelevant person, since the difference between match and mismatch does not reach significance ($M(match - mismatch) = .06$, $SD = .35$), $F(1, 89) = 2.45$, $p = .121$, $\eta_p^2 = .03$.

In experiment 1 we found a STI and a STT effect, meaning the trait is inferred from the behavioral description and it is linked to the person in the photo. We also found a difference in the magnitude of these two effects, with a larger effect when the person in the photo is the actor than when the person is said to be randomly paired with sentence, replicating previous finding in the literature (Brown & Bassili, 2002; Skowronski et al., 1998; Goren & Todorov, 2009).

Regarding the fillers, that is, the trials where the trait was explicitly provided as part of the sentence, a different pattern of results was observed. There was a strong link between the trait with the actor, but there seems to be no significant link between the provided trait and the irrelevant person, a result to which we'll come back later on in the manuscript.

4.3 EXPERIMENT 2

After replicating the difference in magnitude between the STI and STT, we conducted a second experiment in order to test the idea that the inference of the trait from the sentence is the same in both conditions, applying the forced recognition paradigm. This same paradigm was used in a similar format by Ferreira and collaborators (2012). In their studies, after presenting behavioral descriptions, participants' task was a forced choice between two versions of trait implicative descriptions that were identical in every aspect except the fact that one included the implied trait while the other did not. Only one was actually presented in the study phase, and the participants had to choose the one that was. The main reasoning behind the task is that, if the trait is spontaneously activated, then the internally generated traits will be confused with the externally generated traits and that will lead to more false recognitions.

4.3.1 *Method*

Participants

Sixty-seven participants took part in the this experiment from which 9 were males. Their average age was 19.10 years old. The sample size was defined by the number of show-ups in a week time.

Material

The sentences used in the second experiment were exactly the same as the ones used in experiment 1. We also used 96 colored photos of people with neutral expressions.

4.3.2 *Procedure*

The experimental design in this study is: 2 Relevance (actor versus irrelevant person) $\times$ 2 Test (memory for sentences versus memory for faces) $\times$ 2 Type of trial (experimenta versus fillers) with all the independent variables being within-subject and the dependent variable being the accuracy in the memory tests.

The learning phase and the instructions were the same as in experiment 1 except the fact that, this time, half the trials had sentences where the implied trait was actually part of the sentence (instead of $1/3$ from experiment 1) and the other half just had the trait implied. Just like in experiment 1, half the trials were STI and the second half were STT. After studying the material there was a short distractor task followed by two memory tests. The first test was a forced choice between two sentences, a sentence where the implied trait was part of the sentence and a sentence without the trait (the trait being just implied). Only one of the two had actually been seen before and the participant had to choose which one it was. Half the sentences were labeled as A and half as B and the participant had to indicate which one (by pressing "A" or "B") they saw in the learning phase. A 100 ms black screen was presented in between the trials. The assignment of label A or B to the correct sentence was randomized for each participant. The second test was a recognition test where 48 old and 48 new photos were presented and the participants were

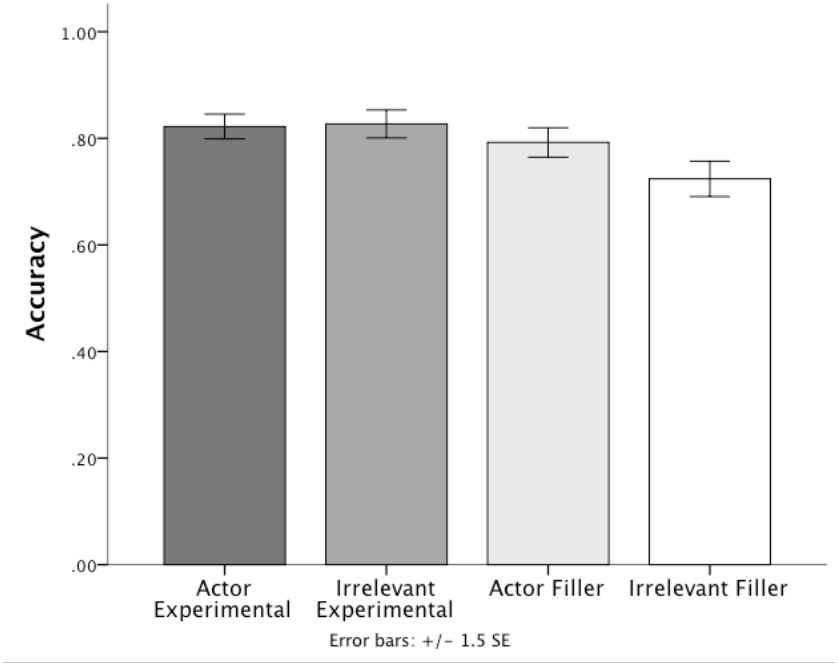

Figure 16: Accuracy for the memory test regarding the Sentences in function of relevance and type of trial in experiment 2. In the experimental trials the trait were only implied in the sentence whereas in the filler trials the trait was part of the sentence.

instructed to indicate as quickly and accurately as possible if the photo was presented in the learning phase (pressing the "S" key) or not (pressing the "N" key). Half of the old faces were from the STI condition and the other half from the STT condition. The inter-trial interval was again 100 ms.

4.3.3 *Results and Discussion*

Two different analysis were conducted, one for each type of test.

The first analysis is a repeated measure ANOVA for the performance in the forced choice recognition test. The dependent variable was the accuracy and the independent ones were the relevance (actor versus the irrelevant person) and the type of trial (fillers versus experimental). A main effect of relevance, $F(1,66) = 5.21$, $p = .026$, $\eta_p^2 = .07$, and a main effect of type of trial, $F(1,66) = 9.01$, $p = .004$, $\eta_p^2 = .12$, were found. Also, an interaction between the relevance and the type of trial was found, $F(1,66) = 5.47$, $p = .022$, $\eta_p^2 = .08$. As it can be seen in figure

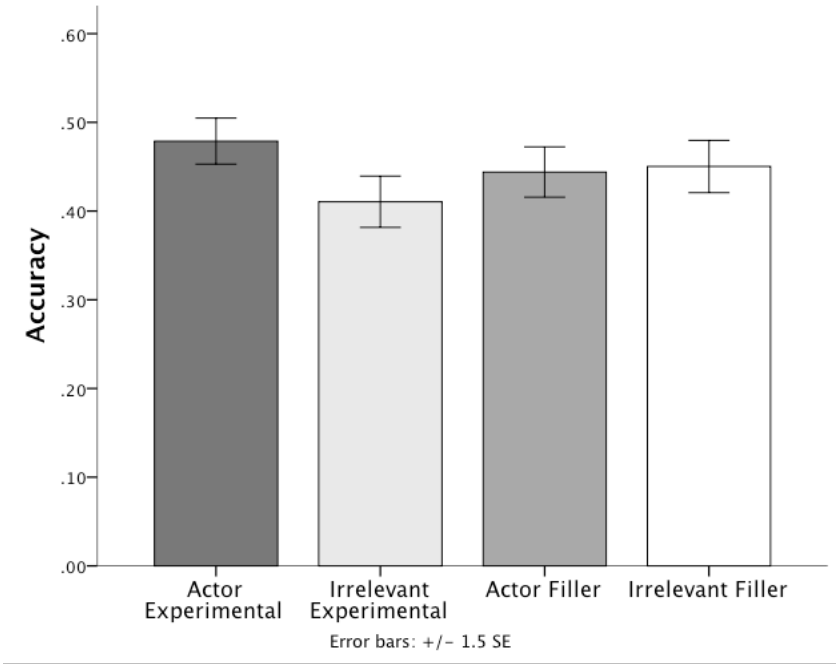

Figure 17: Accuracy for the memory test regarding the Faces in function of relevance and type of trial in Experiment 2.

15, there is no difference in accuracy between the actor ($M = .82$, $SD = .13$) and the irrelevant person condition ($M = .83$, $SD = .14$) in the experimental trials, $F < 1$. Whereas in the filler trials the difference between the actor ($M = .79$, $SD = .15$) and the irrelevant person ($M = .72$, $SD = .18$), is significant, $F(1, 66) = 9.32$, $p = .003$, $\eta_p^2 = .12$.

In the memory test for faces, the accuracy rate was overall low ($M = .47$, $SD = .06$) which can be explained by the large number of trials in the learning phase. The performance for the new faces were not included in the following analysis. An ANOVA was performed with dependent variable being, this time, the accuracy in the test regarding the memory for the faces. As it can be seen in figure 16, there is an interaction between the type of trial and the relevance, $F(1, 66) = 4.90$, $p = .030$, $\eta_p^2 = .07$. There is a significant difference between the actor ($M = .48$, $SD = .14$) and the irrelevant person ($M = .41$, $SD = .16$), in the experimental trials, $F(1, 66) = 9.09$, $p = .004$, $\eta_p^2 = .14$, whereas the difference between actor ($M = .44$, $SD = .15$) and irrelevant ($M = .45$, $SD = .16$) in the filler trials is not significant, $F < 1$.

As predicted by our hypothesis, concerning the way the behavioral description is processed, no differences were found between STI and STT, meaning that the relevance of the person does not

affect the way the trait is inferred from the behavioral description. However, people do accept the actor's face as being presented before more often than they do for the irrelevant face. That might suggest that there are differences in the way the two persons are processed. And that might happen because the actor is more salient, and thus more attention is paid to it, something that we suggest in Chapter 2 and 3. Even though we cannot infer a causal link between the additional processing of the face and the larger magnitude of the STI effect, we do believe that these two are related.

4.4 GENERAL DISCUSSION

In the present work we tested the idea that the difference between STI and STT is not due to the way the trait is activated from the behavior but due to the way the actor and the irrelevant persons are processed and how the trait is linked to the representation of the person.

Moreover, as shown in Chapter 2, there are some indications that the actor's face receives more attention from participants in a false recognition paradigm setting than does the irrelevant person's face. Thus, we further speculated that, because the actor's face is more attended and more processed, participants, when presented with actors' faces in a recognition test, would find them more familiar than the irrelevant faces.

Thus, besides replicating a stronger association between the trait and the actor than the association between the trait and the irrelevant person, in this paper, we tried to test two hypothesis: 1) the activation of the trait from the behavior is similar in STI and STT and 2) the face of the actor is more processed than the face of the irrelevant person.

In the first experiment, by applying the false recognition paradigm, we found a stronger link between the inferred trait and the actor of the behavior implying that trait than between the implied trait and the irrelevant person that was said to be paired with the behavior randomly by the computer. When the trait was actually part of the sentence presented with the actor/irrelevant person, a very strong actor-trait link was detected, whereas no irrelevant person - trait link was verified. This result, speculatively, can be interpreted as an inhibition that subject applied to the provided trait that does not correspond to the person presented. In other words, when the trait,

instead of implied is explicitly provided the subject might have more control over its binding to the irrelevant person.

This paradigm, whereas very informative in respect to the link created between the trait and the person, is not very enlightening in respect to the way the trait is extracted from the behavior in function of the relevance of the person accompanying the behavior.

In experiment 2 we used the forced choice recognition paradigm in order to investigate participants' memory of the sentence, and in particular regarding the presence of the trait. We found no difference between the STI and STT conditions in this task. However, we did find a difference in the memory performance for the faces of actors and irrelevant persons, with more familiarity for the actors' faces.

These results are compatible with the results obtained in the simulations presented in Chapter 2. The model was created with the same assumption that the behavior activated the trait in equal matter independently of the relevance of the person, and what varied was the activation of the face that received higher values input when it was an actor's face than when it was the face of an irrelevant person. Consequently, this difference in the input for the person leads to different strengths in the connections between the person and the trait nodes (with a stronger connection when the person is an actor).

Finally, on a concluding note, STT is usually thought to occur due to spatial and temporal contiguity, by encoding the behavior and the photo of a person simultaneously. However, contrary to STI where the trait is thought to affect the representation of the actor, the trait is not expected to actually be integrated in the the representation of the irrelevant person. We cannot conclude from these experiments that this difference in the links and in the processing of the person can be translated into differences in terms of integration in the representation of the person. But we believe that by knowing that the locus of the difference between STI and STT might be in the way the person is processed and the consequent link with the trait and not in the way the trait is extracted from the behavior, can be useful for future investigation in STI.

4.5 REFERENCES

Bassili, J. N., & Smith, M. C. (1986). On the spontaneity of trait attribution: Converging

evidence for the role of cognitive strategy. *Journal of Personality and Social Psychology*, *50*(2), 239–245.

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.

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.

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*(3), 466–472.

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.

Ferreira, M. B., Garcia-Marques, L., Hamilton, D., Ramos, T., Uleman, J. S., & Jerónimo, R. (2012). On the relation between spontaneous trait inferences and intentional inferences: An inference monitoring hypothesis. *Journal of Experimental Social Psychology*, *48*(1), 1–12.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, *27*(2), 222–248.

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

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. *Advances in Experimental Social Psychology*, *2*, 219–266.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 15). Lincoln.

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition*, *33*(1), 20-66.

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.

Srull, T. K., & Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social psychology*, *37*(10), 1660–1672.

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.

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.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.

Uleman, J. S., Rim, S., Adil Saribay, S., & Kressel, L. M. (2012). Controversies, questions, and prospects for spontaneous social inferences. *Social and Personality Psychology Compass*, *6*(9), 657–673.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252.

5

ACKNOWLEDGING THE ROLE OF WORD-BASED ACTIVATION IN SPONTANEOUS TRAIT INFERENCES

Diana Orghian, Tânia Ramos, Joana Reis and Leonel Garcia-Marques

Under review

This Chapter introduces a discussion about a confounder present in studies that explore the topic of Spontaneous Trait Inferences (STI). In these studies the material employed are, most of the times, written trait-implying behavioral descriptions. When researchers are preparing these materials, their main objective is to create sentences that strongly activate personality traits about the actors performing behaviors. A potential limitation of this material is the possibility of the trait becoming activated by specific words in the sentence (word-based priming) and not, or not only, by an inference made based on the comprehension of the behavior and the sentence as a whole (text-based priming). This aspect has been recurrently ignored in many studies in the field. In the present Chapter, we suggest a way of overcoming this confounder. We created a set of 122 trait-implying sentences and their correspondent control versions. These control sentences have approximately the same words as the trait-implying sentences, but the words are rearranged in such a way that the sentences gain a new meaning and no longer imply the target traits. Keeping the words constant controls for the eventual activation coming from individual words in the sentences. Thus, the second goal was to provide a solution for the problem raised and present an initial set of material that researchers interested in investigating STI with the Portuguese population can use in their research. It is also discussed how the word-based priming may have obscured the interpretation of previous results in the literature.

5.1 INTRODUCTION

Individuals have the fascinating talent of making inferences and reading between lines. This is an efficient and effortless skill that plays a crucial role in our comprehension of the world. If someone says that "she likes apples", we automatically know that the person is referring to eating apples. We don't need to retrieve all the possibilities that would match this same context (*e.g.*, "she likes the color of apples", "she likes planting apples", "she likes reaping apples") instead, we just immediately select the most appropriate one ("she likes eating apples"). Inferential thinking is useful in many different contexts; one of them is the social context. If you know that "John did not smoke at home while his roommate was trying to quit" it is very likely you'll infer that John is "considerate". Research in social cognition has shown that such trait inferences may occur spontaneously, in the absence of any specific intention to form impressions or to infer traits (*e.g.*, Winter & Uleman, 1984; Uleman, Rim, Adil Saribay, & Kressel, 2012, for a recent review).

Inferring personality traits from behavioral descriptions seems intrinsic to our comprehension, but in experimental terms there are some recurrent difficulties. One of them has been in demonstrating that the inference is made during the encoding and is not based on retrieval processes. STI research has put much effort on creating better paradigms to demonstrate how and when trait inferences occur. But less attention has been paid to the type of material used in these paradigms, in particular to how the sentences can activate the target trait without inferential processes being necessarily involved. A plausible alternative explanation for the activation of a trait are word-to-word processes. That is, the activation of the trait can be the result of an association with very specific words in the sentence, and not the result of the understanding of the entire sentence. This word-based priming cannot be considered an inference, but its result may perfectly mimic that of an inference.

The present Chapter has two main goals. First, to initiate a discussion about the eventual influence of word-based priming in the STI literature. Second, to present a solution along with a set of material that controls for the word-based priming and that researchers conducting studies with Portuguese speaking participants can use in their own investigations.

We start by discussing the importance of distinguishing word-based priming from true trait inferences both theoretically and methodologically. We then show how this issue has been ne-

glected in many of the past studies in the trait inference literature. Finally, we present a set of sentences, which were created and initially validated for the Portuguese language, that can be further developed and applied by researchers in future STI studies.

5.2 WORD-BASED PRIMING

Keenan and colleagues (Keenan, Potts, Golding, & Jennings, 1990; Keenan & Jennings, 1995) highlighted the problem of how word-based priming might contaminate the interpretation of results in studies that examine inferences during text comprehension. They noticed that two sources of priming, word-based priming and sentence-based priming (*i.e.*, text inferences), had been confounded in many previous studies in the field, rendering the evidence for inferences somewhat inconclusive. These authors claim that word-based priming is based on intra-lexical associations and it is insensitive to the meaning of the whole sentence, whereas an inference is based on the meaning of the text in combination with perceiver's knowledge of the described situation. Thus, it was crucial to question whether results that had been taken as evidence of elaborated, higher level inferential processes, were not simply a result of word-based associations.

Within the STI research, the main aspect we want to emphasize is that trait activation, usually interpreted as an inference, can be a consequence, not of the (spontaneous) interpretation of the behavioral description, but of the presence of certain words that are semantically associated with the trait. It is common for trait-implying sentences to contain words highly associated with the trait concept. Take, for example, the following behavioral description employed by Carlston and Skowronski (1994) in one of the most cited papers in the STI literature: "I am 18 years old and a doctor. I received my medical degree from Harvard. In my spare time I enjoy doing research at the Mayo Clinic.", a behavioral description that implies the trait "intelligent". There are several words in the sentence ("degree", "Harvard" and "research") that may activate the trait intelligent, independently of the meaning of the entire sentence. And note that, if the sentence is processed in a shallow manner (if the participant is under cognitive load, tired or just unmotivated for the task), he still can pick up the presence of these words and activate the trait without deeply processing the meaning of the sentence. The need of experimentally controlling for these associations is

crucial, not just because of their inference mimicking power, but also because they might interact with sentence-based inferences in unknown ways.

One efficient solution to this problem would be to eliminate all word-based associations that could possibly activate the trait. However, it is very difficult, if not impossible, to eliminate all the associations (Keenan et al., 1990; Kintsch & Mross, 1985; Forster, 1981). A more viable solution is to equate the word-based associations present in the trait-implying behavioral descriptions with the control versions by using the same words but rearranged in such a way that the meaning of the control sentence changes so that it no longer implies the target trait. It is then possible to verify if the trait gets more activated in the inference version than in the rearranged version. Because the trait-implying sentence and the control version have similar words, word-based priming effects should be of similar magnitudes in both sentences. Any differences in the activation of the trait between the two sentences can be attributed to the processing of the meaning of the sentence, that is, to text-base inferences. The material presented in this Chapter is essentially the implementation of this second solution.

Surprisingly enough, neither the problem nor the solution are new for the text comprehension researchers (McKoon & Ratcliff, 1986; Potts, Keenan, & Golding, 1988). For example, in order to study the occurrence of predictive inferences, McKoon and Ratcliff (1986) applied a probe recognition task. The authors, in order to test for the existence of inferences, presented paragraphs that included either a predicting sentence ("The director and the cameraman were ready to shoot closeups when suddenly the actress fell from the 14th story.", where "dead" should be the inferred prediction) or a rearranged sentence ("Suddenly the director fell upon the cameraman, demanding closeups of the actress on the 14th story." where "dead" shouldn't be inferred). Note that roughly the same words were used in both sentences. After reading the last sentence in the paragraph, a word was presented and the participants had to indicate, as quickly and accurately as possible, whether the word was presented in the paragraph. In the critical trials the target word was the predicted event ("dead"). If the predicted event was inferred during the reading of the predictive sentence, then the correct response ("no") to the target should be slower and more inaccurate than in the case in which the predicted event followed a rearranged sentence. This was exactly what was found.

5.3 SPONTANEOUS TRAIT INFERENCES

The first paradigm used in the STI field was the cued-recall paradigm (Claeys, 1990; Uleman, Moskowitz, Roman, & Rhee, 1993; Winter & Uleman, 1984; Winter, Uleman, & Cunniff, 1985). In this paradigm, participants are usually presented with trait-implying behaviors, under memory instructions. After a distractor task, participants are asked to recall the previous behaviors under various cue conditions. The central assumption is that if the trait is inferred during the encoding of the sentence, it will lead to a better recall of the behavior when given as a cue, compared to the no cue condition.

Although none of the studies applying the cued-recall paradigm used rearranged sentences, Winter and Uleman (1984) revealed concern about the potential interference of word-based associations with their results. These authors conducted a series of rigorous pre-tests that controlled for some of the word-based activations in the sentences. For example, they controlled for semantic associations with the actors of the sentences (that were described via their occupations, like for instance, reporter and librarian), eliminating those actors that were strongly associated with the target trait. The authors wanted to rule out the possibility that traits presented at test activated the actor in the sentence on the basis of semantic associations, increasing the likelihood of retrieving the sentence, independently of trait inference processes taking place during encoding. Critically, while these researchers pre-tested the materials extensively, they did not explore all the associations between the trait and the words in the sentences, remaining inconclusive the extent to which word-based associations are affecting the results in the cued-recall task. This problem could have been overcome by using rearranged sentences to verify whether the recall of the sentence given the trait is better when the sentence implied the trait versus in its rearranged version. Other common paradigms that have been applied to explore STI are the savings in relearning (Carlston & Skowronski, 1994) and the false recognition (Todorov & Uleman, 2002). In both of these paradigms participants are initially presented with a series of photos of actors, each one paired with a trait-implying behavior. In the false recognition task, the previously seen photos are again presented at test, paired with trait-words. Some of the pairs are created by pairing the actor with the trait previously implied by his behavior (match pairs), while other pairs are created by presenting the actor with a trait previously implied by the behavior of another actor

(mismatch trials). The participant is instructed to indicate whether the trait was part of the sentence presented previously with that actor. A higher rate of false recognitions of traits in match than in mismatch condition is taken as evidence of occurrence of STI during the encoding of the behavior and evidence of the binding of the traits to the actors' representation. In the savings in relearning paradigm, in a second phase, participants are presented with face-trait pairs. Some of these pairs are "relearning pairs" (faces are paired with the corresponding implied traits) while others are new (faces are paired with new traits). In a final phase, the photos are presented as cues and participants have to recall the corresponding traits paired with those photos in the second phase. Results typically show that the recall of the traits is superior for relearning pairs than for new pairs (the co called "savings effect"), indicating that participants had inferred the traits in the initial phase of the study. Studies using the false recognition and the savings paradigms never included rearranged control versions of the trait-implying sentences. As such, the degree to which performance in these tasks is being affected by specific intra-lexical associations between the to-be-inferred traits and words from the sentences remains unclear.

To our knowledge, the only STI study that originally included rearranged sentences was conducted by Uleman and collaborators (1994) with the recognition probe paradigm. The recognition probe paradigm was borrowed by the authors from text comprehension literature (*e.g.*, McKoon & Ratcliff, 1986) and has been popular since then in the STI literature (*e.g.*, Ham & Vonk, 2003; Newman, 1991, 1993; Ramos, Garcia-Marques, Hamilton, Ferreira, & Van Acker, 2012; Uleman, Hon, Roman, & Moskowitz, 1996; Van Overwalle, Drenth, & Marsman, 1999; Wigboldus, Dijksterhuis, & Van Knippenberg, 2003; Wigboldus, Sherman, Franzese, & van Knippenberg, 2004). In their study, Uleman and collaborators (1996) used both types of sentences: trait-implying sentences ("He took his first calculus course when he was 12 years old.", a sentence that implies the trait "smart") and rearranged sentences ("He took his first calculus course when he was 42 years old." a sentence that does not imply the trait "smart"). Results showed that participants made more errors (experiment 1), or took more time to provide a correct response (experiment 2), after reading trait-implying than control rearranged sentences.

Note, however, that Uleman and colleagues' paper (1996) is an exception and not the rule. As far as we know, none of the other papers (Ham & Vonk, 2003; Newman, 1991, 1993; Van Overwalle et al., 1999; Wigboldus et al., 2003, 2004) using the recognition probe paradigm had

rearranged controls, making it impossible to know whether they were dealing with real trait inferences or not. The control trials in many of the recent studies using the paradigm are neutral paragraphs that are non-related to the implying sentences. This kind of neutral control is not supposed to lead to the inference of the trait. However, it does not allow researchers to disentangle real inference from word-based priming effects.

The goal of the present Chapter is to present a set of material that would allow researchers to control for the impact of word-based priming effects on STI occurrence. We created a set of trait-implying behavioral descriptions. The control versions were created in such a way that the control sentence incorporates as many words from the trait-implying sentence as possible.

5.4 EXPERIMENT

5.4.1 *Pretest 1: Trait-implying sentences*

An initial pre-test was conducted with the purpose of obtaining trait-implying sentences illustrative of 223 personality traits. Two hundred and ninety-three subjects, from which 115 were males, took part in the pre-test. The average age of the sample was 29.31 years old. The pre-test was conducted online, using Qualtrics Survey Software and the participants were recruited using social media tools (*e.g.*, Facebook groups dedicated to data collection for social psychology research) and email invitations. Each participant was presented with 15 personality traits randomly chosen from the initial list of traits and their task was to generate an illustrative behavior for each of the presented traits. Participants were instructed to think about people they knew and to give concrete examples of their behaviors. Participants were also instructed to take maximum 1 minute and 30 seconds per trait, to avoid using adjectives, and to be as specific as possible in their behavioral descriptions.

Two independent judges analyzed the collected data. Each judge received half of the traits and the correspondent generated behavioral descriptions. The first step of the analysis consisted in eliminating answers that were not behavioral descriptions (*e.g.*, definitions of the traits or traits' synonyms) as well as redundancies between participants' answers. Then, the judges selected 2 or 3 behavioral descriptions that better illustrated each trait (*e.g.*, the sentence "Mostrou-se a favor

do casamento entre pessoas do mesmo sexo." was generated for the trait "aberta"), and in cases in which none of the descriptions were behaviors, the judges created a sentence for that trait. Traits with similar behavioral descriptions, usually synonyms, were grouped under the same trait label (*e.g.*, "chata" and "enfadonha"). This grouping resulted in a total of 154 traits and their corresponding behavioral descriptions.

Two new judges received half of stimuli each (77 pairs) and were asked to select the best behavioral description for each trait. At this point the main concern was to choose the sentence that implied the trait the most. These same two judges, also created rearranged control versions for each of the 154 trait-implying sentences. Critically, an effort was made to use all the words from the trait-implying sentences in their rearranged versions. In some sentences keeping exactly the same words was easier (*e.g.*, the rearranged sentence "A nota mais alta que conseguiram tirar na sua cadeira no semestre passado foi 14" of the trait-implying sentence "A nota mais alta que conseguiram tirar na sua cadeira no semestre passado foi 19.", that implies the trait "exigente") than in others (*e.g.*, the rearranged sentence "Para não ir sozinha, adiou a viagem para o próximo mês deste ano." of the trait-implying sentence "Este ano passou um mês a viajar sozinha." that implies the trait "aventureira").

5.4.2 *Pretest 2: Rearranged control sentences*

The 154 rearranged sentences were then presented to four new independent judges. Two of the judges (group A) were asked to write down the first word that came to their mind when reading the sentences and the other two (group B) were asked to evaluate to what extent the sentences presented were related to the critical traits by using a 9-point scale ranging from not related (1) to very related (9). If at least one of the judges from group A generated the target trait or a word that was related with the trait ("esperançoso" when the trait inferred was "optimista"), that pair of rearranged/implying sentences was excluded from the set (based on this criterion 14% of the material was excluded). In addition, from the remaining material, if both judges from group B rated the relation between the rearranged sentence and the trait with a value higher than 5, that pair of sentences was also excluded from the set (8% of excluded material based on this criterion). This resulted in 122 pairs of trait-implying/rearranged sentences which are presented in Table 9.

In this pre-test the inter-judges reliability in Group B was low, $ICC = .360, 95\,\%\, CI\, [.008, .577]$, and that motivated the following pre-test.

Table 9: Trait-implying and rearranged versions (in Portuguese).

Trait-implying (TI) and Rearranged (R) versions
Aberta
TI: Mostrou-se a favor do casamento entre pessoas do mesmo sexo.
R: Não se mostrou a favor do sexo antes das pessoas casarem mesmo.
Activista
TI: Recolheu assinaturas de todos os moradores da cidade para pedirem obras na escola primária.
R: Disse aos moradores que assinou o contrato para as obras na escola primária da cidade.
Agradecida
TI: Pagou a refeição ao desconhecido que veio atrás dela para lhe dar o casaco de que se tinha esquecido no metro.
R: Pagou a refeição e ia a sair quando um desconhecido lhe veio entregar o casaco de que se tinha esquecido na cadeira.
Agressiva
TI: Levantou a mão à empregada porque esta pediu-lhe para baixar o tom de voz.
R: Levantou a mão e com sua voz baixa fez um pedido à empregada
Alegre
TI: Estava a assobiar uma melodia muito conhecida no caminho para o trabalho.
R: Estava a passar uma melodia muito conhecida no caminho para o trabalho.
Ambiciosa
TI: Acabou de ser promovida a chefe mas já está a pensar no que tem de fazer para ser nomeada directora.
R: Pensou que acabariam por promovê-la a chefe mas o que fizeram foi nomeá-la directora.
Ansiosa
TI: Não conseguiu dormir nada de noite porque ia viajar no dia seguinte.
R: Dormiu a noite toda durante a viagem e estava fresca no dia seguinte.
Arrogante
TI: Olhou a pessoa de alto a baixo antes de lhe responder.
R: Ele é uma pessoa baixa e teve de olhar para cima para lhe responder.

Autoritária

TI: Disse ao filho que este ia começar a praticar natação mesmo sem ele querer.

R: Porque o filho queria tanto, disse-lhe que podia começar a praticar natação.

Aventureira

TI: Este ano passou um mês a viajar sozinha.

R: Para não ir sozinha, adiou a viagem para o próximo mês deste ano.

Barulhenta

TI: O vizinho veio queixar-se do volume da sua televisão.

R: Queixou-se ao vizinho do pacote volumoso em que vinha a sua televisão.

Bondosa

TI: Decidiu esquecer que a colega a prejudicou no exame e dar-lhe uma nova oportunidade.

R: Decidiu esquecer o exemplo da colega e fazer o exame das "novas oportunidades".

Calculista

TI: Foi à festa porque sabia que lá ia estar uma pessoa importante e podia dar jeito um dia destes conhecê-la.

R: Como era uma pessoa importante para si disse que podia ir à festa nesse dia.

Calma

TI: Encostou-se à poltrona com um chá a ouvir a sua música favorita.

R: Estava a dar a sua música favorita quando entornou o chá na poltrona.

Carinhosa

TI: Passou uma semana inteira em casa do primo que tinha tido um acidente.

R: Disse ao primo que passou uma semana em casa porque tinha tido um acidente.

Carismática

TI: Depois da sua palestra, várias pessoas vieram pedir-lhe um autógrafo.

R: Depois da sua palestra, várias pessoas vieram pedir-lhe os slides.

Chata

TI: Contou tantas vezes a história aos colegas que eles já a sabem de cor.

R: Porque os colegas pediram ele contou a história que já sabia de cor.

Ciumenta

TI: Não quer que a namorada se dê com outros rapazes que não ele.

R: Não se dá com aquele rapaz e com sua namorada porque eles não querem.

Cobarde

TI: Incitou os colegas a fazer greve e foi o primeiro a desistir quando chegou o chefe.

R: O chefe incitou os colegas a fazerem greve e ninguém chegou a desistir.

Coerente

TI: Disse que era contra cunhas e quando um primo lhe ofereceu um trabalho na câmara ele recusou.

R: Ele recusou a cunha do primo dizendo que já tinha um trabalho na câmara.

Confiante

TI: Foi ao concurso e apesar de haver mais de mil participantes ele estava certo de que ia ganhar.

R: Estava certo de que ganhar num concurso com mais de mil participantes havia de ser difícil.

Confiável

TI: Não contou a ninguém o que o colega lhe contou acerca do passado do seu pai na prisão.

R: Não contou a nenhum dos seus colegas que o seu pai esteve na prisão no passado.

Conflituosa

TI: Criou logo um escândalo durante a refeição só porque a pisaram no bar da faculdade.

R: Ele pisou o almoço o que criou logo um escândalo no bar da faculdade.

Controlada

TI: Bebeu uma cerveja e parou porque sabia que ia ter que conduzir.

R: Como sabia que já não ia ter de conduzir foi beber uma cerveja.

Cooperante

TI: Disse que ajudava a pagar o arranjo do elevador do prédio mas se todos os outros também o fizessem.

R: Fez os arranjos no prédio todo e disse que ia precisar de ajuda para arranjar o elevador também.

Crente

TI: Mesmo face à má época do seu clube, ele ainda acha que podem ter uma vitória.

R: Mesmo face a uma vitória do seu clube, acha que a má época ainda pode continuar.

Criativa

TI: Como não encontrava o livro para ler uma história ao filho, inventou uma e o filho gostou.

R: O filho gosta de histórias inventadas mas ela leu uma de um livro que encontrou.

Cruel

TI: Descobriu que uma das suas empregadas tinha sido toxicodependente e usou isso contra ela.

R: Descobriu que perto do seu emprego havia toxicodependentes.

Cuidadosa

TI: Transportou os copos novos devagar até casa sem partir nenhum.

R: Partiu um copo ao transportá-lo para a sua casa nova.

Culta

TI: Viu um erro de datas no documentário sobre a história militar do Império Romano.

R: Viu um documentário sobre as histórias e as datas dos maiores erros militares do Império Romano.

Curiosa

TI: Procurou informação sobre aquela espécie de cão estranha que viu passar na rua.

R: Na rua que estava à procura viu passar um cão de uma espécie estranha.

Dedicada

TI: Queria entregar a tese em primeira fase e por isso nas últimas semanas passou o tempo todo a escrever.

R: Teve o tempo todo para se inscrever na primeira fase mas só o fez na última semana.

Dependente

TI: Admitiu finalmente que não consegue viver sozinha e nem põe a hipótese de o fazer.

R: Admitiu finalmente ainda não saber se a melhor hipótese era viver sozinha ou não.

Desarrumada

TI: Demorou meia hora a encontrar a outra meia do par no seu quarto.

R: Vestiu a outra meia do par e em meia hora estava a sair do quarto.

Desastrada

TI: Trazia o tabuleiro e deixou cair tudo no chão porque olhou para a mesa de trás.

R: Algo tinha caído para o chão na mesa de trás mas ela não olhou porque trazia um tabuleiro.

Desleal

TI: Aceitou trabalhar para o rival do seu melhor cliente a troco de mais alguns dinheiro por mês.

R: Apercebeu-se que a empresa rival tinha melhores clientes e fazia mais dinheiro por mês.

Desligada

TI: Foi viver sozinha e desde então já vai para 3 meses que não liga aos pais.

R: Ligou aos pais e disse que desde então ia viver sozinha por 3 meses.

Desonesta

TI: Deu troco a menos, como quem se tinha enganado, para ver se o cliente não notava.

R: Recebeu troco a menos porque o cliente não notou que se tinha enganado.

Desorganizada

TI: Não apontou os dias das reuniões e acabou por trocar os dias todos.

R: Acabou por pedir para trocar os dias das reuniões e apontou-as todas na agenda.

Despreocupada

TI: Apesar de ter teste no dia seguinte, ela ainda foi à praia descontrair um pouco.

R: Não conseguiu ficar mais na praia a descontrair porque tinha um teste no dia seguinte.

Desrespeitadora

TI: Passou à frente de três pessoas da fila sem pedir autorização a nenhuma.

R: Pediu para passarem a frente porque ali não havia autorização para fazerem fila.

Discreta

TI: Entrou para buscar a sua mala e depois saiu sem ninguém dar por ela.

R: Depois de ir buscar a mala entrou mas voltou a sair porque não deu com ninguém.

Distraída

TI: Estava à procura dos óculos quando os tinha na sua própria cabeça.

R: Procurou os óculos e pô-los na sua própria cabeça.

Eficaz

TI: Estava a fazer dois trabalhos ao mesmo tempo e conseguiu boa nota em ambos.

R: Em ambos os trabalhos fizeram uma nota sobre o bom tempo que ia estar.

Egocêntrica

TI: Falou tanto de si e das suas férias que não teve tempo para saber como é que a amiga estava.

R: Esteve a falar com a amiga e soube que não tinham estado de férias ao mesmo tempo.

Egoísta

TI: Arranjou o exame dos outros anos mas não contou a ninguém.

R: Falaram-lhe de uns exames de outros anos mas ninguém os conseguiu arranjar.

Encorajadora

TI: Disse ao empregado para não desistir do trabalho mostrando-lhe os seus pontos fortes.

R: Disse que o ponto forte do empregado era mostrar trabalho.

Entusiasmada

TI: Gesticulava muito e falava quase sem respirar ao contar as aventuras na viagem à Ásia.

R: Respirou fundo e contou novamente as aventuras da viagem à Ásia quase sem gesticular.

Esquecida

TI: Voltou a casa para buscar o almoço que tinha deixado no frigorífico de manhã.

R: Deixou comida no frigorífico de manhã para almoçar quando voltasse para casa.

Estudiosa

TI: Deixou de ir a três festas para se preparar para o exame de química.

R: Depois do exame de química preparou-se para ir a três festas.

Exagerada

TI: Descreveu o robalo que pescou como se fosse um tubarão.

R: Descreveu o tubarão e o robalo que pescou.

Exigente

TI: A nota mais alta que conseguiram tirar na sua cadeira no semestre passado foi 14.

R: A nota mais alta que conseguiram tirar na sua cadeira no semestre passado foi 19.

Extravagante

TI: Quando acordou naquele dia, pintou o cabelo de roxo.

R: Quando acordou naquele dia o seu olho estava roxo.

Flexível

TI: Mudou os seus planos para se ajustar aos dos seus colegas.

R: Mudou o plano para melhor enquadrar os seus colegas na fotografia.

Formal

TI: Trata todas as pessoas por "você", mesmo quando lhe são próximas.

R: Disse que apenas iria tratar as pessoas que lhe são próximas.

Forreta

TI: Todos os presentes que ofereceu no Natal são comprados com pontos promocionais.

R: Todos se ofereceram para estar presentes nas acções promocionais deste Natal.

Fraca

TI: Ao ser confrontada com falsas acusações ficou tão afectada que nem se conseguiu defender.

R: Ao ver todas aquelas falsificações ficou tão chocada que não soube o que fazer.

Fria

TI: Não demonstrou afecto num momento em que o marido tanto precisava.

R: Precisava de um momento de afecto com o marido.

Gabarolas

TI: Afirmou várias vezes que tirou 18 no exame e não estudou nada.

R: Afirmou várias vezes que é difícil tirar 18 se não se estuda nada.

Gananciosa

TI: Ganhou uma herança considerável e ainda diz querer ganhar o Euromilhões.

R: Já tinha uma poupança considerável e ainda teve a sorte de ganhar o Euromilhões.

Gastadora

TI: Na primeira semana do mês já tem que pedir dinheiro aos amigos.

R: Tem um encontro de amigos já na primeira semana do mês.

Generosa

TI: A caminho de casa, ofereceu o seu jantar a um sem-abrigo.

R: Foi abordado por um sem-abrigo quando saiu do jantar em casa dos pais.

Habilidosa

TI: Construiu uma pirâmide de cartas com 50 centímetros sem deixar cair nenhuma carta.

R: A pilha de cartas sem nenhuma resposta na sua mesa atingiu já os 50 centímetros.

Hospitaleira

TI: Não se importou de dormir na sala para alojar bem as suas visitas.

R: As suas visitas não se importaram de ficar a dormir na sala.

Humilde

TI: Atribuiu o mérito ao grupo, quando foi ela que encontrou a solução para o problema.

R: Encontraram em grupo a melhor solução para o problema.

Ignorante

TI: Disse que África é um país que fica a sul de Espanha.

R: Disse que África fica a sul de Espanha.

Impaciente

TI: Perguntou 3 vezes à recepcionista se faltava muito para ser atendida.

R: Perguntou à recepcionista se era a terceira vez que essa pessoa faltava.

Imparcial

TI: Não deu razão ao irmão na discussão que este teve com o vizinho por causa da construção da varanda.

R: Percebeu que a razão da discussão que o irmão teve com o vizinho era a construção da varanda.

Impulsiva

TI: Despediu-se do trabalho durante uma discussão com o chefe.

R: Despediu-se do chefe após a discussão de um trabalho.

Incapaz

TI: Chumbou pela terceira vez no exame de condução.

R: Mudou pela terceira vez o exame de condução.

Incompetente

TI: Amputou a perna de alguém porque pegou no relatório médico de outro paciente.

R: Amputou a perna do paciente e foi escrever o relatório médico do mesmo.

Inconsistente

TI: Está sempre a corrigir os outros quanto aos seus hábitos alimentares mas depois só come fritos.

R: Está sempre a dizer aos outros para excluírem os fritos dos seus hábitos alimentares.

Indecisa

TI: Precisou de meia hora para decidir onde ia almoçar nesse dia.

R: Precisou de meia hora para conseguir almoçar nesse dia.

Ineficiente

TI: Perdeu 2 horas a fazer uma tabela que podia ter feito em 15 minutos se utilizasse o novo programa.

R: Perdeu 15 minutos a fazer uma tabela que iria demorar 2 horas se não utilizasse o novo programa.

Ingénua

TI: Não percebeu que toda a conversa e elogios eram porque o rapaz estava apaixonado por ela.

R: Percebeu que o rapaz por quem estava apaixonada não era muito dado a elogios e conversas.

Ingrata

TI: Nem se lembrou dos colegas que o ajudaram no trabalho naquele dia.

R: Ajudou um colega a lembrar-se do trabalho que tinha para fazer naquele dia.

Insegura

TI: Pediu aos colegas que confirmassem se estava a pensar e a fazer bem.

R: Confirmou com os colegas o que tinham a fazer e pensar.

Interessada

TI: Fez 3 perguntas durante a apresentação para melhor compreender o tema.

R: A terceira vez que apresentou o tema, compreenderam-no melhor e não houve perguntas.

Interessante

TI: Mesmo os mais sonolentos ficaram curiosos ao ouvi-lo falar sobre o tema.

R: Ficou mais sonolento à medida que ia ouvindo falar sobre o tema que até lhe despertava curiosidade.

Invejosa

TI: Ficou toda vermelha quando soube que a colega tinha tido melhor nota do que ela.

R: Ficou melhor ao notar que a colega tinha ficado tão vermelha quanto ela.

Irónica

TI: No final de um dia acidentado, disse ao amigo que o dia que não poderia ter corrido melhor.

R: No final de um dia acidentado, disse ao amigo que o dia poderia ter corrido melhor.

Irrealista

TI: Mesmo tendo um salário reduzido, acreditava que ia conseguir comprar um carro desportivo a curto prazo.

R: Mesmo tendo um salário reduzido, conseguiu num curto prazo comprar um carro desportivo.

Irrequieta

TI: Não consegue estar parada mais do que dois minutos, começa a bater o pé e a mexer as mãos.

R: Não consegue estar de pé mais do que dois minutos, sem que as pernas comecem a tremer.

Irritadiça

TI: Bateu com a porta da sala quando soube que teve só 13 no teste.

R: Bateu à porta da sala 13 para saber a sua nota no teste.

Machista

TI: Não contratou a pessoa com melhor currículo porque era uma mulher.

R: A mulher que contratou não era a pessoa com melhor currículo.

Mal-educada

TI: Respondeu à professora quando esta lhe chamou a atenção por causa do barulho que estava a fazer.

R: Por causa do barulho que estavam a fazer, não ouviu a professora quando esta chamou o seu nome.

Manhosa

TI: Perguntou algo que sabia de antemão para testar o seu colega.

R: Perguntou ao seu colega se sabia que estava em contramão.

Medrosa

TI: A sala ficou escura e ela agarrou-se logo ao braço da pessoa ao lado.

R: Ao seu lado na sala estava uma pessoa com uma camisola escura nos braços.

Melancólica

TI: Relembrou um acontecimento do tempo em que seu marido estava vivo.

R: Relembrou o tempo e onde vivia com o marido quando isso aconteceu.

Mentirosa

TI: Contou aventuras que um amigo teve numa viagem a África como tendo sido suas.

R: Quer viajar até África por causa das aventuras que uma amiga conta ter tido.

Mesquinha

TI: Cortou relações com o amigo por causa de 20 cêntimos.

R: Cortou 20 centímetros ao cabelo do amigo.

Mimada

TI: Não falou o dia todo com a mãe porque esta não lhe comprou a camisola que queria.

R: Passou o dia todo a tentar falar com a mãe para saber que camisola esta queria que lhe comprasse.

Misteriosa

TI: Não disse onde ia nessa noite e quando questionado fugiu à pergunta.

R: Perguntou para onde e de onde estava a fugir esse indivíduo de noite.

Namoradeira

TI: Mudou de namorada várias vezes no último ano.

R: Mudou de morada várias vezes no último ano.

Obediente

TI: Ele fez como o polícia pediu, saiu do carro, tirou as mãos dos bolso e pô-las na cabeça.

R: O polícia vinha de mãos nos bolsos, baixou a cabeça e pediu-lhe para sair do carro.

Optimista

TI: Acha que **2014** vai ser melhor em termos económicos.

R: Acha que em **2014** vai ser melhor sermos económicos.

Orgulhosa

TI: Apesar de ter percebido que aquele não era um bom sítio para fazer a festa não deu o braço a torcer.

R: Sem se ter apercebido, torceu o braço enquanto arrumava o sítio onde ia fazer a festa.

Ousada

TI: Foi dançar para a coluna da discoteca.

R: Partiu a coluna quando dançava na discoteca.

Passiva

TI: Soube que a amiga lhe tinha mentido mas não fez nada quanto a isso.

R: Não soube o que fazer à amiga que lhe tinha mentido.

Patriota

TI: Pendurou a sua bandeira à janela durante os jogos olímpicos.

R: Viu a bandeira dos jogos olímpicos a partir da sua janela.

Pensativa

TI: Passou uma tarde inteira a olhar para o mar.

R: Passou uma tarde inteira a tomar banho no mar.

Perita

TI: Foi chamada para dar pareceres sobre a sua área de especialização.

R: Foi chamada para fazer uma especialização fora a sua área.

Persistente

TI: É a quinta vez que se vai candidatar ao curso de medicina.

R: Disse que é a quinta vez que o curso de medicina vai excluir candidatos.

Pessimista

TI: Ainda nem fez o teste e já acha que lhe vai correr mal, apesar de estar preparada.

R: Apesar de achar que já estava preparada, o teste acabou por lhe correr mal.

Picuinhas

TI: As janelas ainda lhe pareciam sujas, mesmo depois de as ter lavado 3 vezes.

R: Lavou de uma vez as 3 janelas porque estas lhe pareciam sujas.

Pontual

TI: Chegou uns minutos antes da hora marcada ao local da reunião.

R: Acabou por marcar a hora da reunião mesmo ali no local.

Poupada

TI: Comparou os preços das várias marcas antes de fazer as suas compras.

R: Ao fazer as suas compras trouxe produtos de várias marcas e preços.

Preguiçosa

TI: Não se levantou do sofá para atender o telefone.

R: Levantou-se do sofá para atender o telefone.

Quieta

TI: Não se mexeu durante a aula toda.

R: Mexeu-se durante toda a aula de dança.

Racista

TI: Disse à filha que a deserdava se ela se casasse com um africano.

R: Disse à filha que gostava que ela se casasse com um vestido africano.

Rápida

TI: Tomou banho em 10 minutos e ainda chegou antes da hora marcada ao jantar.

R: Tomou banho 10 minutos antes da hora marcada para chegar ao jantar.

Rebelde

TI: Fugiu de casa sem os pais saberem para ir ao festival de verão.

R: Contou ao pais tudo o que deviam saber sobre o festival de verão.

Religiosa

TI: Precisou de se confessar 3 vezes nesse ano.

R: Confessou que precisou de ir ao casino 3 vezes nesse ano.

Respeitadora

TI: Apesar de as pessoas da reunião falarem todas ao mesmo tempo, ela esperou pela sua vez de falar.

R: Quando chega a sua vez de falar numa reunião, as pessoas falam todas ao mesmo tempo.

Rude

TI: Palitou os dentes à mesa de um restaurante muito elegante.

R: Teve uma terrível dor de dentes naquele restaurante muito elegante.

Simpática

TI: Aceitou substituir um colega do trabalho que queria visitar a avó no hospital.

R: Queria visitar a avó no hospital mas nenhum colega aceitou substituí-la no trabalho.

Simples

TI: Foi à festa sem maquilhagem e com roupa do dia-a-dia.

R: No seu dia-a-dia usa maquilhagem mas não roupa de festa.

Sincera

TI: Disse ao amigo para se conter porque estava a ser muito irritante.

R: Disse que tentou conter o amigo que estava a ser irritante.

Sociável

TI: Falou com o estranho que estava ao seu lado.

R: O estranho que estava ao seu lado falou para o motorista.

Solitária

TI: Passou a passagem de ano com o seu gato.

R: Comprou uma passagem para levar o seu gato.

Supersticiosa

TI: Bateu três vezes na madeira para evitar que acontecesse o que acabara de dizer.

R: Acabou por encomendar três ripas de madeiras como aconteceu nas últimas vezes.

Teimosa

TI: Sugeriram-lhe que usasse uma técnica mais moderna, mas ela continua a fazer as coisas à maneira dela.

R: Passou a usar uma técnica mais moderna que lhe sugeriram em vez de fazer as coisas à maneira dela.

Tímida

TI: Corou quando a chamaram ao centro do palco para lhe cantar os parabéns.

R: Chamaram-na ao centro do palco colorido para lhe cantar os parabéns.

Tolerante

TI: Pela terceira vez consecutiva esperou pela namorada mais de 30 minutos sem se zangar com ela.

R: Pela terceira vez consecutiva encontrou a namorada sem ter combinado com ela.

	Trabalhadora
TI: Ficou no escritório mais 1 hora, mesmo sabendo que não ganharia mais por isso.	
R: Ficou no escritório mais 1 hora, mesmo sabendo que não havia nada para fazer.	
	Vingativa
TI: Meteu-se com o namorado da amiga porque esta no passado tinha-se metido com um namorado seu.	
R: Meteu-se com a amiga dizendo que no passado se tinha metido com o seu namorado.	

5.4.3 *Pretest 3: Trait-implying and rearranged control sentences*

Four new independent judges were presented with half of the pairs of sentences (66 trait-implying sentences and 66 rearranged sentences) and the correspondent traits. For each sentence-trait pair, the judges were asked to indicate how well the trait described the person performing the behavior. They were instructed to use a scale ranging from 1 ("the trait does not describe the person at all") to 9 ("the trait describes very well the person"). Moreover, for each sentence, the judges had to indicate how easy it was to comprehend the sentence, again by using a 9 point scale (1 – not easy at all, 9 – very easy). Each judge was presented with one of the two sentences (the trait-implying or the rearranged) related to the same trait. Thus, 2 judges were presented with a set of sentences (Group C) and the other 2 with a different set (Group D). For the trait ratings, in both Group C, $ICC = .868$, 95 % CI $[.811, .907]$, and Group D, $ICC = .818$, 95 % CI $[.735, .874]$, we obtained a high inter-rater reliability. As expected, the ratings for how much the trait describes the person in the rearranged sentence is much lower ($M = 2.84$, $SD = 2.08$) than in the trait-implying sentence ($M = 8.19$, $SD = 1.07$), $t(121) = 22.87$, $p < .001$. For the comprehension, in both Group C, $ICC = .354$, 95 % CI $[-.042, .590]$, and Group D, $ICC = .386$, 95 % CI $[.120, .572]$, the inter-rating reliability was low. The comprehension ratings for the rearranged sentences ($M = 7.37$, $SD = 1.88$) were significantly lower than for the trait-implying sentences ($M = 8.68$, $SD = .70$) , $t(121) = 6.71$, $p < .001$. Even though different, the ratings for comprehension are above 5 in both conditions. Table 10 contains the average ratings for each pair of stimuli (trait-rearranged sentence and trait-implying sentence). When using this material,

we recommend researchers to select those pairs with similar levels of comprehensibility for the rearranged and the trait-implying sentences.

Table 10: Mean and standard deviation of the ratings of the trait-implying descriptions and the rearranged versions in pre-test 3.

	Trait-implying Sentence				Rearranged Sentence			
	Describes the actor		Comprehension		Describes the actor		Comprehension	
Traits	**M**	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
aberta	9	0	9	0	1	0	8.5	0.5
activista	8.5	0.5	9	0	7	2	5	2
agradecida	9	0	6.5	0.5	5	0	7.5	1.5
agressiva	8.5	0.5	7.5	1.5	1	0	6	1
alegre	5.5	0.5	8.5	0.5	5.5	0.5	6	2
ansiosa	8	1	9	0	1	0	9	0
arrogante	8.5	0.5	9	0	1	0	7.5	0.5
ateia	6.5	0.5	9	0	1	0	4.5	3.5
autoritária	7.5	1.5	8.5	0.5	3.5	1.5	8.5	0.5
aventureira	9	0	9	0	1	0	5	4
barulhenta	6	3	8.5	0.5	3.5	2.5	9	0
bondosa	8	0	9	0	1	0	4.5	3.5
calculista	8	1	7.5	0.5	3	2	6	3
calma	7	1	9	0	3.5	1.5	5.5	3.5
carinhosa	7	1	7.5	1.5	1.5	0.5	9	0
chata	9	0	9	0	1	0	7	2
ciumento	9	0	9	0	1.5	0.5	8	1
cobarde	8	0	9	0	1	0	4.5	3.5
confiável	9	0	8.5	0.5	4.5	0.5	9	0
conflituosa	8	0	9	0	5	4	1	0
controlada	7	1	9	0	7.5	0.5	9	0
cooperante	7.5	0.5	9	0	1.5	0.5	6.5	1.5

crente	6	1	9	0	3.5	1.5	9	0
criativa	8.5	0.5	9	0	1	0	8	1
cruel	9	0	7	2	3.5	1.5	8	1
cuidadosa	8.5	0.5	9	0	1	0	8.5	0.5
curiosa	8.5	0.5	9	0	5	1	8	1
dedicada	8.5	0.5	9	0	1	0	9	0
desarrumada	8.5	0.5	9	0	6	2	8.5	0.5
desbocada	9	0	9	0	1	0	6	3
desleal	8.5	0.5	9	0	4.5	0.5	9	0
desligada	8	0	9	0	5	4	5	4
desonesta	9	0	9	0	1.5	0.5	6.5	2.5
desorganizada	9	0	9	0	2.5	1.5	6	2
desorientada	7.5	0.5	8.5	0.5	6.5	2.5	8.5	0.5
despreocupada	8.5	0.5	9	0	1	0	9	0
desrespeitadora	4.5	4.5	9	0	3.5	1.5	5	2
discreta	9	0	8	1	1	0	4.5	3.5
distraída	9	0	9	0	6.5	1.5	6.5	0.5
eficaz	9	0	9	0	1	0	1.5	0.5
egocêntrica	9	0	8	1	2	1	8	0
egoísta	9	0	9	0	1	0	6	3
encorajadora	8	1	8	1	8	0	9	0
entusiasmada	9	0	9	0	1.5	0.5	5.5	1.5
esquecida	7	0	8	1	1	0	9	0
estudiosa	8.5	0.5	9	0	3	1	7.5	0.5
exagerada	9	0	9	0	7	0	9	0
exigente	5	3	9	0	4.5	3.5	9	0
extravagante	8	1	9	0	3	2	9	0
falsa	9	0	9	0	1	0	2.5	1.5
flexível	9	0	7.5	1.5	7	1	9	0

formal	8.5	0.5	9	0	1	0	3.5	2.5
forreta	5	3	9	0	4.5	3.5	8.5	0.5
fraca	5	2	9	0	4.5	3.5	8	1
fria	8	1	9	0	3.5	1.5	9	0
gabarolas	9	0	9	0	1	0	8.5	0.5
gananciosa	7.5	1.5	9	0	4	2	9	0
gastadora	9	0	9	0	1	0	6.5	2.5
generosa	9	0	8	1	3	2	9	0
habilidosa	9	0	9	0	1	0	7.5	1.5
hospitaleira	9	0	9	0	4	0	9	0
humilde	8	1	9	0	3	0	7.5	1.5
ignorante	9	0	9	0	1	0	9	0
impaciente	9	0	9	0	1	0	8	1
imparcial	5.5	1.5	6	2	3.5	1.5	9	0
impulsiva	8	1	9	0	1	0	4.5	3.5
incansável	6	2	8.5	0.5	5	0	6	0
incapaz	8	1	9	0	1.5	0.5	7.5	0.5
incompetente	8.5	0.5	6.5	1.5	1.5	0.5	9	0
inconsistente	8.5	0.5	9	0	1	0	9	0
indecisa	9	0	8	1	4.5	0.5	9	0
ineficiente	9	0	9	0	1	0	4.5	3.5
ingénua	7	1	9	0	1.5	0.5	9	0
ingrata	8.5	0.5	9	0	1	0	9	0
insegura	8.5	0.5	8	1	2	0	8.5	0.5
interessada	8.5	0.5	9	0	1.5	0.5	6	1
interessante	7	1	8.5	0.5	3	2	9	0
invejosa	9	0	9	0	1	0	6	2
irónica	7	2	5.5	2.5	1.5	0.5	9	0
irrealista	9	0	9	0	1	0	6.5	1.5

irrequieta	9	0	9	0	8.5	0.5	9	0
irritadiça	6.5	0.5	9	0	1	0	9	0
justa	8	1	8.5	0.5	6	3	9	0
machista	9	0	9	0	3.5	2.5	8	1
mal-educada	8	1	8.5	0.5	5	4	8.5	0.5
manhosa	9	0	9	0	1	0	7	2
medrosa	9	0	9	0	1	0	9	0
melancólica	6.5	1.5	9	0	6	2	6	3
mentirosa	9	0	9	0	1	0	9	0
mesquinha	9	0	9	0	2	1	8	1
mimada	9	0	9	0	4	3	9	0
misteriosa	8.5	0.5	9	0	1	0	4	3
namoradeira	8	1	8.5	0.5	3	2	9	0
obediente	9	0	9	0	2	1	4.5	3.5
orgulhosa	8.5	0.5	7	2	4	3	9	0
ousada	9	0	9	0	1	0	5	4
passiva	7	1	7.5	1.5	6.5	0.5	8.5	0.5
patriota	9	0	9	0	1	0	5	4
pensativa	6.5	2.5	9	0	3.5	2.5	9	0
perigosa	8	0	9	0	1	0	4.5	3.5
persistente	9	0	9	0	3	2	9	0
pessimista	9	0	9	0	1	0	9	0
pontual	7.5	1.5	9	0	5.5	2.5	9	0
poupada	9	0	9	0	3	2	9	0
preguiçosa	8	1	9	0	1.5	0.5	9	0
quieta	8.5	0.5	9	0	1	0	6	3
racista	9	0	9	0	3	2	8.5	0.5
rápida	9	0	9	0	5.5	0.5	7	2
rebelde	9	0	9	0	1	0	9	0

religiosa	9	0	9	0	1	0	7.5	1.5
respeitadora	8.5	0.5	9	0	1.5	0.5	6.5	2.5
rude	9	0	9	0	1	0	5.5	3.5
sensível	9	0	9	0	8	1	9	0
simpática	9	0	9	0	3	2	8	1
simples	6.5	2.5	6.5	2.5	8	1	9	0
sociável	8.5	0.5	9	0	2.5	1.5	5	2
solitária	8	1	9	0	5	1	9	0
supersticiosa	8.5	0.5	9	0	1	0	4.5	3.5
teimosa	9	0	7.5	1.5	1	0	9	0
tolerante	8.5	0.5	9	0	1	0	5	3
trabalhador	9	0	9	0	4.5	3.5	9	0
vingativa	9	0	9	0	3.5	2.5	5.5	0.5

5.4.4 *Discussion*

One of the goals of the present Chapter was to discuss the potential role of word-based priming in the occurrence of spontaneous trait inferences, an issue that has remained highly unexplored in STI research. Typically, in STI studies, trait-implying sentences are presented during encoding. During the test, different measures are used to detect whether participants had unintentionally inferred the implied trait at encoding. Evidence of STI has been extensively obtained. However, from the current literature is difficult to conclude whether previous results are being influenced by simple word-based priming effects. This is particularly problematic because trait-implying sentences frequently contain individual words that are strongly related with the trait, and that can activate the trait regardless of text-based inferences taking place. One efficient way of overcoming this confounder it by comparing the trait-implying sentences with control sentences containing the same words. Thus, the second goal of the Chapter was to create a set of trait-implying sentences and the rearranged versions of those sentences. The rearranged versions have similar

words to the trait-implying sentences and as such, word-based priming effects are controlled. Any differences in the performance found between the two sentences should be due to real inferences based on processing of the entire meaning of the sentences. We highly recommend the use of these pairs of sentences in future studies investigating trait inferences in Portuguese language. It should be noted that we consider the present Chapter as a first step in approaching this issue. Future studies might be necessary to further test the present material with larger samples, and regarding other variables, such as valence, familiarity, or ease of comprehension (since our results were not very clear regarding this measure).

After discussing the important role that word-based priming might have in the STI literature, our first recommendation for future studies is to include rearranged sentences, besides the trait-implying ones. However, even when including rearranged sentences, because the calculation of STI effects relies on the difference between critical and control sentences, we should guarantee beforehand that word-based activation does not play a significant role in the material. Not controlling the material for this aspect before the actual experiment carries the risk of underestimating the effects or obtaining false negatives (*e.g.*, strong associates can lead to a ceiling effect). Thus, our second recommendation is to test the material before the experiment, and only select those pairs of trait-implying and rearranged sentences where the activation of the trait is significantly stronger in the critical sentence when compared with its control version. In order to pretest the material, we suggest the use of activation measures such as the lexical decision task. In the lexical decision task, words and non-words are presented to the participants, and the activation of the trait can be accessed via the speed of the decision when the trait-implying sentence versus rearranged sentence precedes the presentation of the trait word. If the response to the trait is faster after reading the trait-implying version than after reading the rearranged version, then it can be deduced that participants generated a "true" trait inference from the trait-implying sentence, beyond any word-based priming effects.

Finally, an important aspect that is still unclear, is how word-based priming processes might differentially impact performance in different tasks used to study STI. For that, it would be necessary to clearly understand the mechanisms underlying word-based priming and text-based priming, and also how both processes interact during text comprehension. For example, if word-based priming is a shorter lived effect and text-based priming has a longer duration, then the

effect of word-based priming would be particularly problematic for activation measures as the probe recognition paradigm. In contrast, memory measures, as the savings in re-learning and the false recognition tasks, should be less affected, or maybe not affected at all, by this contamination. Unfortunately, although there are models that can be used to make predictions about the processes that underly lexical and text level effects (Kintsch, 1988; Sharkey & Sharkey, 1992), we still don't understand how lexical activation emerging from single words interacts with more elaborate and holistic processing of the text (Stafura & Perfetti, 2014). Thus, knowing to what degree this confounder influences performance in different STI tasks remains an empirical question, that should be carefully addressed in future studies.

In sum, in the current work we addressed the role of word-based priming in STI studies, we discuss how this issue might affect and distort the interpretation of previous findings, and finally, we propose specific steps that should be taken by future research in order to avoid this problem.

5.5 REFERENCES

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*(5), 840–856.

Claeys, W. (1990). On the spontaneity of behaviour categorization and its implications for personality measurement. *European Journal of Personality*, *4*(3), 173–186.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, *33*(4), 465–495.

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*(5), 434–447.

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

Keenan, J. M., Potts, G. R., Golding, J. M., & Jennings, T. M. (1990). Which elaborative

inferences are drawn during reading? a question of methodologies. In D. A. Balotta, G. B. F. d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (p. 377-402). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*(2), 163–182.

Kintsch, W., & Mross, E. F. (1985). Context effects in word identification. *Journal of Memory and Language*, *24*(3), 336–349.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, memory, and cognition*, *12*(1), 82–91.

Newman, L. S. (1991). Why are traits inferred spontaneously? a developmental approach. *Social cognition*, *9*(3), 221–253.

Newman, L. S. (1993). How individualists interpret behavior: Idiocentrism and spontaneous trait inference. *Social Cognition*, *11*(2), 243–269.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, *27*(4), 399–415.

Ramos, T., Garcia-Marques, L., Hamilton, D. L., Ferreira, M., & Van Acker, K. (2012). What i infer depends on who you are: The influence of stereotypes on trait and situational spontaneous inferences. *Journal of Experimental Social Psychology*, *48*(6), 1247–1256.

Sharkey, A. J., & Sharkey, N. E. (1992). Weak contextual constraints in text and word priming. *Journal of Memory and Language*, *31*(4), 543–572.

Stafura, J. Z., & Perfetti, C. A. (2014). Word-to-text integration: message level and lexical level influences in erps. *Neuropsychologia*, *64*, 41–53.

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.

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*(4), 377–394.

Uleman, J. S., Moskowitz, G. B., Roman, R. J., & Rhee, E. (1993). Tacit, manifest, and in-

tentional reference: How spontaneous trait inferences refer to persons. *Social Cognition*, *11*(3), 321–351.

Uleman, J. S., Rim, S., Adil Saribay, S., & Kressel, L. M. (2012). Controversies, questions, and prospects for spontaneous social inferences. *Social and Personality Psychology Compass*, *6*(9), 657–673.

Van Overwalle, F., Drenth, T., & Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or to the action? *Personality and Social Psychology Bulletin*, *25*(4), 450–462.

Wigboldus, D. H., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484.

Wigboldus, D. H., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition*, *22*(3), 292–309.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252.

Winter, L., Uleman, J. S., & Cunniff, C. (1985). How automatic are social judgments? *Jouranl of Personality and Social Psychology*, *49*(4), 904-917.

6

ACTIVATION IS NOT ALWAYS INFERENCE: TESTING THE INFLUENCE OF WORD-BASED PRIMING ON IMMEDIATE AND DELAYED TESTS OF SPONTANEOUS TRAIT INFERENCES

Diana Orghian, Tânia Ramos and Leonel Garcia-Marques

Submitted

Throughout the years, several studies have provided evidence for Spontaneous Trait Inferences (STI). Upon the presentation of a behavior, people spontaneously infer a trait about the actor performing the behavior. However, a methodological confound has remained overlooked in the literature: the distinction between word-based priming and text-based priming effects. The activation of a trait is considered a "real" inference when it results from the processing of the behavioral sentence as a whole (*i.e.*, text-priming). However, a trait can also be activated by intra-lexical associations with individual words in the sentences (*i.e.*, word-priming). The main goal of the present experiments is to clarify the influence of word-priming on some of the most popular STI paradigms and to shed light on how this confound has affected previous findings in the literature. In experiment 1, by using a lexical decision task, we tested trait-implying sentences and control versions of those same sentences for the activation of the target trait. In three additional experiments, we examined how word-based priming influences results on an immediate probe recall task (experiment 2), on a delayed false recognition task (experiment 3), and on an explicit trait judgment task (experiment 3). Results showed that when immediate tests of STI are used, word-based priming plays a considerable role and can lead to false conclusions about the occurrence of STI. However, when the testing is delayed or explicit, the role of this spurious

activation is negligible and the real STI can be more easily detected. Possible explanations for the findings are discussed.

6.1 INTRODUCTION

When we perceive someone's behavior we go far beyond their concrete behavior – we infer traits regarding their personality. Inferring traits about others serves fundamental social functions and it can be useful in many situations. For example, when we are deliberately trying to form an impression about a new candidate to the presidency, when we are perceiving the behavior of someone we have just met, or when we are trying to decide whether to turn down a second date or not. Based on trait inferences, we can predict what other types of behavior to expect from the people around us, and we can adjust our own behavior in function of those predictions. In other words, inferring traits is a powerful way to categorize and gain control (or at least perceived control) over the vast social world we live in. Moreover, research on spontaneous trait inferences (STI) has demonstrated that people infer traits from behavioral descriptions, even in the absence of intentions to form impressions (Winter & Uleman, 1984). If you read the following behavioral description "She won the chess tournament", you will spontaneously infer the trait "intelligent" during its comprehension or encoding.

Evidence in favor of STI occurrence has accumulated over the years at an impressive rate (for a review see Skowronski & Hartnett, 2008; Uleman, Newman, & Moskowitz, 1996; Uleman, Rim, Adil Saribay, & Kressel, 2012). A large part of the research has been dedicated to show that the trait inference occurs at encoding (as opposed to retrieval) and to explore the processes underlying the link between the inferred traits and the actors (*e.g.*, Carlston & Skowronski, 1994, 2005; Orghian, Garcia-Marques, Uleman, & Heinke, 2015; Todorov & Uleman, 2002). However, the processes that lead to the activation of the trait from the behavioral description has not been examined to the same extent. While the generalized assumption is that the trait is activated based on the meaning of the behavioral sentence as a whole, an alternative possibility is that the trait is being activated by specific words in the sentence, through word-based priming mechanisms. Going back to the previous example, in the sentence "She won the chess tournament", it is possible that the trait "intelligent" is being automatically activated by the word "chess", since the

two words are semantically related. We suspect that word-based priming is a common problem in STI studies because the use of strong associates might emerge as natural consequence of the structure of our cognition during the process of creation of the behavioral descriptions by the researchers. This means that some of the previous results showing STI can, in fact, be a result of associations between specific words in the behavioral sentences and the traits. Since this aspect has not been systematically controlled, the degree to which previous findings have been influenced by word-based priming remains unknown. The goal of the experiments described in this Chapter is to investigate to what extent word-based priming plays a role in both immediate and delayed measures of STI. In order to do that, we used a set of trait-implying sentences and control versions of those sentences that were created by using roughly the same words as the words in the trait-implying sentences but rearranged in such a way that they did not imply the traits anymore. Since these two sentences contain almost the same words, they are equated for word-based priming effects. Thus, any difference in the likelihood of inferring a trait from the two sentences should be due to sentence-based priming. In experiment 1 we applied a Lexical Decision Task to a set of 48 pairs of trait-implying and control sentences in order to distinguish the sentences where a text-based inference is detected aside from any word-based priming effects, from those where word-based priming overrides the detection of any inferences. Te result of this experiment was the creation of two sets of pairs of sentences: one that allowed for STI detection (more facilitation in the lexical decision task when the trait is preceded by the trait-implying sentence in comparison with the rearranged sentences) and a second set where text-based inferences are not detectable beyond word-based priming effects (*i.e.*, both rearranged and trait-implying sentences lead to a similar amount of facilitation). These two sets of material were then used in three more experiments: one where the measurement of the inference happened immediately after reading the behavioral description (experiment 2), one with a delayed measurement of STI (experiment 3) and one with an explicit trait judgment measure where no word-based priming is expected since the participants are explicitly instructed to infer personality traits (experiment 4). The present results are expected to shed light on the role of word-based priming on STI findings and to clarify the importance of controlling for this aspect when using different methods.

6.2 SPONTANEOUS TRAIT INFERENCES

Initial evidence for STI was provided by Winter and Uleman (1984) with the cued-recall task. In their studies participants memorized trait-implying behaviors and later had to recall the behaviors under different cueing conditions. Results showed that providing the implied trait as a cue improved the recall of the sentence compared with a no cue condition, and the trait was equally or more effective than strong semantic associates of words presented in the sentence. These findings were taken as evidence that participants had spontaneously inferred the traits during the behavior comprehension, and that the behavior and the corresponding trait had been encoded in memory together. As a consequence, when the trait is presented as a cue, it facilitates the retrieval of the behavior. In these studies, Winter and Uleman revealed some concern about how specific words in the initial sentences could distort their results. For instance, through a pretest, they intentionally chose actor nouns that did not lead to the generation of the critical traits. This was to prevent the trait (*e.g.*, "thrifty") to serve as a retrieval cue to the actor noun (*e.g.*, "businessman") at test, and indirectly facilitate the retrieval of the sentence as a whole (*e.g.*, "doubled his investment in business venture"). However, associations between the trait and the other words in the sentences were not systematically controlled (*e.g.*, "business").

Since Winter and Uleman's original findings, other tasks have been developed in order to investigate whether traits are spontaneously inferred during behavior encoding. These tasks can be divided into two categories: immediate tasks and delayed tasks. With immediate measures (also called activation measures) the trait inference occurrence is measured immediately after reading the sentence. The underlying logic is straightforward. If the trait was inferred during the comprehension of the behavior, then the activation level of the trait will increase after reading the sentence. The probe recognition task (McKoon & Ratcliff, 1986) is an example of an immediate measure that has been widely applied to the study of STI (Ham & Vonk, 2003; Newman, 1991, 1993; Ramos, Garcia-Marques, Hamilton, Ferreira, & Van Acker, 2012; Uleman, Newman, & Moskowitz, 1996; Van Overwalle, Drenth, & Marsman, 1999; Wigboldus, Dijksterhuis, & Van Knippenberg, 2003; Wigboldus, Sherman, Franzese, & van Knippenberg, 2004; Wang, Xia, & Yang, 2015). In this procedure, each behavioral sentence is immediately followed by a probe word. Participants have to indicate as fast as possible if the word was part of the sentence.

In critical trials, the trait-implying sentence is followed by the implied trait. It is assumed that if the trait was inferred during reading, it will be harder for the participant to correctly indicate that it was not present in the sentence. Other immediate measures used to study STI in the past include the lexical decision task (Zárate, Uleman, & Voils, 2001) and the word-stem completion task (Whitney & Williams-Whitney, 1990). Naming and Stroop tasks also measure the activation of inferences immediately after its occurrence, however, so far, they have been applied mainly to discourse comprehension research (*e.g.*, Dosher & Corbett, 1982; Forster, 1981; Potts, Keenan, & Golding, 1988).

In delayed measures of STI (also called memory measures), there is an interval between the reading of the sentence and the measurement of the trait inference. In other words, there is a delay between the moment the inference is generated and its measurement. In order to perform a delayed memory task, participants need to access their mental representation of the text or event. As examples of delayed measures in STI literature, Carlston and Skowronski (1994) developed the savings in re-learning procedure and Todorov and Uleman (2002) the false recognition task. In both cases, trait inferences are accessed after an interval of time, instead of immediately after the reading of the sentence. In both of these tasks the participant is first exposed to behavioral descriptions and photos of actors who performed those behaviors. In the false recognition paradigm, during a later memory test, participants falsely recognize the trait as being presented in the sentence previously paired with the actor. In the saving in re-learning paradigm there is an advantage in re-learning trait-actor pairs after participants have learned the trait-implying sentence/actor pairs first, when compared with new trait-actor pairs. These results are interpreted as the trait having been inferred during the reading of the behavioral sentence and linked to the mental representation of the actor. Several studies have provided support for these assertions (Carlston, Skowronski, & Sparks, 1995; Crawford, Skowronski, Stiff, & Scherer, 2007; Todorov & Uleman, 2004).

In sum, STI have been demonstrated across different laboratories and a variety of tasks. Previous findings suggest that STI is a robust and largely automatic phenomenon. They are observed when: concurrent tasks are imposed during behavior encoding (Todorov & Uleman, 2003; Wells, Skowronski, Crawford, Scherer, & Carlston, 2011); each behavior is present only for 2 seconds (Todorov & Uleman, 2003); behaviors are presented as distractors (Winter, Uleman, & Cunniff,

1985; Uleman, Newman, & Winter, 1992); under shallow processing conditions (Todorov & Uleman, 2003), and despite the fact that participants do not report any awareness of such inferences (Winter & Uleman, 1984; Uleman & Moskowitz, 1994). Importantly, the fact that STI have been demonstrated with both immediate and delayed measures suggests that traits are being activated while reading the behavior and are included in the cognitive representation of the event.

However, as mentioned before, very little is known about the mechanisms of text processing by which the trait is inferred. Specifically, whether the trait is being activated at the sentence-level, at the word-level or both. Since behavior descriptive sentences are pretested to imply traits, one obvious mechanism through which the trait can be activated is via text-level processing. In this case, the interpretation of the meaning of the sentence leads to the activation of the trait. An alternative mechanism involves word-based priming. It is surprising that this alternative mechanism has been so overlooked in the previous research on trait inferences, especially because the importance of controlling for word-based priming has been widely acknowledged in the study of inferences in the discourse comprehension field (*e.g.*, Forster, 1981; Keenan, Potts, Golding, & Jennings, 1990; Keenan & Jennings, 1995; Sharkey & Sharkey, 1992).

6.3 EFFECTS OF WORD-BASED PRIMING

Inference generation involves a close interaction between what is explicitly provided in the text or observed by the perceiver, and their general knowledge. Since the majority of the studies that explored the occurrence of inferences used textual descriptions of episodes, it is essential to understand the mechanisms that leads to this integration of information provided by the text and the prior knowledge of the subject. Keenan and colleagues (Keenan et al., 1990; Keenan & Jennings, 1995) proposed that, as a result of this interaction, an inference concept can be activated via two very different processes. One way is via text-based processing, that is, by inferring the concept from the meaning of the sentence as a whole. This process corresponds to the definition of inference according to which "an inference is the generation of (new) semantic information from (old) semantic information" (Rickheit & Strohner, 1985, p. 8). An alternative way of activating a concept, is through word-based priming mechanisms, that is, from reading related words in the sentence. This type of activation is not generally considered a "true" inference,

because it involves nothing more than a passive spreading of intra-lexical activation, without more elaborate semantic or propositional processing taking part.

In general, researchers studying inferences are interested in the first process and not in the second. Inferring a personality trait from a behavioral description clearly involves applying a higher order type of processing. The problem is that, although word-based and text-based inferences are different processes, their effects can be indistinguishable. In fact, word-based priming effects may perfectly mimic text-based inferences. Our argument is that ignoring the distinction between these two sources of activation may lead to erroneous conclusions about the occurrence of trait inferences. Specifically, evidence of intra-lexical activations can be wrongly taken as evidence of STI.

Keenan and colleagues (1990) offered two alternatives to control for word-based priming effects. The first consists in eliminating any word in the sentence that is highly related to the concept to be inferred. However, as the authors recognized, this is an extremely difficult task. Obviously, the meaning of the whole sentence, that will lead to the trait inference, must derive to a certain degree from the meaning of the individual words in the sentence and thus, we should expect some amount of association between the meaning of individual words and the trait.

A better alternative suggested by the authors is to create appropriate control versions of the implying sentences. These control versions should contain the same words as the implying sentences, but the words should be rearranged in such a way that the resulting sentence no longer implies the concept of interest. Because the two sentences are equated in terms of individual words, if the activation of the target trait is shown to be higher following the encoding of the implying sentence than following the encoding of the control version, then the activation has to be due to text-based inferences.

This solution has been frequently implemented in the discourse comprehension field (Calvo, Castillo, & Schmalhofer, 2006; McKoon & Ratcliff, 1986; Otten & Van Berkum, 2009). A representative example is the study of McKoon and Ratcliff (1986) in which they explored the occurrence of predictive inferences. In this study, a predicted word (“dead”) was presented immediately after a predicting sentence (“The director and the cameraman were ready to shoot closeups when suddenly the actress fell from the 14th story”) or after a rearranged sentence that contains all the words that might be related to the inference concept (“Suddenly the director fell

upon the cameraman, demanding close-ups of the actress on the 14th story"). Results showed that participants had more difficulty in rejecting the predicted word after reading the predictive sentence than after reading the control version. As the word-based priming is controlled, these discrepancies can only be explained by text-based inferences.

The word-based priming effects have been recurrently neglected in the STI literature. The exception is the research conducted by Uleman, Hon, Roman, and Moskowitz (1996). In this research, participants were presented with trait probes (*e.g.*, "lazy") after reading a trait-implying sentence (*e.g.*, "He drove to the newsstand, only half a block away") or after reading a control version created by using roughly the same words (*e.g.*, "He drove to the only newsstand, twenty blocks away"). In study 1, results showed no differences in reaction times (RTs) between conditions. However, significant differences between conditions were obtained in the error rates (but only in the first half of a long set of trials). In a second study, a feedback was introduced after each trial. In this case, RTs were reduced and differences emerged between trait-implying sentences and control sentences in the predicted direction (again only in the first half of the trials). To our knowledge, no other study in the STI literature has included rearranged sentences. Thus, it remains ambiguous how have previous results been affected by this confound.

For sake of clarity, we do not claim that all the previous results reporting STI are exclusively driven by intra-lexical associations. While it is extremely unlikely that word-based priming could account for all previous results obtained in the STI literature, we also believe that word-based priming is only a problem for some paradigms used in the field. This aspect is further discussed below.

6.4 WORD-BASED PRIMING ON IMMEDIATE AND DELAYED TESTS

STIs have been investigated with both immediate and delayed measures. The type of methodology chosen is an important matter, however one method cannot be considered "better" than the others. While some researchers would argue that it is essential to examine the state of activation immediately after the inference takes place (*i.e.*, on-line), others would claim that what is important is to access if the inference is incorporated in a more stable mental representation. Immediate measures tend to capture the process of inference generation, while delayed measures

capture the product of the inference process. In essence, immediate and delayed methods access the inferential process at different points in time, and thus, the two can be seen as complementary rather than conflicting.

Central to our discussion is that immediate and delayed methods may not be equally affected by word-based priming effects. To clarify this subject we need to know the time course of word-based priming on one hand, and the time course of text-based inferences on the other. Currently, there is no clear answer to this question (see Simpson, 1994). Nevertheless, there are both theoretical and empirical reasons to believe that word-based priming might be relatively short lived compared to text-level inferences.

According to Kintsch's Construction-Integration model of text comprehension (1988), when we read a sentence, each word activates, in parallel, concepts in our general knowledge network in a rather promiscuous way. As a consequence, the initial representation of the text includes several plausible concepts, including concepts that are "incorrectly" activated due to spreading activation mechanisms. In other words, the initial process of activation is context-free. This is seen as the price to pay for the tremendous flexibility of the comprehension system. The model further states that this initial noisy and incoherent text-representation goes through cycles of activation until it stabilizes on the representation that is more coherent with the context of the sentence. During this stabilization process, which according to the model corresponds to an integration phase, meaning-inappropriate concepts are deactivated and excluded from the text representation. Thus, on the basis of Kintsch's model, it is possible to predict that priming effects at the lexical level will impact performance on immediate tests, but not necessarily on delayed measures of STI where more stabilized representations are accessed.

There are also empirical findings that converge to the same prediction. Several studies, using ambiguous words, have found that all possible meanings of a concept are immediately activated as the word is encoded, however, after some time only the contextually appropriate meaning remains activated (Conrad, 1974; Kintsch & Mross, 1985; Lucas, 1987; Oden & Spira, 1983; Onifer & Swinney, 1981; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982; Tanenhaus, Leiman, & Seidenberg, 1979). These results suggest that this initial access to multiple meanings is automatic and context-independent. Text-based inferences, by contrast, would only intervene at a later stage of processing after the implausible meanings have been deactivated. Onifer and

Swinney (1981), for example, used a cross-modal priming technique to explore the effect of biased contexts on meaning activation of lexical ambiguous words. Participants attended to sentences that contain an ambiguous word (*e.g.*, "bank"), either in a context that favored the primary meaning of the words (*e.g.*, "All the cash that was kept in the safe at the bank was stolen last week when two masked men broke in."), or in a context that favored the secondary meaning of the word (*e.g.*, "A large piece of driftwood that had been washed up onto the bank by the last storm stood as a reminder of how high the water had actually risen."). Immediately after reading each sentence, words were visually presented for lexical decision. The word was either related to the primary meaning (*e.g.*, "money"), to the secondary meaning (*e.g.*, "river"), or control words. Results showed that lexical decisions were facilitated for words related to both the primary and the secondary meanings, regardless of the semantic context. However, when the presentation of the probe word was delayed for 1.5 seconds, facilitation was observed only for the word related to the context appropriate meaning. Thus, even when the primary meaning of a word is favored by the context, the secondary meaning still gets activated. It is only at a later stage, that the context leads to the selection of the most appropriate meaning.

Regarding the trait inference, if a trait-implying sentence and its rearranged version contain words highly associated with the target trait, we expect all the possible meanings of the sentences and of its words to be active (via word-based and via text-based activation) at an early stage of encoding. At this stage a large overlap of activation is expected between the trait-implying and the rearranged sentences (mostly due to word-based priming), and this overlap may be enough to cloak the extra activation coming from the inference in the trait-implying sentences. Later on, after the different elements in the sentence are integrated, only the relevant activations will remain, such is the case of the "real" inference. At this stage, after context-irrelevant and spurious meanings are deactivated, the overlap between the rearranged and the trait-implying sentences should be none or much lower than initially. Thus, it should be easier to detect the difference between the two sentences, *i.e.*, the true inference.

Our argument according to which word-based priming affects the performance on immediate tasks but, because it dissipates quickly, are innocuous on delayed tasks, finds support in previous research. But this view is far from linear, and is not without counterarguments. In fact, it is not possible to state definitively that the effects of activation based on individual words will

not have any influence on delayed tasks. There are several reasons for that. First, the so called exhaustive views of lexical access, including the construction-integration model (Kintsch, 1998), predicts that mainly concepts that are inconsistent with the context are inhibited over time, remaining in the text-representation only the meanings that are in line with the constrains of the context. However, in the case of trait-implying behaviors , the "problematic" words do not activate inconsistent meanings. Quite on the contrary, the intra-lexical activation is in line with the text-based inference – the trait is activated by both processes. For example, in the sentence "She won the chess tournament", the activation of the word "intelligent" is supported both by processing at the text-level and by word-level associations (*i.e.*, due to the presence of words like "won" and "chess"). The affinity between the activation spreading from individual words and the inference itself makes it plausible for the word-based priming to endure over time. Secondly, the duration of word-based priming effects may depend on the number, or proportion, of words in the sentence that are related with the same concept (*i.e.*, trait). Since trait-implying behaviors often have more than one word that is associated with the trait, it is important to consider cumulative effects of multiple words. Yet, little is known about the duration of priming resulting from multiple convergent words in the context of sentence processing. And finally, we don't know how lexical level processes interact with text level processes during comprehension, but these are certainly not independent processes. For example, intra-lexical associations may increase the likelihood of inferring the same trait concept at the text level. It seems reasonable to assume that activations at the lexical level could serve as a base for inferences at the text level, without which STI would remain invisible. In other words, word-based priming may have an indirect instead of a direct effect. All this uncertainty regarding the processes and duration of word-based priming effects, in the context of sentence processing, makes the need to control for this factor urgent.

In sum, there are two alternative views regarding the influence of word-based priming effects on immediate and delayed tasks. On one hand, it is plausible to predict that word-based priming affects immediate measures, but not delayed measures of STI. Both, the construction-integration model and numerous previous results report short durations for intra-lexical activations, which are consistent with this prediction. On the other hand, the duration of intra-lexical effects may be different for the STI due to the peculiarity of its materials, thus making it possible to speculate that word-based priming will affect performance on both immediate and delayed STI tasks.

Furthermore, the effect of word-based priming might depend not only on the time when the inference is measured, but also on the type of task participants are asked to perform. In other words, the distinction between immediate and delayed tasks may not be sufficient to predict differences of word-based priming effects across tasks. Specifically, when the task *explicitly* requires the participant to engage in semantic processing of the behavior, text-based inferences may override effects of lexical priming, even if the task takes place immediately after reading the sentence. This should be the case when participants are asked to make explicit trait judgments. For example, when asked to rate the actor of a behavior (*e.g.*, "She won the chess tournament") on a trait rating scale (*e.g.*, that goes from 1= unintelligent to 9 = intelligent), participants should be able to ignore residual activations derived from word-based associations, and focus on the relevance of the entire behavior to the trait being rated. Thus, according to our theorizing, regardless of the timing of the test, word-based activations should not be a major problem in explicit trait judgment tasks, as they are for the implicit measures. This hypothesis is in agreement with the literature on implicit and explicit measures. This literature shows that implicit measures provide a proxy for more automatic associations in memory and these are not necessarily reflected in explicit judgments (*e.g.*, Gawronski & LeBel, 2008). This is especially true when people have motivation to deliberate upon the material given (Gawronski & Bodenhausen, 2006) as we believe happens in explicit trait judgements.

6.5 OVERVIEW OF EXPERIMENTS

In experiment 1 we used a lexical decision task to indirectly measure the level of activation of critical traits after reading trait-implying sentences or control versions. The results of the lexical decision task allowed us to divide the stimuli sentences into two groups based on the difference in the activation of the trait after reading a trait-implying sentence versus a control version. In one group, the lexical decision about the trait-word was faster after reading the trait-implying sentence than after the reading the rearranged version of the same sentence. The pairs in this group were labeled as *strong pairs* since the evidence for text-based priming in these pairs is strong. Given that the words are identical or very similar in both sentences, the additional facilitation verified after reading the trait-implying behavior can only be explained by

trait inferences at text level. Thus, even if there is word-based activation in this group, it is not strong enough to override the detection of text-based priming. In contrast, for the second group of pairs of sentences, the trait-implying and the rearranged version lead to the same amount of facilitation in the lexical decision task. In these pairs, labeled *weak pairs*, the evidence for text-based priming is weak, because no inference is detected beyond word-based activation, meaning that this activation completely masks an eventual text-level inference.

These two groups of materials were used in three experiments, each using a different paradigm. The first goal was to test whether there is evidence of STI in both groups or only in the strong group. The second goal is to investigate how this interacts with the type of paradigms. Thus, first, we applied a probe recognition task (experiment 2) with the purpose of comparing the strong and weak groups of sentences in an immediate STI measure. If word-based priming plays a role in this measure, as we predict it will, then, a significant interference for trait probes should be observed in the trait-implying condition when compared to the rearranged one (*i.e.*, an STI effect), but only for the strong pairs. Next, we conducted an experiment using the same materials, but applying the false recognition task, a delayed measure of STI (experiment 3). If the effect of word-based priming does not fade away during the delay, then a STI effect should be observed only for the strong pairs and not for the weak pairs. On the other hand, if there is a decay of this type of activation over time, then it should be possible to detect STI even in the weak group of sentences. Finally, since our claim about the effect of word-based priming is strictly regarding implicit measures, and because we don't believe that this passive activation can override explicit judgments about personality traits, we don't expect a difference between strong pairs and weak pairs in an explicit trait rating task. The measurement in this experiment is immediately after the reading of the sentence, so if there would be a word-based priming effect we would be better able to detect it in this experiment than we would with a delayed judgement measure (experiment 4).

6.6 EXPERIMENT 1

In the first experiment, trait-implying sentences and their rearranged versions were presented to participants under memory instructions. After each sentence, a word was presented for a lexical decision task. The purpose of this experiment was to separate those sentences that lead to a

strong word-based activation of the trait from those where this confound is insignificant, making it easier to detect text-based inferences. The material tested in this experiment was also used in the next three experiments, in order to explore the role of word-based activation across different STI paradigms.

6.6.1 *Method*

Participants

108 undergraduate students participated in this experiment, 19 males and 89 females with an average age of 19.95 years old. The sample size was determined based on another experiment that was being conducted in the same session, and that experiment had a requirement of 100 valid participants minimum.

The data, not only in this experiment but in all the experiments in the present Chapter, were only analyzed when the samples were complete, and no additional data were collected afterwards.

Material

48 pairs of trait-implying behavioral sentences and their rearranged versions were selected from a previous study conducted by Orghian and colleagues (Orghian, Ramos, Reis, & Garcia-Marques, 2016). The rearranged sentences had as many words from the trait-implying sentences as possible, but did not imply the critical traits. There were also 96 neutral filler sentences that did not imply any traits (*e.g.*, "Bananas and mangoes are important sources of potassium and magnesium."). None of the neutral sentences conveyed behavioral information. These sentences were included in order to disguise the goal of investigating trait inferences. The 48 critical traits (*i.e.*, traits implied in the trait-implying behavioral sentences) served as target words in the lexical decision task. Besides the 48 critical traits, we also had 24 non-trait words (*e.g.* "sword") and 72 non-words taken from Domingos and Garcia-Marques' norms for Portuguese non-words (2013).

6.6.2 *Procedure*

The experimental design of this study is: 2 Type of sentence (trait-implying sentence versus rearranged sentence) $\times$ 2 Version (version A versus B of the material), with the second factor being the only between-subjects factor. The reaction time to decide if the string of letters was a real word or not was the dependent variable.

All the experiments in this manuscript were build and conducted by using Opensesame, a graphical experiment builder (Mathôt, Schreij, & Theeuwes, 2012). Participants were all recruited with ORSEE tool (Greiner, 2015) (except for experiment 3).

Participants completed the experiment in groups of maximum 8 people. As a cover story, participants were told that the study was about multitasking. Specifically, they were told that they had to memorize sentences and to perform an unrelated lexical decision task simultaneously. Immediately after each sentence, a string of letters was presented and their task was to decide as quickly and accurately as possible whether it was a word or a non-word. Each trial started with a 500 milliseconds fixation cross, then the sentence appeared on the screen until the participant pressed the space bar. After pressing the space bar, a string of letters was presented for lexical decision, and it remained on the screen until the participant gave the answer. Participants had to press the letter "b" key if they considered that the target was a word and the "n" key if they considered that the target was not a word. Participants started with 6 practice trials that were followed by 144 experimental trials.

In 24 trials of the 144 trials, subjects memorized trait-implying sentences and in other 24 they memorized rearranged sentences. These 48 trials were followed by the presentation of the correspondent traits. There were also 24 trials with neutral sentences followed by non-trait words and 72 neutral sentences followed by non-words. Note that for half of the trials the correct response was "yes", that is, to say that the string was a word (48 traits plus 24 non-trait words) and for the other half of the trials the correct response was "no" because the letter string was a non-word (72 non-words). Each sentence and each string of letters appeared only once to each participant.

From the 48 trials where the target given for lexical decision was a trait, half of the preceding sentences implied that trait, and the other half were rearranged versions of non-presented trait-implying sentences. Thus, the same participant never saw the two sentences (the trait-implying

and the rearranged) corresponding to the same trait. In order to do this, two versions of the experiment were created, and each participant was randomly assigned to one of them. In the first version, half of the traits (set 1) were preceded by sentences implying those traits, and the other half (set 2) was preceded by rearranged sentences. In the second version, the traits from set 1 in the first version were preceded by the corresponding rearranged sentences, while the traits from set 2 were preceded by trait-implying sentences. The experiment had a duration of approximately 15 minutes.

6.6.3 *Results and Discussion*

Only the trials where the target for lexical decision was a trait were analyzed. To assure than only the motivated people are included in this analysis, and because the success of the following experiments will depend on the result of this experiment, we decided to trim the less accurate 5% of the participants, that corresponds to 5 participants. The average accuracy of these 5 participants was 77.5%, whereas the accuracy of the remaining sample is 96.17%. Only the RTs of the correct responses are considered in the following calculations. As an attempt to exclude those answers that do not represent the most immediate response of the participants, RTs larger that 2500 ms were eliminated (1.44% of the correct responses). Because this study was conducted with the purpose of distinguishing between the material where the trait is more activated by the trait-implying sentence than by the rearranged version from those cases where this difference is null or in the opposite direction, all the analysis in the manuscript were conducted per item, as opposed to subject.

The mean RT for each trait in both conditions (when preceded by a trait-implying versus rearranged sentence) was computed. Next, the difference between the two versions was calculated for each trait (trait-implying minus the rearranged; see Appendix A for the statistics of each pair). And by using the median split based on these differences, two groups were obtained: a group with the most positive differences (ranging from 44 to -15 ms), that is, a group where the trait-implying sentence did not lead to facilitation when compared with the rearranged version, and a second group with the most negative differences (ranging from -16 to -169 ms), that is, a group where facilitation for the trait-implying sentence is detected when compared with

its rearranged version. As mentioned before, we call the first group the weak pairs group and the second the strong pairs group. As discussed in the introduction section, the reason why we think some of the pairs (the weak pairs) do not work in the expected way (*i.e.*, greater facilitation for the trait-implying than for the rearranged version) is due to the fact that word-based priming effects are overriding the trait inference effect. Thus, for the weak pairs, the trait inference is not revealed beyond word priming effects, whereas for the strong pairs the trait inference is strong enough to be detected regardless of intra-lexical priming effects. These two groups of sentences were used in the following experiments as a way to analyze to what extent word-based priming effects influence the results in different paradigms typically used in the STI literature. To certify the robustness of the conclusions obtained based on the median split, we performed a different split based on terciles. The analyses and the graphs for the three groups obtained are presented in Appendix B.

6.7 EXPERIMENT 2

In experiment 2 we wanted to investigate how the two groups of sentences would perform in an immediate measure frequently used in the STI literature, the probe recognition task. If the hypothesis that word-based activation influences immediate measures is true, then we should be able to observe a similar pattern to the one that was found for the lexical decision task (Experiment 1) with the probe recognition paradigm. In other words, a greater activation of the trait after reading the trait-implying sentences is expected when compared with the activation of the trait following the rearranged sentences, but only in the strong material group.

In the recognition probe task, participants read a sentence and immediately afterwards they are presented with a probe word. Their task is to indicate, as quickly and accurately as possible, whether the probe word was presented in the sentence previously seen. This is an immediate measure of STI because the trait inference is tested immediately after the reading of the sentence. Note that in this paradigm, larger RTs are associated with a greater activation of the trait probe (contrary to what happens in the lexical decision task where the shorter the RT the more activated is the trait). This means that if the probe is being activated by the sentence (while not being presented in the sentence), then it should take longer for the participants to provide a correct

response, because it would be more difficult to indicate that the probe word was not part of the sentence. We predict that participants will respond slower to a trait probe after reading a sentence that implies the trait than after reading its control version, and that would indicate that a text-based inference occured. However, this pattern should be observed only for the strong pairs of sentences, and not for the weak pairs. In Appendix B we present additional analyses where instead using the median to split the materials in two groups, we defined terciles.

6.7.1 *Method*

Participants

94 undergraduate students (32 males) took part in the study. Their average age was 23 years old. By using the G*Power 3.1 tool (Faul, Erdfelder, Lang, & Buchner, 2007), for a medium partial eta-squared of .06 and a power of .80, a minimum of 24 participants is required to detect the interaction between the type of material and the type of sentence. For a small effect size of .01 (.80 power) 138 participants are required. We tried to collect a sample size somewhere in between the two suggested samples, taking into account the abundance of participants in the lab at the time the experiment was conducted.

Material

The same 48 critical traits and the 48 pairs of sentences from Experiment 1 were used in this Experiment. 24 of these pairs correspond to the strong group and 24 correspond to the weak group. 56 filler trials were added. The filler trials consisted of neutral sentences and 56 non-trait words.

6.7.2 *Procedure*

The experiment had a 2 Type of sentence (trait-implying sentence versus rearranged sentence) $\times$ 2 Type of material (weak material versus strong material) design. All the factors were within-subject and the reaction time was the dependent variable.

Participants were told that this study aimed to investigate working memory abilities. Initially, participants completed 8 practice trials, that were followed by 104 experimental trials. Each trial started with the presentation of a fixation dot for 700 ms, followed by the presentation of a sentence. Participants were instructed to press the space bar when they finished reading the sentence. As they press the space bar, a 100 ms blank screen followed. Next, the probe was presented and remained on the screen until a response was given. Participants were instructed to press the letter "q" key on the keyboard if the answer was "no" (*i.e.*, to indicate that the word was not part of the previous sentence) and to press the "w" key if the answer was "yes" (*i.e.*, to indicate that the word was part of the previous sentence). In 32 of the trials the probe was actually presented in the sentence and they were non-trait words, while in the remaining trials (24 non-words trials plus the 48 trait trials) the probe was new, that is, it was not part of the sentence. As in Experiment 1, there were two versions of the experiment and only one was randomly assigned to each participant. If a trait was preceded by its trait-implying sentence in one version, it was preceded by its rearranged sentence in the other version, and vice versa.

6.7.3 *Results and Discussion*

The overall accuracy rate was 91%. Since there were no significant effects obtained for this dependent variable, accuracy results will not be mentioned in the following analyses. For the analyses regarding the reaction times, only the correct responses were considered. All the responses with RTs lower that 250 ms or larger than 2500 ms were eliminated (less than 1% of the data). A mixed effects ANOVA was conducted, with RTs per item as the dependent variable, type of sentence (trait-implying sentence versus rearranged sentence) as the within-item factor, and type of material (strong pairs versus weak pairs) as the between-item factor. A significant interaction was observed between the two factors, $F(1,46) = 11.90$, $p = .001$, $\eta_p^2 = .21$. For the strong pairs, longer RTs were observed for the trait-implying sentences ($M = 625$, $SD = 34$) than for the rearranged sentences ($M = 611$, $SD = 24$), $F(1,24) = 8.11$, $p = .009$, $\eta_p^2 = .25$. This result means that it was more difficult to reject the trait probe when it was preceded by a sentence that implied that trait than when it was preceded by a rearranged version. Thus, we found evidence for the occurrence of STI in the strong pairs condition. In the weak pairs, however,

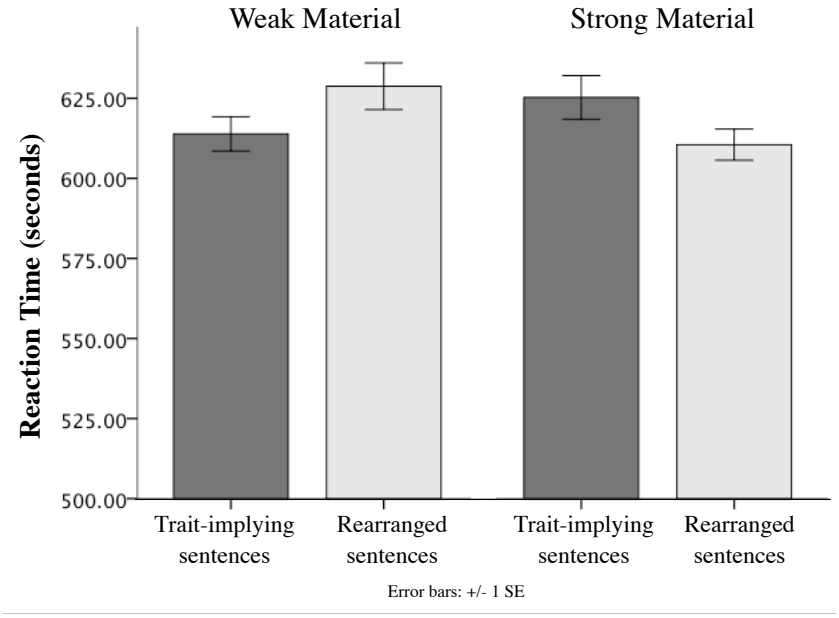

Figure 18: Mean RTs in function of type of material (weak or strong pairs of sentences) and type of sentence (trait-implying and rearranged) in Experiment 2.

there was no evidence of STI. In fact, the opposite pattern was observed, with longer RTs for the rearranged versions ($M = 629$, $SD = 35$) than for the trait-implying sentences ($M = 614$, $SD = 26$), $F(1, 23) = 4.57$, $p = .044$, $\eta_p^2 = .17$ (see Figure 18). Thus, for the weak pairs, the trait-implying sentences were not more likely to activate the trait than the rearranged versions. We don't have a definitive explanation for why the pattern reversed for the weak pairs but we suspect it might have to do with the extra analytical thinking required by the trait-implying sentences that might be interfering with the intra-lexical activations. A different but related explanation, has to do with the way the traits are being activated in the trait-implying sentence; it is very likely that we don't activate only one trait per sentence, but a semantic area (or multiple closely related traits) while the word-based priming only activates a unique trait. Thus, there might be more competition among the multiple traits activated by the trait-implying sentence than among the traits activated via word-based priming. In sum, as predicted, we found a strong interaction between the type of material and the type of sentence, with a larger interference in the trait-implying condition when compared to the rearranged condition in the strong pairs, but not in the weak ones.

6.8 EXPERIMENT 3

In Experiment 3, we used the same material used in the previous study, but applied to a delayed memory task commonly used in STI research, the false recognition task (Todorov & Uleman, 2002). Typically, in this task, participants see pairs of sentences and photos under a memory instruction. Each sentence describes a behavior performed by the person in the photo. In a delayed test phase, participants are presented with the same photos and each photo is paired with a trait word. The task is to indicate whether the word was present or not in the sentence that was paired with that photo in the study phase. In some trials, the trait presented is the one that was implied by the sentence paired with that photo (match trials), whereas in others, the trait presented was implied by a behavior previously paired with some other photo (mismatch trials). It is assumed that if the trait was inferred from the sentence, and was specifically linked to the right actor, then participants will exhibit a larger rate of false recognitions in the match than in the mismatch condition. Note that in the mismatch condition, a lower rate of false recognitions should be verified since no link was created between the presented trait and the person in the photo during the first part of the experiment. Instead of presenting only trait-implying sentences, as it is usually done in this paradigm, we also included control rearranged sentences. Again, our main goal was to explore the occurrence of STI in the strong and weak sentence pairs. If the influence of word-based priming is only a problem for immediate measures, then the type of sentence should not make a difference in this paradigm. That is, we should observe evidence of STI for both types of sentences, strong and weak. However, if word-based priming lasts long enough to affect the delayed test phase, then the pattern of results should be similar to the one in the previous experiment, that is, STI will only be detected for the strong sentence pairs.

6.8.1 *Method*

Participants

29 students (9 male) took part in the experiment. The average age was 17 years old. The participants were high school students visiting the university, and thus, the sample size was

defined by the number of visitors. Participant's parents gave their consent in what regards their children's participation in the current study.

Material

The same 48 pairs of sentences and their correspondent 48 traits used in our previous experiments were also used in the current study. Moreover, 24 of trait-implying sentences were used as fillers. In the filler sentences, the trait implied by the behavior was presented in the sentence (*e.g.* "She is so superstitious that knocked on wood after mentioning that something bad could happen"). These sentences were only included to ensure that the correct response to the question made at test (*i.e.*, whether the word presented was part of the sentence seen with that person) wasn't always "no".

6.8.2 *Procedure*

The experiment had a 2 Type of sentence (trait-implying sentence versus rearranged sentence) $\times$ 2 Type of material (weak material versus string material) $\times$ 2 Pairing (match trials versus mismatch trials) design. All the factors were within-subject and the false recognition rate was the dependent variable.

The experiment was conducted in the laboratory and each session had a maximum of 8 participants. Each participant worked individually on one computer with dividers separating their work stations. The experiment was presented as a memory study. Participants were told that the experiment consisted of two phases, a first phase where they would have to memorize pairs of photos of people and sentences with information about the people in the photos. The second phase was said to be a memory test of the material learned during the first phase (the type of test was not specified). Each trial started with a 500 ms fixation cross, followed by a photo presented in the middle of the screen and a sentence below the photo. Each pair was presented on the screen for 8 seconds. The study phase started with 4 practice trials. For each participant, our experimental software randomly chose 24 rearranged and 24 trait-implying sentences from the 48 critical trials, with the only criteria being that two versions of the same trait could not be presented to the same participant. Overall the first phase consisted of 72 experimental trials.

Besides the trials where the trait was implied by the sentence (48 critical trials), in one third of the 72 trials the trait was actually presented in the sentence (24 filler trials). The order of the trials was randomized for each participant and the pairing between the sentence and the photo was also randomized individually for each of them. After this phase, participants completed a 3 minutes distracting task. Next, participants completed the test phase. In the test they were presented with the previous photos, each one being paired with a trait. Their task was to indicate whether the word was part of the sentence presented in the learning phase with that person. They had to use one key (letter "s") to indicate "yes", and another key ("l") to indicate "no". Each trial in the test phase started with a 500 ms fixation cross, followed by the presentation of the trait and the photo, which remained on the screen until a response was given. After a response was given, a blank screen followed for 500 ms. This phase started with 4 practice trials, followed by 72 experimental trials. There were two types of critical trials in the test phase: the match trials, where the trait presented was implied in the sentence presented with that photo, and the mismatched trials, where the trait presented was implied in a sentence but the sentence was previously paired with a different photo. From the 48 photos in the critical trials, half (24) have been paired in the first phase of the study with rearranged sentences and from these 24, 12 were match trials and 12 mismatch trials. The same was true for the trait-implying sentences. The remaining 24 trials corresponded to the fillers. As mentioned before, for the fillers, the trait was actually presented in the sentence, so for those the correct response was "yes".

6.8.3 *Results and Discussion*

Only the critical trials were included in the following analysis. Again the material was analyzed in accordance with the groups created in Experiment 1. A mixed effects ANOVA was conducted, with the proportion of false recognition being the dependent variable, the type of sentence (implying sentence versus rearranged sentences) and the pairing (match versus mismatch) being the within-item factors and the type of material (strong group versus weak group) being the only between-item factor. A main effect of type of sentence was found, $F(1,46) = 9.81$, $p = .003$, $\eta_p^2 = .18$, with more false recognitions for the trait-implying sentences ($M = .37$, $SD = .19$) that for the rearranged versions ($M = .27$, $SD = .15$), indicating that the trait was more acti-

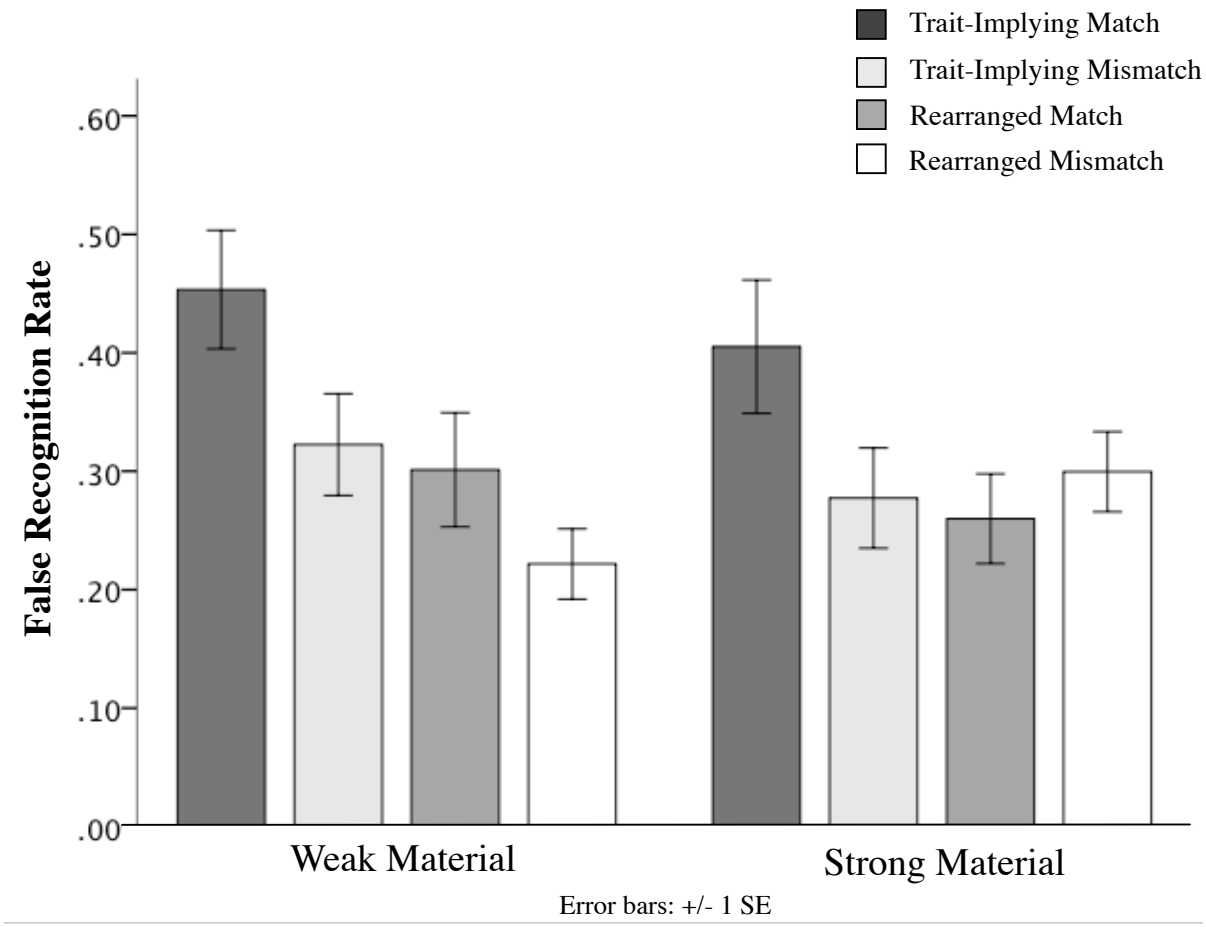

Figure 19: Mean RT in function of type of material (weak and strong), type of sentences (trait-implying and rearranged) and pairing (match and mismatch) in Experiment 3.

vated by the trait-implying sentences than by their control versions. A main effect of pairing was also observed, $F(1,46) = 9.03$, $p = .004$, $\eta_p^2 = .16$, with more false recognitions in the match condition ($M = .36$, $SD = .18$) than in the mismatch condition ($M = .28$, $SD = .12$), meaning the trait was inferred *and* linked to the person presented in the photo. A marginal interaction between pairing and type of sentence was also found, $F(1,46) = 3.70$, $p = .061$, $\eta_p^2 = .08$, with more false recognitions observed in the match conditions ($M = .43$, $SD = .26$) than in the mismatch condition ($M = .30$, $SD = .21$) for trait-implying sentences, $F(1,47) = 9.58$, $p = .003$, $\eta_p^2 = .17$. The same was not true for the rearranged sentences, since the rate of false recognitions in the match condition ($M = .28$, $SD = .21$) was not significantly different from

the rate of false recognitions in the mismatch condition ($M = .26$, $SD = .16$), as it can be seen in figure 19, $F < 1$. Moreover, there is no significant three-way interaction, $F < 1$, meaning that this pattern was similar in the weak and strong conditions. This result suggests that word-based priming effects are particularly impactful on immediate measures of STI, but not on delayed measures as the false recognition. In other words, independently of the type of material, we did find evidence of STI, indicating that the trait was inferred from the sentence and linked to the actor when the sentences was the trait-implying and not when the sentence was rearranged.

One may say that our conclusion is based on a null effect (the lack of a three-way interaction), however, while it is not surprising that people make inferences for the strong material, it is very informative that they do make inferences for the weak material and that is a non-null effect that strongly supports our hypothesis. More precisely, for the trait-implying condition in the strong group a significant inference was found, $F(1,23) = 4.40$, $p = .047$, $\eta_p^2 = .16$, with more false recognitions in the match condition ($M = .41$, $SD = .28$) than in the mismatch condition ($M = .27$, $SD = .21$). The same pattern was found for the trait-implying condition in the weak group, $F(1,23) = 5.01$, $p = .035$, $\eta_p^2 = .18$, with more false recognitions in the match condition ($M = .45$, $SD = .25$) than in the mismatch condition ($M = .32$, $SD = .21$). When the same comparisons are conducted for the rearranged conditions, no inference is found in the strong group, $F < 1$, with a similar amount of false recognitions in the match condition ($M = .26$, $SD = .19$) and in the mismatch condition ($M = .30$, $SD = .17$). A similar pattern is found for the weak group in the sense that the difference was not significant, $F(1,23) = 3.02$, $p = .095$, $\eta_p^2 = .12$, but the rate of false recognitions in the match condition ($M = .30$, $SD = .24$) seems to be higher than in the mismatch condition ($M = .22$, $SD = .15$) when the sentence does not imply the trait.

One may also argue that the sample size of this experiment is small, however we are not very concerned with it because the inference effect sizes in the false recognition paradigm are usually strong. For example, in the false recognition tasks reported by Todorov and Uleman the *r* varies from .72 to .87 (2003). In this same paper, for experiments very similar with ours, the sample size varies from 27 to 38 participants.

6.9 EXPERIMENT 4

Our concern with word-based priming effects is not strictly related with the timing of the measurement, it also has to do with the implicit nature of the measures. In order to show that this uniquely applies to implicit measures of inference, in this experiment, we asked participants *explicitly* to infer traits from behavioral descriptions. In this case the influence of word-based priming is expected to be very low or none, since people are consciously analyzing the sentence as a whole, with an explicit goal of inferring the trait.

6.9.1 *Method*

Participants

58 undergraduate students (12 male) took part in this study. Their average age was 22 years old. The sample size in this experiment was defined by an experiment that run in the same session, and thus, it was defined *apriori* but based on a arbitrary criteria in regard to the current experiment.

Material

The same 48 critical pairs of trait-implying and rearranged sentences (24 weak pairs and 24 strong pairs), and the correspondent traits were also used in this experiment. Moreover, no neutral sentences were used this time, since there was no need to disguise the purpose of the study to the participants.

6.9.2 *Procedure*

The experiment had a 2 Type of sentence (trait-implying sentence versus rearranged sentence) $\times$ 2 Type of material (weak material versus string material) design. All the factors were within-subject and the ratings corresponded to the dependent variable.

The experiment was presented to the participants as a pilot study to pre-test material for future experiments. There were 48 trials and each trial started with a fixation cross presented for 500

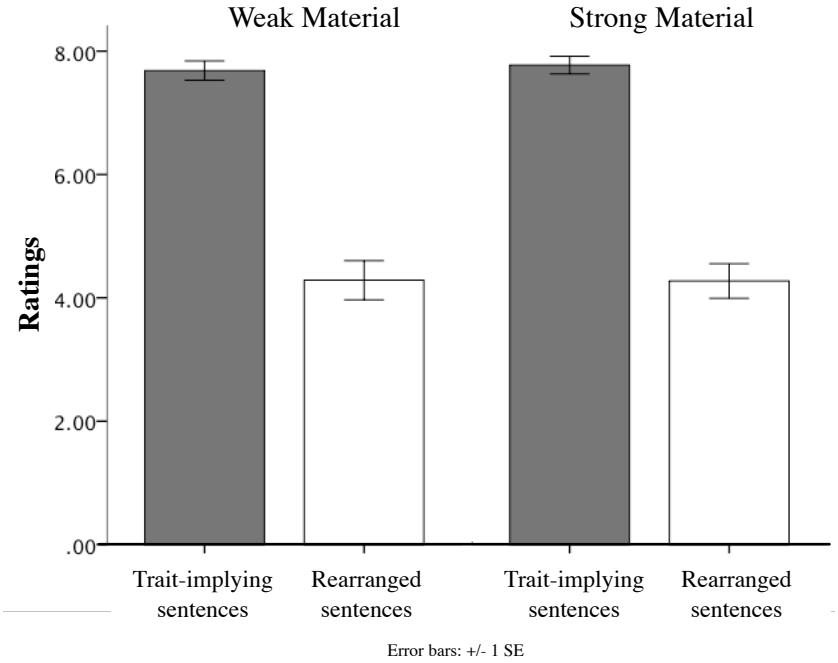

Figure 20: Mean rating in function of type of material (weak and strong) and type of sentence (trait-implying and rearranged) in Experiment 4.

ms, followed by the presentation of a sentence. On the same screen, the trait was presented with a rating scale ranging from 1 to 9. The participants' task was to indicate the extent to which the actor described in the sentences possessed the trait, on a scale ranging from 1 (not at all) to 9 (totally). Again, we had two different versions of the experiment, such that only one of the sentences in each pair was presented to each participant (either the trait-implying or the rearranged version).

6.9.3 *Results and Discussion*

A main effect of type of sentence was found, $F(1,46) = 194.63$, $p < .001$, $\eta_p^2 = .81$, with higher ratings in the trait-implying condition ($M = 7.73$, $SD = .72$) than in the rearranged one ($M = 4.28$, $SD = 1.44$). No other effects were observed, and critically as it can be seen in figure 20, there is no difference between the weak and strong material, $F < 1$. These results suggest that word-based priming has no role in explicit measures, even if they are immediate.

6.10 GENERAL DISCUSSION

The main goal of the present research was to test the impact of word-based activation on different measures of spontaneous trait inferences. To study STI, researchers usually create behavioral descriptions in the form of sentences or paragraphs. These descriptions imply traits, and the more they imply a trait, the better the material is considered to be. However, an overlooked aspect is the way the trait is being activated during the reading of the behavioral description in these stimuli. Keenan and colleagues (Keenan et al., 1990; Keenan & Jennings, 1995) suggested that there are two ways of a concept getting activated during the comprehension of a sentence. One based on pure associations between specific words in the sentence and the target concept to be inferred, and a second that derives from the processing of the sentence as whole. Only the second type can be considered a real inference, since the implied concept results from a combination of our knowledge with the global meaning of the sentence (that is, more than the sum of the meanings of the individual words). Thus, when in a sentence, both, word-based activation and sentence-based inference lead to the activation of the same trait and if appropriate controls are not used, it is impossible to tear them apart.

Unfortunately, the whole STI field is built on studies that do not take into account the existence of these two forms of trait activation. The consequence is obvious, we don't know whether the research on STI is actually about inferences. Thus, we strongly suggest, for future studies, the use of control materials that would allow us to distinguish the two sources of activation. A possible control is creating new sentences by rearranging the words in the trait-implying sentences in such a way that they don't imply the traits anymore. This is the approach we have followed in the present manuscript. By contrasting the activation of the trait in the trait-implying and in the rearranged sentence, we can conclude whether the effect is driven by a real inference or by pure word-based priming.

There is a large number of paradigms that were developed to study STI. These paradigms vary in the type of material used, in the time during which the inference is tested and in the type of dependent variables. The employment of one paradigm over another depends mainly on the objective of the researcher. However, different paradigms might be differently susceptive to confounders. Four different paradigms were investigated in the present research: the lexical

decision, the probe recognition (immediate measures), the false recognition (delayed memory measure) and an explicit trait rating measure.

Fortunately, if our conclusions are correct, the delayed measures of STI, that are frequently used in the field, are less affected by the word-based activation. This means that not controlling for such a confounder is less of a problem for these paradigms.

The same cannot be said about the immediate measures. Our results suggest that for immediate measures, like the lexical decision and the probe recognition paradigm, the word-based activation obscures the detection of real trait inferences. One possible explanation for this finding is the duration of the word-based priming effect. The time course of the word-based activation might be shorter due to dissipation, while the time course of the inference might be longer since it requires more elaboration and higher-level integration. As defended by the Construction Integration Model (Kintsch, 1988), there is an explosion of activations when a sentence is being read, but as the person continues to process the sentence and the different elements in the sentence are integrated, some information will be left behind. In addition, for this integration to be possible, and also in order to avoid incoherence, the person must access knowledge previously stored in their memory and combine it with the information being processed, selecting only the relevant information and incorporating it into a more global representation of the event.

One limitation of the current studies is the fact that when we compare, for example, the results in experiments 2 and 3, besides the timing of the measurement, the task also changes drastically. For a stronger argument in favor of our hypothesis that the timing of the measurement matters, we would need to conduct a study using the same paradigm and varying only the time when the inference is measured. Not all the paradigms can, however, be applied in delayed and immediate modes. Neither the lexical decision can be easily applied in a delayed mode, nor the false recognition can be applied in a immediate mode. A plausible candidate is the probe recognition paradigm that can actually be applied in both, immediate and delayed fashion (for more see McKoon & Ratcliff, 1986).

However, even without completely clarifying the aspect regarding the timing of the inference measurement, we believe the debate about the impact of the word-based priming on trait inference is a very necessary and urgent debate in the field.

Finally, we would also like to stress that besides the important methodological issue this manuscript discusses, this research has some relevant theoretical implications for spontaneous inference and inference overall. This represents an advance in our understanding of the difference between ephemeral semantic activation and true inferences that change our representation of the events over time.

In sum, word-based priming effects are likely to play a role in typical measures of STI, especially those where the effect tends to be measured at a very early point in respect to the encoding moment. With these studies, we hope to initiate a debate about the effect of word-based activation on the research done on inferences. And, as a consequence, we hope this leads us to an improvement and a refinement in the way trait inferences are studied methodologically and conceptually, so that our knowledge about how we perceive others becomes more and more precise.

6.11 REFERENCES

Calvo, M. G., Castillo, M. D., & Schmalhofer, F. (2006). Strategic influence on the time course of predictive inferences in reading. *Memory and Cognition*, *34*(1), 68–77.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*(5), 840–856.

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.

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: Ii. on the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, *69*(3), 420–436.

Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory and Cognition*, *2*(1), 130–138.

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.

Domingos, A., & Garcia-Marques, T. (2013). Normas de valência e familiaridade de "não-palavras" portuguesas. *Laboratório de Psicologia*, *6*(1), 49–74.

Dosher, B. A., & Corbett, A. T. (1982). Instrument inferences and verb schemata. *Memory and Cognition*, *10*(6), 531–539.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *The Quarterly Journal of Experimental Psychology*, *33*(4), 465–495.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731.

Gawronski, B., & LeBel, E. P. (2008). Understanding patterns of attitude change: When implicit measures show change, but explicit measures do not. *Journal of Experimental Social Psychology*, *44*(5), 1355–1361.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, *1*(1), 114–125.

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*(5), 434–447.

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

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

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*(2), 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Pressniversity press.

Kintsch, W., & Mross, E. F. (1985). Context effects in word identification. *Journal of Memory and Language*, *24*(3), 336–349.

Lucas, M. M. (1987). Frequency effects on the processing of ambiguous words in sentence contexts. *Language and Speech*, *30*(1), 25–46.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314–324.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, memory, and cognition*, *12*(1), 82–91.

Newman, L. S. (1991). Why are traits inferred spontaneously? a developmental approach. *Social cognition*, *9*(3), 221–253.

Newman, L. S. (1993). How individualists interpret behavior: Idiocentrism and spontaneous trait inference. *Social Cognition*, *11*(2), 243–269.

Oden, G. C., & Spira, J. L. (1983). Influence of context on the activation and selection of ambiguous word senses. *The Quarterly Journal of Experimental Psychology*, *35*(1), 51–64.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory and Cognition*, *9*(3), 225–236.

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition*, *33*(1), 20-66.

Orghian, D., Ramos, T., Reis, J., & Garcia-Marques, L. (2016). Acknowledging the role of word-based activation in spontaneous trait inferences. *Manuscript submitted for publication*.

Otten, M., & Van Berkum, J. J. (2009). Does working memory capacity affect the ability to predict upcoming words in discourse? *Brain Research*, *1291*, 92–101.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, *27*(4), 399–415.

Ramos, T., Garcia-Marques, L., Hamilton, D. L., Ferreira, M., & Van Acker, K. (2012). What i infer depends on who you are: The influence of stereotypes on trait and situational spontaneous inferences. *Journal of Experimental Social Psychology*, *48*(6), 1247–1256.

Rickheit, S. W., G, & Strohner, H. (1985). The concept of inference in discourse comprehension. In G. Rickheit & H. Strohner (Eds.), *Inferences in text processing* (pp. 3–49). Amsterdam: North-Holland.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, *14*(4), 489–537.

Sharkey, A. J., & Sharkey, N. E. (1992). Weak contextual constraints in text and word priming. *Journal of Memory and Language*, *31*(4), 543–572.

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

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 427–440.

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.

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.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*(4), 377–394.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, *66*(3), 490–501.

Uleman, J. S., Newman, L., & Winter, L. (1992). Can personality traits be inferred automatically? spontaneous inferences require cognitive capacity at encoding. *Consciousness and Cognition*, *1*(1), 77–90.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology*, *28*, 211–279.

Uleman, J. S., Rim, S., Adil Saribay, S., & Kressel, L. M. (2012). Controversies, questions, and prospects for spontaneous social inferences. *Social and Personality Psychology Compass*, *6*(9), 657–673.

Van Overwalle, F., Drenth, T., & Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or to the action? *Personality and Social Psychology Bulletin*, *25*(4), 450–462.

Wang, M., Xia, J., & Yang, F. (2015). Flexibility of spontaneous trait inferences: The interactive effects of mood and gender stereotypes. *Social Cognition*, *33*(4), 345–358.

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.

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

Wigboldus, D. H., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484.

Wigboldus, D. H., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition*, *22*(3), 292–309.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252.

Winter, L., Uleman, J. S., & Cunniff, C. (1985). How automatic are social judgments? *Jouranl*

of Personality and Social Psychology, *49*(4), 904-917.

Zárate, M. A., Uleman, J. S., & Voils, C. I. (2001). Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition*, *19*(3), 295-323.

6.12 APPENDIX A

6.12.1 *Material - Experiment 1*

Table 11: The 48 traits and the statistics corresponding to the differences in RTs in the lexical decision task between the trait implying-and the rearranged conditions. The difference corresponds to the RT(rearranged sentence) minus the RT(trait implying sentence). In bold are those that are part of the strong group whereas the remaining are part of the weak group.

Trait	t-test	df	p-value	M dif.	SE	95% CI
open-minded (aberto)	.545	99	.587	23	43	[-62, 108]
joyful (alegre)	-1.178	100	.242	-47	40	[-127, 33]
arrogant (arrogante)	.019	97	.985	1	47	[-92, 93]
adventurous (aventureiro)	-.170	95	.865	-9	51	[-110, 93]
kind (bondoso)	-1.066	98	-289	-51	48	[-145, 44]
calculating (calculista)	.237	101	.813	16	69	[-120, 152]
calm (calmo)	-.331	96	.741	-12	37	[-86, 62]
affectionate (carinhoso)	-.597	99	.552	-31	52	[-114, 72]
boring (chato)	-.057	101	.955	-3	57	[-117, 111]
jealous (ciumento)	-.520	98	.604	-29	55	[-139, 81]
coherent (coerente)	-.478	99	.634	-26	54	[-132, 81]
confident (confiante)	-1.253	99	.213	-55	44	[-143, 32]
cooperative (cooperante)	.209	94	.835	14	67	[-119, 148]
religious (crente)	.714	99	.477	44	62	[-79, 167]
knowledgeable (culto)	-.848	101	.398	-39	46	[-130, 52]
messy (desarrumado)	.642	99	.522	35	54	[-73, 142]

low-profile (discreto)	-.373	101	.710	-16	42	[-100, 68]
effective (eficaz)	.032	99	.974	1	37	[-71, 74]
enthusiastic (entusiasmado)	-.617	101	.538	-37	59	[-154, 81]
demanding (exigente)	.791	100	.431	39	50	[-59, 138]
flexible (flexível)	.332	100	.741	13	39	[-64, 90]
cold (frio)	-1.935	101	.056	-85	44	[-172, 2]
greedy (gananciosa)	-.404	99	.687	-24	59	[-140, 93]
waster (gastador)	-.736	93	.464	-49	66	[-179, 82]
impatient (impaciente)	.434	100	.666	24	55	[-86, 134]
incapable (incapaz)	-.619	100	.537	-28	45	[-116, 61]
inconsistent (inconsistente)	.325	92	.746	24	74	[-123, 172]
ungrateful (ingrato)	-.648	96	.519	-43	66	[-173, 88]
interesting (interessante)	-.212	99	.832	-11	54	[-118, 95]
sarcastic (irónico)	-.083	100	.934	-5	60	[-124, 114]
unrealistic (irrealista)	-989	100	.325	-78	79	[-233, 78]
sexist (machista)	.062	98	.950	4	59	[-114, 121]
sly (manhoso)	-1.697	96	.093	-108	64	[-235, 18]
liar (mentiroso)	.083	99	.934	-5	63	[-129, 119]
niggardly (mesquinho)	-1.892	97	.061	-94	50	[-192, 5]
mysterious (misterioso)	-.519	101	.605	-30	58	[-146, 85]
optimistic (optimista)	.101	101	.920	5	47	[-89, 98]
passive (passivo)	-2.631	99	.010	-159	61	[-280, -39]
patriotic (patriota)	-.167	100	.868	-7	44	[-95, 80]
punctual (pontual)	-.984	99	.328	-46	46	[-138, 46]

economical (poupado)	-1.236	97	.220	-67	54	[-174, 41]
racist (racista)	-2.253	101	.026	-169	75	[-318, -20]
fast (rápido)	-.342	99	.733	-15	45	[-104, 74]
respectful (respeitador)	.397	99	.692	26	65	[-104, 155]
nice (simpático)	-.033	101	.974	-1	38	[-77, 75]
social (sociável)	-1.003	101	.318	-45	45	[-134, 44]
stubborn (teimoso)	.032	99	.974	1	56	[-109, 113]
shy (tímido)	-1.507	101	.135	-70	46	[-162, 22]

6.13 APPENDIX B

6.13.1 *Additional Analyses*

Additionally to data based on the median split that we presented in the body of the article, in order to show that our conclusion are robust and do not depend on this specific method of obtaining the groups, we partitioned the data in three groups, each containing one third of the data. Thus, in Group 1 (weak) the difference in RTs (trait-implied minus the rearranged) varied from 44 to -3 seconds, in Group 2 (intermediate) from -3 to -37 and in Group 3 (strong) from -39 to -169.

6.13.2 *Experiment 2*

The split of the material based on terciles were used to conduct a mixed ANOVA. No main effect of type of sentence was detected, $F < 1$, but a significant interaction between the type of sentence and the group, $F(2,93) = 4.440$, $p = .014$, $\eta_p^2 = .09$, was found. Next, three separate repeated measures ANOVA were conducted, one for each group.

In Group 1, and as represented in figure 21, a significant difference between the two types of sentences was found, $F(1,31) = 4.236$, $p = .048$, $\eta_p^2 = .12$, with lower RTs in the trait-implying conditions ($M = 613$, $SD = 26$) than in the rearranged one ($M = 624$, $SD = 23$).

In Group 2, no significant difference between the two types of sentences was found, $F(1,31) = .048$, $p = .828$, $\eta_p^2 = .002$, with similar RTs in the trait-implying conditions ($M = 623$, $SD = 30$) and in the rearranged one ($M = 625$, $SD = 40$).

In Group 3, a significant difference between the two types of sentences was found, $F(1,31) = 11.293$, $p = .002$, $\eta_p^2 = .27$, with larger RTs in the trait-implying conditions ($M = 623$, $SD = 35$) than in the rearranged one ($M = 610$, $SD = 25$).

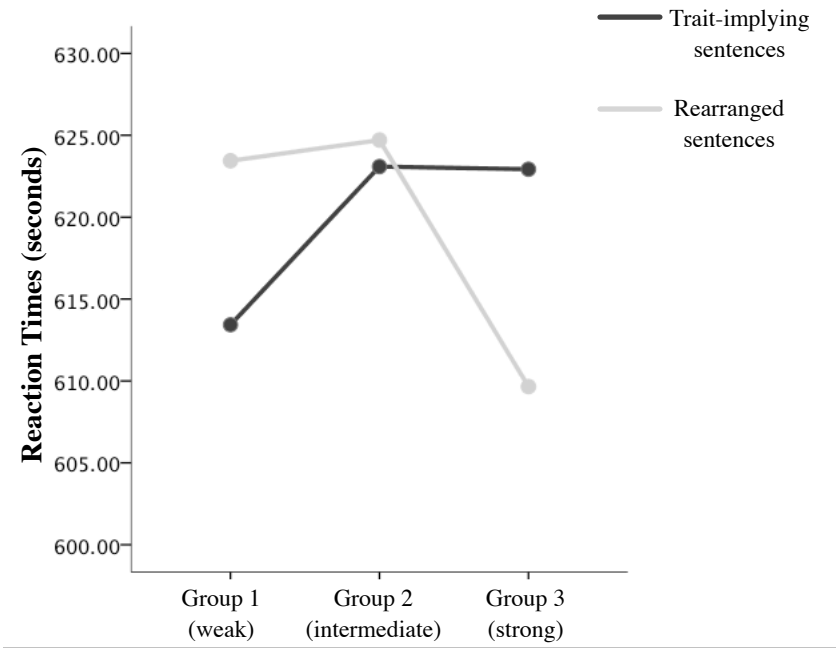

Figure 21: Mean RT in function of type of material (weak, intermediate and strong pairs of sentences) and type of sentence (trait-implying and rearranged) in experiment 2.

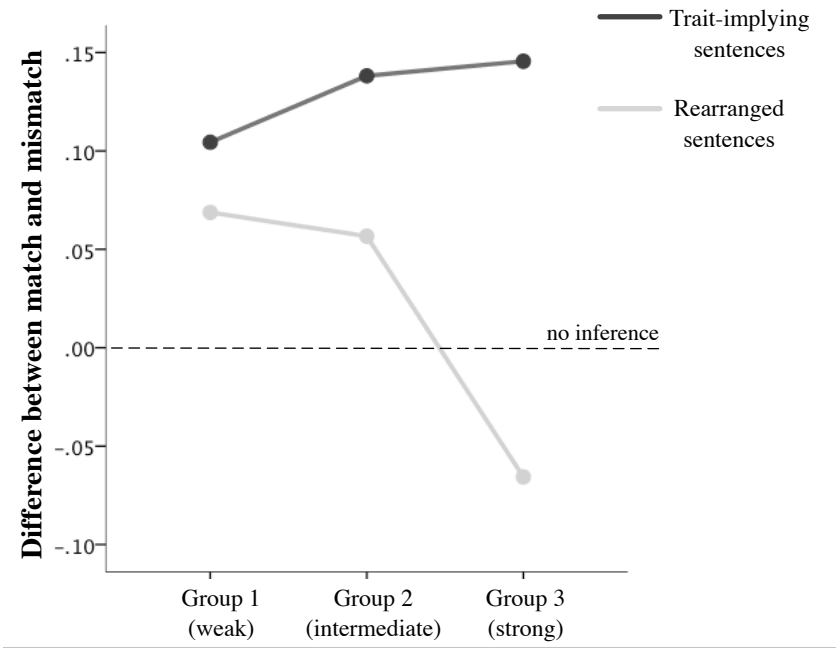

Figure 22: Difference between match and mismatch in the false recognition rate in function of type of material (weak, intermediate and strong pairs of sentences) and type of sentence (trait-implying and rearranged) in experiment 3.

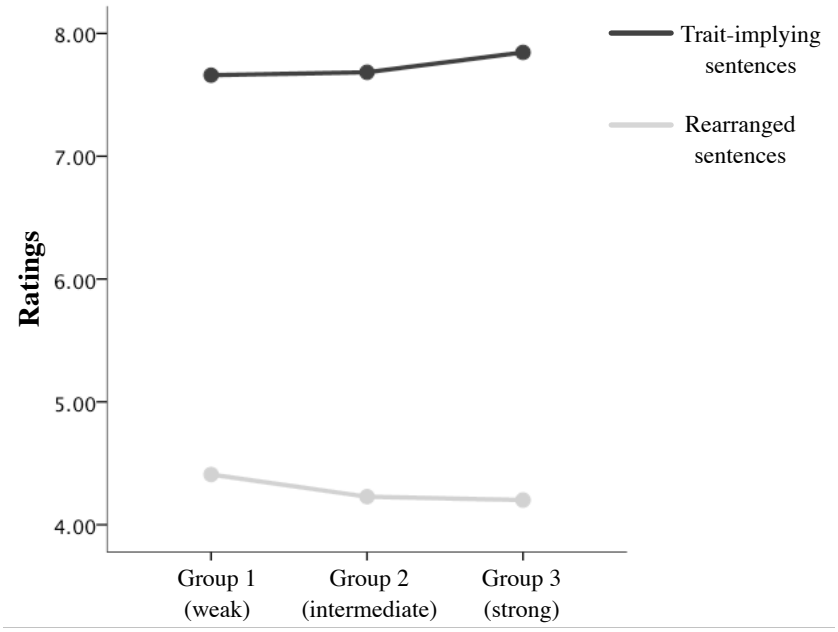

Figure 23: Mean rating in function of type of material (weak, intermediate and strong pairs of sentences) and type of sentence (trait-implying and rearranged) in experiment 4.

6.13.3 *Experiment 3*

For this experiment, two mixed ANOVAs were performed, one for the trait-implying material and one for the rearranged. For the trait-implying condition, a main effect of paring was found, $F(1,45) = 9.21$, $p = .004$, $\eta_p^2 = .17$, with higher false recognition rate in the match condition ($M = .43$, $SD = 26$) than in the mismatch condition ($M = 30$, $SD = 20$). No other effects were found, $F < 1$.

When the same analysis is conducted for the rearranged sentences, no significant effect was found, with the interaction between the pairing and the material's group showing the higher F value, $F(1,45) = 1.69$, $p = .195$, $\eta_p^2 = .07$, as suggested by the the figure 22.

6.13.4 *Experiment 4*

As represented in figure 23, main effect of type of sentence was found, $F(1, 45) = 191.99$, $p < .001$, $\eta_p^2 = .42$, with higher ratings for the trait-implying sentences ($M = 7.73$, $SD = .73$) than for the rearranged ($M = 4.28$, $SD = 1.45$). No other effects were found, $F < 1$.

7

CAPTURING SPONTANEOUS TRAIT INFERENCES WITH THE MODIFIED WORD ASSOCIATION TASK

Diana Orghian, Anna Smith, Leonel Garcia-Marques, and Dietmar Heinke

Submitted

After realizing that word-based priming might be affecting many of the paradigms used to investigate STI, a more systematic analysis of the different types of measures and their characteristics was conducted. In the current Chapter, we discuss the main existing paradigms used to study spontaneous trait inference. The main focus is on comparing activation and memory measures and in introducing the distinction between data-driven and conceptually-driven tasks.

Finally, a novel measure is proposed as a way of overcoming the limitations of, simultaneously, memory measures and their contamination problem and activation measures that are mainly dependent on data-driven performance. Our measure is based on the modified free association task proposed by Hourihan and MacLeod (2007) and presented by the authors as a pure conceptual implicit memory measure. In the present experiments we used the modified free association task to detect spontaneous trait inference and also to compare its results with the naming task and the modified Stroop task. We find this measure to be more reliable and more appropriate to study an elaborate phenomenon as STI. Moreover, we demonstrate that it can be useful to investigate other debates in the STI literature such the difference between STI and STT (discussed in the first four Chapters of this dissertation).

7.1 INTRODUCTION

Spontaneous trait inference (STI) is a social inference that occurs when, for example, the man who holds the door for someone is categorized as "friendly" or the woman who jumps the queue as "rude". These personality impressions shape the way we interact with others – perhaps avoiding the rude woman or seeking out the friendly man. Winter and Uleman (1984) provided the first empirical evidence for STI when participants in their study unintentionally inferred personality traits after reading behavioral descriptions. The researchers have shown that trait inference is a spontaneous process; occurs outside of awareness (Lupfer, Clark, & Hutcherson, 1990); occurs at encoding (*e.g.*, Carlston & Skowronski, 1994; Winter & Uleman, 1984; Uleman, Newman, & Moskowitz, 1996); is maintained across more demanding experimental settings including short stimulus presentation timings and cognitive load (*e.g.*, Carlston & Skowronski, 1994; Todorov & Uleman, 2003) and lingers after explicit recall has faded (Carlston, Skowronski, & Sparks, 1995).

STI has most commonly been investigated by using one of four paradigms: cued-recall (Winter and Uleman, 1984; Winter et al., 1985), probe recognition (Uleman, Hon, Roman, & Moskowitz, 1996), savings in relearning (Carlston & Skowronski, 1994), and false recognition (Todorov & Uleman, 2002, 2003, 2004). As discussed later, all of these paradigms share a characteristic, they are all memory based, *i.e.*, they require the subject to recollect a past event and contrast a probe against this recollection or, to recall the past event with the help of a cue.

Over time, the paradigms have been refined due to new concerns regarding what an inference is, when it occurs, or whether is it actually spontaneous. A paradigm is considered a good one when a set of requirements are fulfilled. 1) A paradigm that measures a spontaneous process implicitly should direct participant's attention to the behavioral information without encouraging explicit trait impression goals (Uleman, 1999). To accomplish this requirement, participants are usually instructed to familiarize themselves with the material (*e.g.*, Skowronski, Carlston, Mae, & Crawford, 1998) or to memorize it (*e.g.*, Winter & Uleman, 1984; Todorov & Uleman, 2002). Note, however, that even if the instruction is to memorize the material and not explicitly form impressions, explicit inferences can still be used in a strategic way in order to help the performance in the memory task, and even more so when the inference actually can help the performance in the

memory task. This is especially true when all the material used are trait-implying descriptions making it easier for the subject to become aware of the researcher's goal in studying impressions. 2) The inference should be measured upon the unconscious recollection of earlier experience, as opposed to explicit recollection (Schacter, 1987). This is also known as the "contamination problem" of the implicit measures with explicit recall (Jacoby, 1991). 3) Related with point 2, the paradigm should discourage strategies likely to induce inference at the moment of retrieval and uniquely measure inference at encoding (Keenan, Potts, Golding, & Jennings, 1990; McKoon & Ratcliff, 1986; Wyer & Srull, 1986). 4) It should assure that the activation of the trait is a consequence of text-based priming and not word-based priming. In other words it should assure that the inference is a result of processing the meaning of the behavior and not a result of reading specific words in the sentence that are individually associated with the trait (Orghian, Ramos, & Garcia-Marques, 2016). 5) It should be a task that is sensitive to top-down phenomena, meaning it should be able to detect concepts resulting from meaning-based processing. Roediger and Blaxton named these kind of measures *conceptually-driven tasks* (1987). Using a conceptually-driven task is important because trait inference occurs via conceptually-driven mechanisms that allow new information to be generated based on the interpretation of the meaning of the behavior. And finally 6) the measure should be possible to apply in a delayed mode, meaning the inference should be possible to be measured with some delay in respect to sentence reading. That will allow us to detect inferences that occur later on in the integration process (Kintsch, 1988).

In the following section we describe the paradigms used in STI literature and some others used in the text comprehension literature and their limitations. We will also present a new paradigm, discuss its advantages while contrasting it with the paradigms already used in the field.

7.2 PARADIGMS AND THEIR LIMITATIONS

According to Keenan and colleagues (1990), inferences are traditionally measured by using memory or activation tasks. In memory measures, participants are asked to recall a past event or to state whether or not a target word featured in some earlier text. Difficulties, error, or delays in giving an answer are said to demonstrate that inference occurred. These authors describe memory measures as presenting a "serious problem" (1990, p. 382) because the nature of the

task encourages participants to compare their mental representation of the text to a target and this can lead to the generation of inference at the moment of testing as opposed to inference occurring at encoding.

7.2.1 *Memory-based Measures*

The cued-recall paradigm was the first measure used in STI literature and because of that it has been one of the most recurrently used (Winter & Uleman, 1984; Winter, Uleman, & Cunniff, 1985; Claeys, 1990; Uleman, Moskowitz, Roman, & Rhee, 1993; Uleman, Winborne, Winter, & Shechter, 1986; Uleman & Moskowitz, 1994; Bassili & Smith, 1986). Winter and Uleman (*e.g.*, 1984) were the first ones to show that a dispositional cue implied by (but not presented with) a trait-implying sentence was a more effective recall cue for that sentence than semantic associates of words presented in the sentence. For example, the recall of the sentence "The librarian carries the old woman's groceries across the street" was better cued by the trait disposition word "helpful" than by the semantic cue "books" or by no cue at all. The greater efficacy of the trait cue was presented as evidence that participants had spontaneously inferred the trait upon reading the sentence. Yet, some researchers questioned the interpretation that traits are inferred during the reading of the sentence, by arguing that the trait word given at test can prompt the recall of the behavior, through backward associations from the trait to the behavior. Presenting the trait during the memory test can trigger the generation of typical behavioral exemplars that would cue the recall of the correct behavior (*e.g.*, Corbett & Dosher, 1978; Srull & Wyer, 1989). This happens because the subject is asked to explicitly retrieve the sentence. Another limitation of this paradigm is the lack of control for word-based priming (Keenan et al., 1990), that is, the trait can become activated via the presence of very specific words that are part of the sentence without any consideration for the meaning of the sentence as a whole.

The recognition probe paradigm is another memory measure (Ham & Vonk, 2003; Newman, 1991; Uleman, Hon, et al., 1996; Van Overwalle, Drenth, & Marsman, 1999; Wigboldus, Dijksterhuis, & Van Knippenberg, 2003; Wigboldus, Sherman, Franzese, & van Knippenberg, 2004) that was imported (Uleman, Hon, et al., 1996) from discourse comprehension literature (*e.g.*, McKoon & Ratcliff, 1986). In this paradigm participants are presented with trait-implying sen-

tences under memory instructions and immediately after the reading of the sentence a probe words is presented. In the critical trials the probe word is the trait implied in the sentence and the instruction is to indicate if the word was or was not part of the sentence. The logic of the task is that if the trait is inferred from the sentence it will be more activated and thus, it will be more difficult (longer reaction times - RTs) to say that it was not part of the sentence and/or the responses will be less accurate. This paradigm, when used with the appropriate control sentences solves the problem of word-based priming. Uleman and colleagues (1996) used control sentences that were rearrangements of the words from the trait-implying sentences but did not imply the trait any longer (note that the used of these controls is not the norm but an exception). This ensures that the two sentences (control and trait-implying) are equal in terms of activation coming from specific words, since the words used in each are closely matched. If some difference is found between the two versions (say longer RTs to reject the trait in the trait-implying condition), it has to be due to the processing of the meaning, or, in other words, due to the inference making present in the trait-implying sentence and absent in the rearranged version. Also, contrary to cued-recall, in this task the participants won't infer the trait as a strategy to improve memory, because the inference leads to a worse, rather than better, performance. The main limitation of this paradigm is that we don't know whether the inference was drown during the encoding or during the test, due to the contamination problem. The task requires the subject to check the probe against the sentence. While checking the probe against the sentence, the participant can be led to draw the inference. The inference being drawn at the test moment can lead to slower responses, just like it would be expected if the trait had been inferred while reading the sentence.

Carlston and collaborators (Carlston & Skowronski, 1994; Carlston et al., 1995; Carlston & Skowronski, 2005) developed a different memory based paradigm - the saving in relearning. In this paradigm participants are instructed to familiarize themselves with trait-implying paragraphs describing behaviors paired with photos representing the actors of those behaviors. Later, after a distraction task, they have to memorize photos of people presented with personality traits. Some of the traits were implied in the paragraphs presented with those actors in the first phase, whereas other are new. The argument is as follows: if the traits are inferred about the actors during the encoding of the paragraphs, then pairing the same photo with the implied trait would reinstate the initial inference which will lead to faster learning of these pairs than learning of new pairs

(a result called saving in relearning). In the test phase the photo is presented again, this time, as a cue to recall the trait. Recall is expected to be better in the relearning trials than in the newly paired trials. In this paradigm it is also difficult to know whether the trait is inferred during the reading of the behavior. In the relearning phase, where the actor and the trait are presented for memorization, the subject might strategically retrieve the behavior to improve their memorization of the trait and thus, draw the inference during this retrieval.

Finally, false recognition is a memory based paradigm developed by Todorov and Uleman (2002, 2003, 2004). In this paradigm participants are first exposed to pairs of photos of people and sentences describing their behaviors, under memorization instruction. After being exposed to these material, they are presented again with the same photos together with traits. The task here is to indicate if the trait was part of the sentence presented with that photo in the first phase of the study. Higher false recognition rate is expected if the trait was inferred from the sentence in the initial phase. This happens because, once the trait is inferred, it is part of the representation of the event and, as such, it is difficult for the participant to distinguish what was actually presented from what was inferred. Again, since the participant is required to contrast the trait against the sentence memorized in the first phase, reconstructive processes are also possible in this paradigm. When the participant tries to remember the behavior, the inference can be drawn. Nether the saving in relearning, nor the false recognition paradigm uses rearranged controls, which makes it hard to evaluate the influence of word-based priming.

7.2.2 *Activation Measures*

Lexical decision is an activation measure that can give us access to the activated trait without explicitly asking the participant to retrieve the behavioral description. Zárate, Uleman, and Voils (2001), in order to avoid the reconstructive process criticism associated with memory measures, used the lexical decision to access the inferences during encoding. The authors presented two types of trait targets for the lexical decision task to their subjects. One type that was preceded by behavioral descriptions that implied those target traits, and a second type where the target was preceded by a control sentence that was unrelated to the trait. Two limitations can be pointed out to this paradigm. The first one is that the type of control does not control for word-based

priming. Second, there is some evidence that the relationship between the sentence and the trait affects the lexical decision latencies, suggesting that the two are contrasted one against other (Balota & Chumbley, 1984; West & Stanovich, 1982; Neely, Keefe, & Ross, 1989). It seems that the backward associations between the target and a prime affect lexical decision latencies. The context checking seems to happen after the lexical access in order to facilitate the decision process and if the target is somehow related with the prime/sentence the decision to say "yes is a word" will be encouraged.

The word stem completion is another activation measure used to study STI. Whitney and Williams-Whitney (1990) used constrained word stems to detect online (*i.e.*, drawn at encoding) inferences. Participants in this paradigm start by reading trait-implying or control paragraphs and then complete a word stem (*e.g.*, F R _ _ _ _ _ _ for FRIENDLY) with the first word that comes to their mind that fits into the blanks. If the trait is activated by the trait-implying sentence, then the trait will be more often used to complete the stem than in the control condition (*e.g.*, Whitney & Williams-Whitney, 1990; Bassili, Smith, & MacLeod, 1989). This measure is usually applied with an effective control condition for word-based priming, *i.e.*, with rearranged versions of the trait-implying paragraphs that contain the same words but do not imply the critical traits (Whitney & Williams-Whitney, 1990). Moreover, this measure was shown to be unaffected by backward associations and thus, less likely to be driven by retrieval processes. Because of that it can be considered a better measure than lexical decision task (Whitney, Waring, & Zingmark, 1992). There is, however, one limitation we would like to point out to this paradigm, that also applies to all the previously mentioned paradigms. The answer the participant is required to give is a direct measure of the activation of that particular trait. These traits are usually pre-tested so that they are inferred by most of the subjects from the inferential sentences. It is not very obvious, however, that all the participants infer the pre-tested traits. For example, in Whitney and colleagues' pilot study (1992) participants were asked to generate traits for behavioral descriptions. The traits used as targets in the stem completion task were given by the subjects in the pilot study in average 54% of the time (with the least consensual being generated 35% and the most consensual 90%). This clearly shows that the participants do not all infer the same traits. However, the same target trait (or stem) is presented for all the participants at test. When a specific trait is used in the test phase (making the effect depend on the lexical activation of that

specific word) and if, for example, a synonym is activated at encoding instead, the lexical facilitation is not expected to occur anymore (or inhibition might occur due to competition between lexical formats). However, at the conceptual level, the same representation might be activated. In more extreme case, the activated representation might not even overlap with the meaning of the target expected to be given as a stem completion.

Activation measures like lexical decision and word stem completion are data-driven tasks as opposed to conceptually-driven tasks like recall is. The distinction between data-driven and conceptually-driven tasks regards the distinction between tasks in which subjects rely more on physical features of the stimulus, and those where the focus on the physical features is minimal and where conceptually-based and top-down processes are used (Roediger & Blaxton, 1987). In this perspective, it is assumed that any task can rely on both types of processes. This means any tasks can be situated on the continuum between data-driven processes on one pole and conceptually-driven processes on the other pole. This view puts a lot of focus on the processing operating in each task and it is claimed that there are "memory benefits to the extent that the operations required at the test recapitulate or overlap the encoding operations during prior learning" (Roediger, Weldon, & Challis, 1989, p. 16). By applying this logic to trait inferences, it is clear that the process taking place at encoding is conceptually-driven since the meaning of the whole sentence has to be comprehended in order for the trait to be inferred. Moreover, no perceptual features can be processed because the trait is only implied and not actually presented at encoding. So, a conceptual task would be more sensitive to such a process because it only minimally depends of the perceptual features of the target.

Some memory measures like free recall can be considered conceptual tests, because the person directly accesses the representation in memory. However, memory measures are contaminated due to explicit retrieval. Activation measures solve, in part, the contamination problem, but they rely heavily on data-driven processes, which, as argued before is not a good fit to a top-down mechanism like STI. Word stem completion and lexical decision are far closer to the data-driven pole of the continuum than to the conceptual pole (*e.g.*, Hamann & Squire, 1996; Roediger, Weldon, Stadler, & Riegler, 1992) and thus, are not suited to STI.

A conceptually-driven measure that is not based on memory would solve both, the contamination problem and the conceptual sensitivity problem.

7.2.3 *Modified Free Association – a new conceptually-driven activation measure*

We suggest a new measure to study STI. Hourihan and MacLeod's (2007) modified word association paradigm is a conceptually-driven measure of implicit memory that appears to overcome some of the above limitations. In their study, in the learning phase participants either generated words from meaningful cues (*e.g.*, "the piece of furniture used for sitting –c?") or merely read the words (*e.g.*, "chair"). In a test phase participants performed a word association task where they had to speak aloud the first word to come to mind upon sight of a prompt word that has been either generated or read in the learning phase. Generating a word at encoding primes that word in implicit memory and this priming or activation extends to neighboring words in the semantic network (Nelson & Goodmon, 2002). Thus, when the prompt was a word generated by the participant, as opposed to a word that was just read, the generated word gets more easily available and is faster to access and so is its semantic network (Light, Prull, & Kennison, 2000). Translated to trait inference research, the premise of the paradigm is that reading a trait-implying sentence primes the trait word and its semantic neighbors. Thus, when an inferred trait word is encountered in the free association task, this delivers faster production of an associate compared to an un-primed trait. Reading a control sentence primes no trait word and therefore responses are slower in this condition since the network of the trait was not pre-activated.

Hourihan and MacLeod presented the paradigm as a conceptual implicit memory test, that provides an isolated measure of conceptual priming free from the influence of explicit retrieval. They claim that if there is nothing to retrieve from the study episode, the explicit retrieval would not be a problem. The authors argue that if a concept was semantically processed (as we think the traits are when inferred), its semantic characteristics should be activated above a baseline, and some activation should spread to semantically associated neighbors (*e.g.*, activation of the trait "friendly" might lead to the activation of the trait "nice"). The activation of the semantic network of a trait is detected via the latencies to generate an associate. Thus, the more the semantic network of the trait was activated by the reading of the behavior the faster and easier will be for the subject to generate an associate. The dependent variable in this paradigm is not the content of the response but the speed of the response. Moreover, there is research that has demonstrated that RT is a valuable measure of conceptual implicit memory (*e.g.*, Light et al., 2000). This is the

first paradigm that allows us to measure the speed to access the semantic network of an inference. The idea of spreading activation through the network is, furthermore, in agreement with the halo effect usually found in the STI literature (Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007; Crawford, Skowronski, Stiff, & Scherer, 2007; Skowronski et al., 1998).

This task reflects the lexical access to the trait, the access to an associate in the semantic network of the trait, and the articulation of the associate. Like in lexical decision, the inferred trait is provided to the participants, but contrary to what happens in lexical decision the accuracy is no longer important here, because there is no correct or wrong answer. The lack of correct answer makes it unnecessary to worry with the memorized material and as a consequence with explicit retrieval. Even though the instruction that the subjects are given in the study phase are to memorize the material (in order to avoid activating impression formation goals), this is an activation measure as opposed to memory measure, which again makes it less sensitive to explicit retrieval strategies since the participant is not required to contrast the trait with past information.

We cannot nevertheless, totally exclude the possibility that backward associations from the trait to the sentence is not affecting the access to the semantic network. However, as we shown later, the inference is also detected in a delayed free association test (experiment 2 and 3). In a delayed test the backward association is less likely to occur because the subject does not have immediate access to the sentence anymore.

Next, we describe two more paradigms: the naming task and the modified Stroop task. Even though these tasks are not usually used in the STI literature we decided to approach them because they are seen by some researchers as good measures of inference and as such we decided to compare them with the new modified free association task in our second experiment.

Naming

In the naming task, after reading inference or control sentences, participants are presented with words and instructed to read the words out loud as quickly as possible. The naming task can be decomposed in lexical access and articulation and because these steps happen so quickly it is difficult to imagine how the relatedness between the target and the text, *i.e.*, the backward association, can have an influence. But, for the same reason it is difficult to imagine that it can be sensitive to text-based inference (Norris, 1986). Thus, whereas there are some studies showing that, un-

like the lexical decision, naming is not affected by backward association (Seidenberg, Waters, Sanders, & Langer, 1984), there are authors that question its ability to detect more elaborated inference. Researchers that have a modular view of language believe that the lexical access can only be influenced by intra-lexical associations and not text-level factors (Fodor, 1983). Thus, even if the inference in drawn during encoding there will be no effect on lexical access because "the lexicon is informationally encapsulated from top-down effects" (Keenan et al., 1990, p. 392). Potts and colleagues (1988) found inference with naming, however the authors admit that this measure might be less sensitive than others, which is in agreement with the small effects they reported (11 and 17 ms). A second limitation of naming task is the possibility of being accomplished without any access to the lexicon, via grapheme to phonemes translation rules (Keenan et al., 1990). Also, note that just like in stem completion and lexical decision, this effect depends on the activation of a very specific lexical format (the word presented at test), and so, the inference of other similar traits (*e.g.*, synonyms) might be difficult to be captured with this measure.

Modified Stroop task

The modified Stroop task is another activation measure popular in text comprehension research. It involves naming the color of the ink in which words are written. In this task participants usually read a text and then name the ink color of test words (Conrad, 1974; Dosher & Corbett, 1982; Whitney & Kellas, 1984; Whitney, 1986). The logic of the task is as follows: if the trait is inferred from the reading of the behavioral sentence, and presented for ink color naming, the participant will take longer to name the color of the inferred trait than he would take to name the ink color of an unprimed trait. That happens because, if the word is previously activated, it is more difficult to suppress its articulation in order to say the ink color.

The task is affected by lexical access, by color identification and color articulation. It is very implausible that the color identification and the articulation are affected by the relatedness between the target and the text, meaning this measure might be a purer measure of the lexical access. Keenan and colleagues argue that it might be a better measure than naming, because the latencies for Stroop are longer, avoiding eventual floor effects (Keenan et al., 1990). These same authors claim that this measure allows to determine if people process the inference at conceptual or lexical level. If the word is primed at conceptual level (the target "doctor" when it

was preceded by a text talking about “hospitals” as opposed to a text talking about “subway”) then it takes longer to name the color ink. However, when the word is primed at lexical level by explicit presentation of the word “doctor”, then facilitation is expected for ink naming (Dosher & Corbett, 1982; Whitney, 1986). Stroop task, just like naming is a data-driven task. Also, the answers in this task are very dependent on specific traits becoming activated. When the word given at the test is different from the one inferred, the articulation of the word at test will not interfere so much with color ink naming because that specific lexical format was not activated before.

In the modified free association paradigm (MFAP), if the trait provided at test is the one inferred, then facilitation should be observed at both lexical level (*e.g.*, “helpful” inferred and “help” generated as an associate) and conceptual level because the trait is part of a larger activated network (“helpful” inferred and “concerned” or “aid” generated). If the trait inferred is different from the one provided at test, any obtained effect cannot be related with that specific lexical form, but there is still an effect that can be obtained due to conceptual priming, that is due to the fact that the concept is part of a larger activated network.

In three experiments we tested MFAP and its sensitive to STI. In experiment 1, we predicted that the reaction time to generate an associate for a trait prompt would be faster when following trait-implying than control sentences. This is due to the fact that by priming a trait word, its semantic network is also being primed and the accessibility of related words in implicit memory increases. In experiment 2 we compare the MFAP with the Stroop task and with naming task. In experiment 3, we are showing that the MFAP is also useful to study the link between the trait and the actor. Moreover, we test the sensitivity of the free association task to differences between STI and spontaneous trait transference (STT), a cognitive error that results in the transference of the inferred trait to a person that is not the actor (Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007).

If sensitive to trait inference, the free association task will be a fruitful contribution to trait inference research as it overcomes some of the limitations of the existing paradigms. First, in order to avoid explicit impression formation goals, the material is presented under memorization instruction for a later unspecified test. Second, the memorization phase and the free association phase are presented as separate tasks, making it more difficult for the subject to be aware of the

relation between the two. Third, because there is no need to explicitly retrieve past information, contamination is less likely to occur. Fourth, it is possible to control for word-based priming by including rearranged versions of the trait-implying material. Fifth, it is a conceptually-driven task, meaning the operations happening at test recapitulate at some extent the top-down operations taking place at encoding. And, finally, it is easy to apply MFAP in a delayed fashion, so that the subjects read all the behavioral descriptions first and are tested in a separate phase. Furthermore, in MFAP participants remain unaware of what is being tested as the task has no explicit success criteria, and indeed the word content produced is irrelevant.

7.3 EXPERIMENT 1

The aim of experiment 1 was to validate MFAP by demonstrating that it is sensitive to the occurrence of STIs. Participants in each trial, were presented with a sentence with a behavioral description followed by a free association prompt word. Upon sight of the prompt they were asked to speak aloud the first word that came to their mind. In the critical trials, the sentence was either trait-implying or control (*e.g.*, trait-implying sentence: "He phoned for help while the others just screamed." and control: "He screamed for the others to help find the phone.") and the prompt was the trait word implied in the trait-implying sentence (the trait "calm"). Following Hourihan and MacLeod's (2007) logic, we predicted that the trait inference would prime the semantic network of the trait word and consequently lead to faster generation of an associate when compared to an un-primed trait from the control condition.

7.3.1 *Method*

Participants

39 undergraduate students participated for academic credits. 6 were males and the average age of the sample was 19.46 years old. In this experiment we used as baseline the sample size reported by Hourihan and MacLeod (2007) in their first study (N = 30). The data in the three

experiments reported in this paper, were analyzed only after the sample size was complete and no more data were collected afterwards.

Material

24 pairs of sentences from Uleman, Hon, Roman and Moskowitz's (1996) study were used in the current experiment. For each trait-implying sentence, there was a control sentence that had approximately the same words but rearranged in such a way that the trait was not implied anymore. To validate the material for the population in the United Kingdom, we conducted a pilot study where the same material was adapted (American expressions changed to British) and given to participants for a rating task. 72 English native speakers living in the UK took part in the pilot study. Each of them was presented with 12 trait-implying sentences and 12 control sentences. The two versions of the same trait were never presented to the same participant. In each trial the participant was provided with a sentence and a trait, and was instructed to indicate to what extent they believed the trait belonged to the person performing the action in the sentence by using a slider (the slider could go from 0 (not at all) to 100 (totally)). Next, we calculated the average rating for each trait-implying and control sentence. The average difference between the rating for the trait-implying and the control was 44.51, with the smallest difference being 22.46 (for the trait "ambitious", for which the trait-implying sentence was "She held a full-time job while being a full-time student." and the control was "She couldn't hold a full-time job while being a full-time student.") and the larger difference being 70.67 (for the trait "sociable" with the trait-implying sentence being "He liked parties more than films." and the control being "He liked films more than parties."). All the 24 tested pairs of sentences and the correspondent traits were used in this experiment. We also had other 24 pairs of trait-implying and control sentences, that were not pre-tested; these will be used as filler trials, as explained in the Procedure section. Additionally, we had 36 neutral sentences that were also used as fillers, and are not meant to imply any specific traits ("She rented a summer place where she went the previous year."). For the free association task, we had 24 critical traits (corresponding to the pre-tested pairs of sentences; *e.g.*, "calm"), 12 traits not related with any of the sentences (*e.g.*, "reliable"), and 48 non-trait words (*e.g.*, "drill").

7.3.2 *Procedure*

The experimental design of this study consists of 2 conditions manipulated within-subject, one concerning the trait-implying sentences and a second concerning the rearranged sentences. The reaction time is the dependent variable.

Participants were tested individually in a sound proof lab on a desktop PC with 17-inch CRT monitor. Instructions and stimuli were produced in size 14 pt white font on a black screen. The software used to build the experiment was OpenSesame (Mathôt, Schreij, & Theeuwes, 2012), a multi-platform open source tool. A Sony plug-in head-set and microphone sounded the 'bleep' warning for the prompt in the free association task and recorded verbal reaction times digitally. The experimenter noted response content by hand. In the pilot study presented below, Qualtrics software was used to conduct the study and Prolific platform to publicize it.

Participants were told that the experiment was an investigation about multitasking and how people cope with the multiple requirements of the environment. More precisely they were asked to memorize sentences while doing a concurrent task - performing associations with target words (that are said to be unrelated to the sentences). Participants were instructed to memorize the sentences in order to encourage the attendance of the stimuli and avoid overt impression formation goals. After each sentence, a word was presented for the free association task and the instruction was to say the first word that comes to their mind upon sight of the prompt as fast as possible. In the free association task they were told that there was no wrong or right response. Before the study began, the experimenter helped the participants to fit the headset and microphone and instructed them to read aloud a series of words on the screen to optimize microphone recording threshold. Participants were instructed to speak clearly and to avoid clearing their throats or make any other noises that would invalidate their response. They started with four practice trials that were followed by an opportunity to ask questions.

Each trial began with a 500 ms fixation cross that was followed by the presentation of a sentence. The sentence remained on the screen for 4000 ms. Participants were instructed to read the sentence silently and memorize it. The sentence was replaced by a 500 ms blank screen that was followed by a bleep sound that warned participants that the prompt was about to appear. Between the beep and the prompt, a 40 ms blank screen was presented. The prompt remained on screen

until a vocal response was detected. At this point the word disappeared and a new trial began. In total there were 84 trials. The presentation order of the sentence-prompt pairs was randomized for each participant. Because we did not want the two versions of the same trait to be presented to the same participant, we created two versions of the experiment. Version A where 12 traits were preceded by their trait-implying sentences and the other 12 by controls and version B where the 12 traits that in version A were preceded by trait-implying sentences were preceded, in this version, by controls and the 12 that were preceded by controls in version A, in version B were preceded by trait-implying sentences.

Thus, in each version of the experiment there were 12 critical trait-implying sentences followed by the correspondent implied trait, 12 critical control sentences followed by the correspondent trait (implied in the trait-implying version), 12 non-critical trait-implying sentences followed by non-trait words, 12 non-critical control sentences followed by non-trait words. The last 24 trials with non-critical sentences are fillers and are presented in order to avoid having only trait-implying sentences being followed by personality traits, which would have made the goal of the study easier to detect by the participants. In other 24 of the trials, neutral sentences were followed by non-trait words. Finally, in the last 12 fillers, neutral sentences were followed by traits, so that not all the neutral sentences would be followed by non-trait words. All these fillers have the purpose of hiding the main relationship under investigation (trait-implying sentences and the correspondent implied trait) that is present in only 12 of the 84 trials.

During the free association task, the experimenter coded as invalid those trials that were affected by technical issues or those where participant's 'lip popping' noises were recorded in error as their vocal response. After all trials were completed, to follow the cover story regarding the memory test for the sentences, participants were given a memory test consisting of 20 pairs of similarly worded sentences. Their task was to identify which of the two versions they had seen earlier. The whole experiment lasted approximately 20 minutes.

7.3.3 *Results and Discussion*

The dependent variable in this paradigm is not the response given by the subjects but the RT of the response. To clean the data, we excluded the trials where recording problems were detected

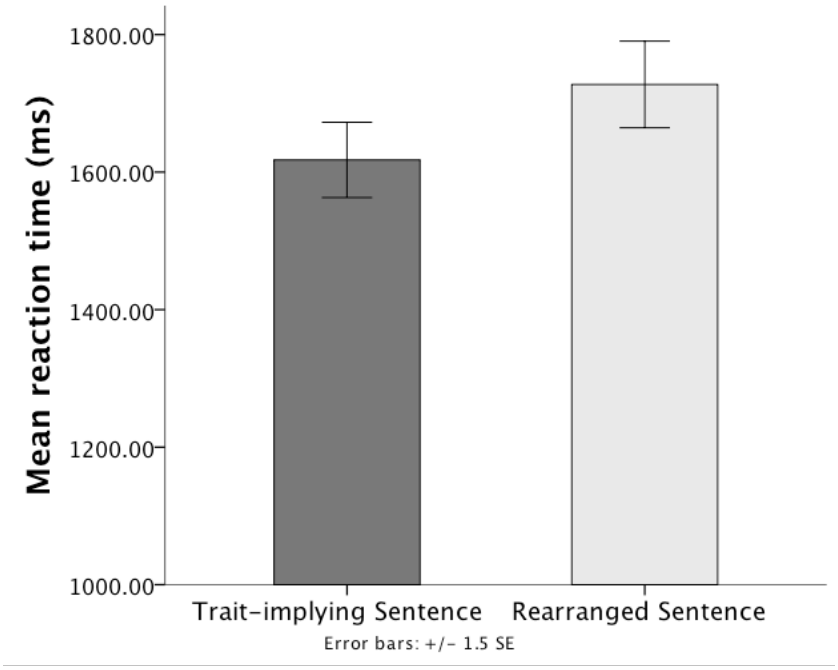

Figure 24: Mean reaction times in function of the condition in experiment 1.

and those flagged by the experimenter (undersensitive or oversensitive microphone), accounting for 109 responses in a total of 936 (11.65%). These 936 responses are concerning only the critical trials, that is, the trials where pre-tested trait-implying and control sentences preceded the trait prompt (12 trials in each condition).

Next, we applied an outlier elimination method based on median absolute deviation (MAD; Leys, Ley, Klein, Bernard, & Licata, 2013; Rousseeuw & Croux, 1993). This method, as opposed to the standard deviation criterion, is not affected by outliers. By using this method we eliminated 63 responses (15.11%) from the trait-implying condition and 50 responses (12.20%) from the rearranged condition, leaving a final total of 354 and 360 RTs, respectively, for the following analysis.

A repeated measure ANOVA, where the only factor was the type of sentence that preceded the trait (trait implying or control/rearranged sentences) was conducted. All the p-values reported are two-tailed, except when replication results are reported (experiment 2).

We found a significant effect of type of sentence, $F(36, 1) = 8.93$, $p = .005$, $\eta_p^2 = .20$. As shown in figure 24, the reaction time to generate a word was significantly shorter when the preceding sentence was implying the trait ($M = 1617.77$, $SD = 222.28$) than when it was a

rearranged version ($M = 1727.39$, $SD = 255.44$). From this result we can conclude that reading the trait-implying sentence, as opposed to the rearranged, leads to the activation of the trait and its semantic network, which facilitates the generation of an associate in the free association task.

Thus, as predicted this measure seems to be sensitive to trait inferences since we found a significant difference in the latencies to generate associates between the trait-implying and the control conditions.

7.4 EXPERIMENT 2

In the second experiment we compared the efficacy of MFAP to other two activation measures used to detect inferences: the naming and the modified Stroop task.

The free association task, in experiment 1, was applied immediately after the reading of the sentence, but testing inference in a delayed mode can give us access to inferences that need more time to develop. Some researchers think that stabilized inferences that become part of the representation of the even occur only on a later stage of information processing (Kintsch, 1988). Thus, it makes sense to test MFAP in a delayed fashion.

In this experiment, we also added a priming manipulation where the trait itself was presented as prime before being presented as target. The goal is to prime perceptual features of the word and thus facilitate the access to it. This is relevant, especially if the inference is only partially encoded as claimed by the minimalist hypothesis theorists (McKoon & Ratcliff, 1986, 1989, 1990). In the minimalist hypothesis, it is claimed that inferences that are not necessary for coherence are minimally or partially encoded. Thus, by providing a subliminal prime, the inference can be instantiated more firmly. Moreover, Strood task is expected to lead to opposite results with and without prime, since repetition prime activates the lexical format of the inference and thus, should lead to facilitation instead of interference in ink color naming.

7.4.1 *Method*

Participants

77 students took part in this experiment in exchange for a 10 euros voucher. 27 were males and the sample's average age was 22.67 years old. In this experiment we used a larger sample size than in the previous experiment because naming is a less sensitive task.

Material

This experiment was conducted with Portuguese participants, so the material was pre-tested for this population. A pre-test was conducted with the purpose of obtaining trait-implying sentences illustrative of a big set of personality traits (223). 293 subjects participated in this pre-test. Each participant was presented with 15 personality traits randomly chosen from this set and their task was to generate a typical behavior for the traits. Participants were instructed to think about people they knew and concrete behaviors. They were also told to avoid using adjectives, and to be as specific as possible in their behavioral descriptions. Two independent judges received half of the presented traits and the correspondent behavioral descriptions generated by the subjects in the pre-test. After eliminating answers that were not behavioral descriptions and redundancies, the judges selected 2 to 3 behavioral descriptions that better illustrated each trait. Traits with similar behavioral descriptions, usually synonym traits, were grouped under the same trait label. This resulted in 154 pairs of trait-behavioral descriptions. Two new judges received these pairs and were instructed to select the best behavioral description, that is, the sentence that implied the trait the most. In a second phase, the same two judges created rearranged versions from the previous 154 trait-implying sentences. An effort was made to use all the words from the trait-implying sentences in their rearranged versions. The newly created rearranged sentences were then presented to four new judges. The first two (group A) were asked to indicate the first word that came to their mind when reading each sentence and the other two (group B) were instructed to evaluate to what extent each sentence was related to the correspondent trait by using a 9-point scale ranging from not related (1) to very related (9). If at least one of the judges from group A generated the target trait for the control sentence, that pair of sentences (rearranged/implying)

was excluded from the final set. Moreover, if both judges from group B rated the relation between the rearranged sentence and the trait with a value higher than 5, that pair of sentences was also excluded from the set. This filtering left us with 122 pairs of trait implying/rearranged sentences. However, from these 122 pairs, only 36 were used in this experiment. The criteria to choose the 36 was based on a lexical decision experiment conducted for a different purpose, but that allowed us to select those pairs where the lexical decision for the trait was faster when preceded by the trait-implying sentence than when preceded by the rearranged version. Beside the 36 critical pairs of trait implying and rearranged sentences, we also used 36 neutral sentences. And besides the 36 traits that corresponded to the critical sentences, we also utilized a set of 36 non-trait words.

7.4.2 *Procedure*

The experiment had a 2 Type of sentence (trait-implying sentence versus rearranged) $\times$ 3 Type of test (naming versus free association versus modified Stroop) $\times$ 2 Prime (with prime versus without prime) design. All the factors, except the factor concerning the presence of the prime, were within-subject and the reaction time to give an associate is the dependent variable.

Participants were told that they were going to take part in 4 unrelated studies. The first one was sentence-memory task for all the participants, and the following three were counterbalanced in terms of the order they were performed in. The three tasks were the naming task (reading the words provided), the modified free association task (say the first word that came to their mind upon the sight of the provided word) and the modified Stroop task (say the color ink in which the word was written). Note that contrary to experiment 1, in experiment 2 the measurements are delayed, meaning the measurement is not immediately after reading each sentence, but during a later phase after memorizing all the sentences. The participant was also randomly assigned to either the prime or no-prime condition. The only difference between the two conditions is that in the prime condition the target words later presented for naming, free association, and modified Stroop, were presented as subliminal primes in the sentence-memorization task before the presentation of the sentences. The purpose of the prime is to strengthen the inferential process in the critical trials. The sentence-memorization task started with 6 practice trials that were followed

by 72 experimental sentences. The selection of the 18 critical trait-implying sentences (and the correspondent traits) from the set of 36 was random for each participant, leaving the remaining 18 critical pairs to be the controls (rearranged sentences and the correspondent traits). This prevented that the same participant sees the two versions (the trait-implying *and* the rearranged sentences) of the same trait.

Thus, each participant memorized 18 trait-implying sentences, 18 rearranged sentences, and 36 neutral sentences that did not imply traits. Some of the neutral sentences were not even behavioral descriptions (*e.g.*, "In the winter the days are shorter"). Including non-behavioral descriptions made it highly unlikely for the subject to become aware of our goal to study STI.

In the no-prime condition each trial started with a 800 ms fixation dot, followed by the presentation of the sentence for 3500 ms, and ended with a 500 ms blank screen. In the prime condition, before the presentation of the sentence, and after the fixation dot, a 150 ms mask preceded the prime which was presented for 16 ms. The prime was followed by another mask that was presented for another 16 ms. Next, the sentence was presented for 3500 ms, followed by the 500 ms blank screen. 37 participants did the prime condition and the remaining 40 did the non-prime condition. After the memorization task, the naming, the free association, and the Stroop task followed. The procedure in the three tasks were identical in the prime and in the no-prime condition (this manipulation regarded only the sentence-memorization phase). All the three tasks started with 6 practice trials that were also used to adjust the microphone to the participant's voice. In the naming task participants were told that the goal of the task was to study the reading skills and they were instructed to read as fast as possible the words presented to them. Before the word, a fixation dot was presented for 800 ms and a 40 ms blank screen followed. The word was presented next and stayed on the screen until the microphone detected a response. The experimenter took notes about the accuracy of the responses. After the response was detected a 500 ms blank screen followed. The procedure was identical in the free association and Stroop task except the instructions that were different and the word in the Stroop task was written in blue, yellow, green, or red ink. In the free association task, the participants were told that the goal of the task was to investigate the speed with which people generate new words. And thus, when the words appeared on the screen they were instructed to say as fast as possible the first word that came to their mind after reading the prompt. They were also told that there was no right or wrong

response and to say the word that came to their mind regardless of how ridiculous it appeared. In the modified Stroop task, the instruction was to say, as fast and accurately as possible, the color ink in which the word was written. And the cover story for the Stroop was that we were interested in investigating the color naming. All the participants did the three tasks, and from the 18 traits that were implied in the 18 trait-implying sentences memorized in the first part of the study, 6 were presented in the naming task, 6 in the free association, and 6 in the modified Stroop task. Besides those, other 6 traits were presented in each task and those corresponded to the rearranged sentences. Moreover, besides the 12 traits (6 in the rearranged and 6 in the trait-implying condition) presented in each task, there were also 24 non-trait words unrelated with any of the sentences presented before. Only the latencies for the traits were analyzed in the following analyses. After participants finished the three tasks, a memory test regarding the sentences was performed, just to keep with the initial cover story. This test consisted of 16 pairs of similarly worded sentences and the task was to identify which of the two versions they had seen earlier. The presentation order of the trials in each of the 5 phases was randomized for each participant.

7.4.3 *Results and Discussion*

After eliminating the recording problems (5.30% responses from the overall set of responses) we applied the same outlier removal method used in experiment 1. In the priming condition, in the association task 37 responses were eliminated with this method only in the trait-implying condition (18.59%) and other 31 responses in the rearranged condition (15%); in the naming task it eliminated 29 responses in the trait-implying condition (12.83%) and 21 responses in the rearranged condition (9.59%): and, finally, in the Stroop task it eliminated 13 responses in the trait-implying condition (6.10%) and 13 responses in the rearranged condition (6.02%). In the no-prime condition, in the association task 25 responses in the trait-implying condition (11.68%) were eliminated and 31 responses in the rearranged condition (14.35%); in the naming task it eliminated 14 responses in the trait-implying condition (6%) and 31 responses in the rearranged condition (13.60%); and in the Stroop task it eliminated 15 responses in the trait-implying condition (6.41%) and 16 responses in the rearranged condition (6.81%).

Next, we conducted individual analysis for each task.

Association Task

Starting with the association task, we conducted a mixed ANOVA, where the between-subject factor was the prime (present or absent) and the within-subject factor was the type of sentence (trait-implying or rearranged). The dependent variable was again the mean RT to generate an associate. As expected and replicating the result in the previous experiment with a delayed measure, a significant effect of type of sentence was observed, $F(1,74) = 4.66$, $p = .017$ (one-tailed), $\eta_p^2 = .06$, with shorter RTs when the sentence is trait-implying ($M = 1758.06$, $SD = 433.46$) than when it is rearranged ($M = 1871.25$, $SD = 425.12$). The priming effect, however, did not reach significance, $F(1,74) = 2.80$, $p = .099$, $\eta_p^2 = .04$, and nor did the interaction, $F < 1$.

Stroop Task

A similar ANOVA was conducted for the Stroop task, with the same independent and dependent variables. The main difference between the free association task and the Stroop task is the fact that if a trait is activated by the sentence, it would interfere with color naming task instead of facilitating. Thus, larger RTs are expected in the trait-implying condition than in the rearranged condition in this task. However, in the prime condition instead of interference, facilitation is expected, because the repetition prime occurs at lexical level as opposed to semantic. A main effect of priming was found, $F(1,75) = 4.60$, $p = .013$, $\eta_p^2 = .06$, and also an interaction between the type of sentence and the priming, $F(1,75) = 6.46$, $p = .013$, $\eta_p^2 = .08$. This interaction is detected due to the interference that occurs for the trait-implying sentences ($M = 693.59$, $SD = 79.60$) when compared with the rearranged sentences ($M = 674.23$, $SD = 84.25$), $F(1,39) = 4.03$, $p = .052$, $\eta_p^2 = .10$, when there is no prime presented, and lack of difference between the trait-implying ($M = 644.86$, $SD = 58.82$) and rearranged sentences ($M = 657.92$, $SD = 60.30$) when there is a prime, $F(1,36) = 2.54$, $p = .12$, $\eta_p^2 = .07$. Thus, contrary to expectations, no facilitation was detected, only interference to indicate the color ink of the trait.

Naming Task

A similar analysis was performed for the naming task, with the same independent and dependent variables. The only effect that reached significance in the present task was the interaction between priming and type of sentence, $F(1,75) = 5.56$, $p = .021$, $\eta_p^2 = .07$. When the participants were primed with the to-be-inferred trait the difference between the trait-implying and the rearranged sentences was not significant, $F(1,36) = 1.99$, $p = .167$, $\eta_p^2 = .05$. This might be happening due to the floor effect mentioned before when the limitations of the naming task were discussed. When there was no prime, contrary to the expected, shorter reaction times were detected for the rearranged sentences ($M = 571.81$, $SD = 61.85$) than for trait-implying sentences ($M = 584.76$, $SD = 72.43$), $F(1,39) = 3.69$, $p = .062$, $\eta_p^2 = .09$, even though this effect was marginal.

The results in our experiment 2 suggest that, as expected, the naming task is the least sensitive task to detect STI, while Stroop task and specially the new paradigm we are introducing, MFAP, performed in the expected way, that is, served to distinguish the trials where inference is expected from those where they are not.

7.5 EXPERIMENT 3

In experiment 3 we explored how the information inferred is integrated with other material presented in the inferential context. Our hypothesis is that the information is differently integrated depending on the relevance of the person performing the behavior.

This final study also allows us to test the paradigm's sensitivity to spontaneous trait transference (STT), the tendency to transfer the inferred trait to a person who just describes the behavior of a third party or that is just simultaneously presented with the behavioral information (*e.g.*, Carlston et al., 1995; Skowronski et al., 1998). In fact, in the case of the superstitious banana, Brown and Bassili (2002) showed that trait transference can even extend to inanimate objects. Evidence shows that trait information is differentially processes in STI and STT. Differences include a stronger effect in STI than STT (*e.g.*, Brown & Bassili, 2002; Skowronski et al., 1998; Goren & Todorov, 2009), the reduction or elimination of STT effect if the actor and informant are

presented concurrently (Crawford, Skowronski, & Stiff, 2007; Goren & Todorov, 2009; Todorov & Uleman, 2004) and a generalization of trait impressions in the form of halo effects in STI but not STT (Crawford, Skowronski, & Stiff, 2007; Carlston & Skowronski, 2005; Skowronski et al., 1998). These differences have led to an ongoing debate about the nature of the cognitive processes that underly them, with some proposing a two-process approach (attribution-based in STI and associative in STI; *e.g.*, Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007) and others a contrasting single-process model based on different strengths of association (Brown & Bassili, 2002; Bassili & Smith, 1986; Orghian, Garcia-Marques, Uleman, & Heinke, 2015). A paradigm that is sensitive to both STI and STT can contribute to this debate. By using the free association task, we predict that the stronger STI effect compared to STT (Brown & Bassili, 2002; Goren & Todorov, 2009; Skowronski et al., 1998) would lead to larger activation of the semantic network of the inferred trait and larger integration with the representation of the person and thus to faster reaction times to deliver an associate word.

7.5.1 *Method*

Participants

51 undergraduate students participated for academic credits. 2 were males and the mean age of the sample is 18.86 years old. Here we also decided to increase the sample size because we are looking for an interaction, and it also corresponds to the show-ups in a week time.

Material

The same 24 trait-implying sentences tested in the pilot study from experiment 1 are used in this experiment. Beside the 24, 48 neutral sentences were given for memorization. For the free association task, the 24 critical traits, 24 non-trait words, and 24 non-trait words that were actually part of the 24 the neutral sentences were used. Besides the words and sentences, 72 black and white photos of people with neutral expressions were used.

Figure 25: Example of one learning trial and two free association trials showing different pairing conditions. Image a) depicts an irrelevant learning trial (identified by border color and pronoun gender); images b) and c) illustrate a match pairing or mismatch (control) pairing respectively.

7.5.2 *Procedure*

The experiment had a 2 Relevance (actor versus irrelevant person) $\times$ 2 Pairing (match trials versus mismatch trials) design. All the factors were within-subject and the reaction time to give an associate is the dependent variable.

Participants were told that the study's goal was to investigate how well people memorize information and multitask. The study consisted of two phases, a learning phase and a free association test phase. In the learning phase people were instructed to memorize photo-sentence pairs of material. In this phase there are two conditions: a relevant (STI) condition where the person presented in the photo is the actor of the action described in the sentence and the irrelevant (STT) condition where the person in the photo is said to be randomly pairing with the behavioral sentence and thus is irrelevant to the behavior. Henceforth, this is referred to as the relevance condition. Photo relevance was made explicit in the learning phase by presenting the photos in one of two colored borders. Instructions explained that when the sentence described the ac-

tions of the person in the photo, it would be presented in a blue border and when the text had been randomly paired with the photo, and was therefore irrelevant to the information, it would be shown in a red border (see figure 25). In the 4 practice trials, a reminder text regarding the meaning of the colors was shown underneath each photo. In the experimental trials, in the irrelevant condition the person in the photo had the opposite gender of the pronoun gender used in the sentence, to emphasize that there was no relation between the person and behavior. Participants were instructed to pay equal attention to all information presented, to photos and to behavioral descriptions regardless of the relevance. The learning phase was presented in a 72 (36 relevant/ 36 irrelevant) trials block. In each trial, a fixation dot lasting for 500 ms appeared followed by the simultaneous presentation of the sentence and the photo in the center of the screen for 6000 ms. A "spot the difference" distractor task lasting three minutes followed the learning phase. Participants read the instructions for the test phase and had an opportunity to ask questions before and after four practice trials. The cover story in this task was: "(. . .) we are interested in people's multitasking abilities. Performing more than one task simultaneously might involve costs in our performance. So, in this phase you will have to perform two different tasks simultaneously. First, you will be presented with the photos from the previous phase again, and thus, you have a second chance to better memorize them. Second, you will have to perform a Free Association task. Free Association is a task where you are presented with a word and your mission is to say, as quickly as possible, the first word that comes to your mind.". The free association trials began with a 500 ms fixation cross followed by a bleep warning. Next, one of the photos presented earlier, this time without an identifying border, appeared in the center of the screen for 1000 ms. The prompt word then appeared directly underneath the photo and remained until a vocal response was detected. Once a response was recorded, a blank screen was shown for 500 ms, followed by the final screen with the sequence 'press C to continue'.

There were different types of trials in the test phase. Half the photos were photos from the irrelevant condition in the learning phase and the other half from the relevant condition. Also the word would be a trait in some trials, and these are the critical trials. The trait was always implied in a trait-implying sentence from the learning phases. When the word wasn't a trait it would be a word that was not related with any sentence from the learning phase, or a word that was part of neutral sentence presented in the test phase. These last trials were included

to prevent the participant from understanding that the goal was about trait inference from the fact that every time the target word was related with the previously seen sentence the word was always a personality trait. Thus, some of the words are non-traits and, not only are related with the sentences, but are actually part of the seen sentences. Note, however, that only the critical trials matter for the analyses reported below, and these are trials where the words were traits.

Among the critical trials there were two types of pairings, the match and the mismatch/control trials. In match trials the photo was presented with the trait implied in the sentence presented with that photo in the learning phase, whereby in mismatch trials the photo was presented with a trait that was inferred in a sentence presented in the learning phase with a different photo. Thus, each trial was one of four relevance-pairing conditions; relevant (STI)/ match, relevant/ mismatch, irrelevant (STT)/ match, irrelevant /mismatch. An example of an irrelevant condition with both pairing versions is shown in figure 25. The difference between the match and the mismatch condition inform us about the integration of the trait inferred into the representation of the actor or the irrelevant person. In the case of the irrelevant person the integration is expected to be much less. Thus, if the trait is inferred from the sentence and bounded to the person, then, in the test phase, when the trait is presented with that person, a facilitation should be expected. In other words, the photo will re-instantiate the inferred trait and its network and thus, making it easier to generate an associate when the trait is actually presented. For each participant there were 24 critical trials, 12 were STI (from which 6 were match and 6 mismatch) and 12 were STT (6 match and 6 mismatch) trials. Image, prompt assignment, and presentation sequence were randomized for each participant. Note that in this case we don't have rearranged sentences, the control is the mismatch condition. We also have shown, in previous research, that not using rearranged controls is not a serious problem when the measure is delayed (Orghian et al., 2016).

7.5.3 *Results and Discussion*

Spoiled trials (recording problems) and RTs below 150 ms were eliminated, making a total of 11.85% of the whole set of responses. Next, the MAD method was applied, eliminating 47 responses in the match relevant condition (16.73%), 34 responses in the mismatch relevant

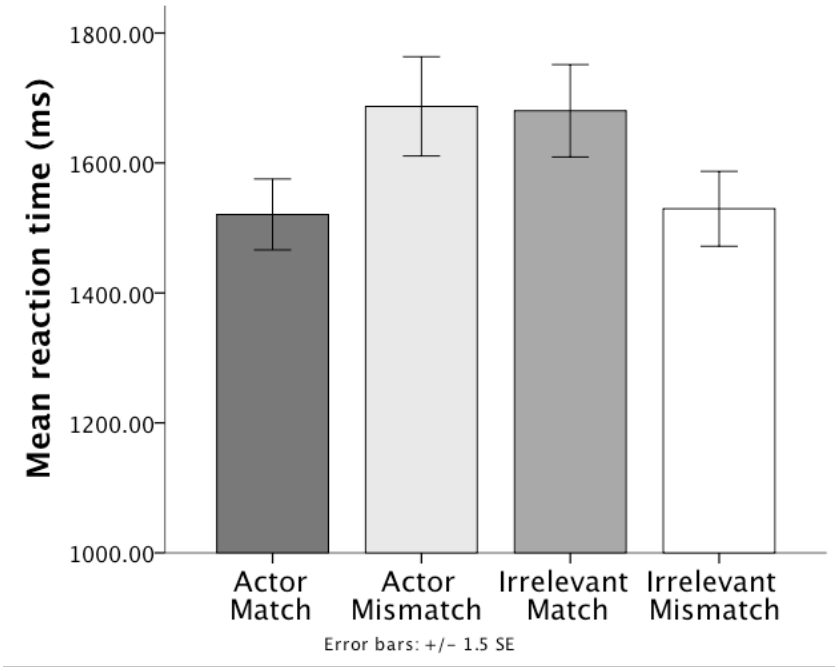

Figure 26: Mean reaction time in function of the relevance and pairing in experiment 3.

condition (12.55%), 28 responses in the match irrelevant condition (10.49%) and 38 responses in the mismatch irrelevant condition (14.18%).

A repeated measure ANOVA was conducted for the critical trials. The critical trials were the trials where traits were presented as targets for the free association task. The first factor is the relevance of the photo (an actor face – STI condition or an irrelevant face – STT condition) and the second factor is the pairing between the face and the trait (match or mismatch). As presented in figure 26, a significant interaction was found between the relevance and the pairing, $F(1,50) = 16.47$, $p < .001$, $\eta_p^2 = .25$, with a larger facilitation for the actor in the match condition ($M = 1520.76$, $SD = 259.39$) than in the mismatch condition ($M = 1687.13$, $SD = 364$), $F(1,50) = 10.33$, $p = .002$, $\eta_p^2 = .17$. The opposite patter is detected in the irrelevant condition, with a larger facilitation in the mismatch condition ($M = 1529.58$, $SD = 274.40$) than in the match condition ($M = 1680.36$, $SD = 338.70$), $F(1,50) = 7.93$, $p = .007$, $\eta_p^2 = .14$. These results suggest that this new paradigm can be used to detect differences between STI and STT and that it is sensitive to differences in relevance.

In this experiment we show that the MFAP can be used to study the link between the trait and the actor. This paradigm also allow to distinguish between relevant actor and irrelevant person via

a dissociation between STI and STT. The pattern found for STT was unexpected, since instead of a smaller effect, we find an opposite effect. In other words, we found more facilitation in the mismatch condition than in the match condition. This opposite pattern might be interpreted as a reflection of an inhibitory effect triggered by the presence of an irrelevant actor. When the person in the photo is not the actor, the subject might unconsciously (or not) inhibit the trait inferred and its linkage to the trait. Thus, in the test phase when the photo is presented, due to the inhibition, the trait inferred from the sentence might be suppressed and so its semantic network. When the trait presented is mismatched, a process similar to the release from inhibition (MacLeod, 1989; Geiselman & Bagheri, 1985) might be taking place.

7.6 GENERAL DISCUSSION

Our studies were motivated by a desire to overcome the limitations of the main paradigms used to study STI. An important aspect on STI paradigms is that all of them need to use cover stories in order to make the subject process the trait-implying material without activating the explicit goal of impression formation. One common way of doing it is by instructing the participants to memorize the material for a later unspecified memory test. Later on, in a memory test, the participants are required to contrast a target (*e.g.*, a trait) against the memorized material (*e.g.*, the behavioral sentence). For such a contrast to be possible the participants have to retrieve the memorized material. However, retrieving the sentence at the test might trigger trait inference mechanisms at that moment in time. That would mean that the inference is not a result of processing the sentence and spontaneously inferring the trait, but a result of retrieval processes happening later on, this is called the contamination problem. Paradigms based on such procedure that requires retrieval of memorized material are called memory based measures. Cued-recall (Winter & Uleman, 1984), probe recognition (Uleman, Hon, et al., 1996), savings in relearning (Carlston & Skowronski, 1994) and false recognition (Goren & Todorov, 2009) are memory based measures that can be easily contaminated by explicit retrieval processes. A way to overcome the contamination problem is by accessing the inference without requiring the participants to explicitly retrieve the learned information (Keenan et al., 1990). Activation measures are usually presented as the solution for the contamination problem, but these measures present their own drawbacks. Sur-

prisingly, activation measures are uncommon in STI research and we suspect that it is because they might be less sensitive to such top-down phenomena as STIs. In other words, activation measures rely on more superficial processing of the material and thus, might be less sensitive to STI. Moreover, some activation measures, like the naming and the lexical decision rely on very fast responses and thus floor effects might be more frequent than in memory measures.

Finally, we introduced a new paradigm, the Modified Free Association Paradigm, a conceptually-driven measure. The measure follows Hourihan and Macleod's (2007) logic that claims that if a word is conceptually processed, its semantic network will also get activated due to the spread of activation. The priming of the semantic network will lead to faster reaction times when the subject has to give an associate of the trait inferred. This task seems capable of detecting activations present in the semantic area that sustains the inference and not just the inferred concept itself.

In experiment 1, after reading trait-implying and rearranged sentences participants were presented with words and instructed to say the first words that came to their mind. The reaction times to generate a word upon the sight of the trait was shorter when it followed a sentence that implied that trait than when it did not imply the trait (while containing roughly the same words). We interpreted this difference in the RTs as being due to the fact that inferences are spontaneously inferred while reading the trait-implying sentence.

In experiment 2, we used a delayed version of the MFAP together with a naming task and a modified Stroop task. We did not find STI with the naming task. As defended by some researchers it might be due to the lexicon, that is encapsulated and thus is not or less sensitive to conceptual activation, (Fodor, 1983) or due to floor effect (Norris, 1986; Keenan et al., 1990). We did find interference of the inferred trait in the performance in the modified Stroop task and we found a facilitation in speed to generate an associate for the trait inferred with the MFAP, replicating the result in experiment 1. Crucially, this experiment demonstrates that, both the modified Stroop task and the MFAP allow for trait inference detection even when applied in a delayed fashion.

The aim of experiment 3 was to validate the paradigm for sensitivity to the relevance of the actor and thus, to STT, a neighboring effect to STI. Due to the stronger effect usually observed in STI (Brown & Bassili, 2002; Goren & Todorov, 2009; Skowronski et al., 1998), we predicted faster reaction time when the person presented together with the behavioral description is the actor of the actions described in the sentence. An opposite effect was found for trials where an

irrelevant person was presented together with the behavioral information. Past evidence shows differential effects in STI and STT across a range of situations (Crawford, Skowronski, & Stiff, 2007; Goren & Todorov, 2009; Todorov & Uleman, 2004). However, none of the evidence was sufficiant to conclude that, in terms of underlying processes, STI and STT are actually different (Orghian et al., 2015). Contrary to this evidence, where the difference is usually in the magnitude of the effect, we found a dissociation between the two phenomena, that might be a more clear indication of dual-processing. And, instead of assuming that some special and complex process takes place in STI (attributional thinking; Skowronski et al., 1998), we can also reason that the process in STI is purely associative and something else is taking place in STT, such inhibition. We can also speculate that the inhibitory effect seen in Study 3 is an attempt to suppress the erroneous STT associations.

Past research shows that the introduction of a new paradigm can uncover new findings, indeed the introduction of the savings in relearning paradigm, which was intended to provide a measure of the actor-trait link, inadvertently led to the identification of spontaneous trait transference (Carlston et al., 1995, Experiment 4). In summary, the experiments presented in the present Chapter suggest that the free association paradigm may be a viable tool in future trait inference research. The task presents characteristics that makes it appealing when to compared to more traditional paradigms. To start with, the subject process the material under memory instruction without encouraging explicit trait inference, a crucial condition to grasp a spontaneous phenomenon. The task is an activation measure, as opposed to memory measure, meaning it is implausible that the inference is a result of contrasting the trait against the memorized sentence. Filler were also used in order for the goal of the task to not become obvious to the participants. A good measure should also be implicit, and even though the memory measures can be seen as implicit since their main objective is not actually to test memory but test inference, they do require an explicit consideration of the past event that contaminates the measure. MFAP does not require such consideration of past event, making it very difficult for the person to intentionally try to remember memorized material. This measure can also be used with the proper control condition (that is, rearranged controls, as done in experiment 1 and 2) that assures us that we are dealing with "real" inference and not just activation coming from word-level priming. An important advantage of this task is the fact that we can measure the access to the semantic net

of a trait and not only to the activation of the trait itself, which is of particular relevance in the context of the minimalist hypothesis (*e.g.*, McKoon & Ratcliff, 1986). Within this hypothesis it is claimed that the activation of an inference does not happen in a none-or-all fashion and that it can be encoded partially. Moreover, contrary to lexical decision (that on average takes between 600 and 700 ms) and naming (that takes between 500 and 600 ms) that occur very fast and are thus, less sensitive to top-down thinking, the MFAP occurs slowly (because generating words is more difficult, it take usually more than 1500 ms to generate a word) leavening more room for top-down influences to be detected. Thus, MFAP is a conceptually-driven task that makes a better fit to a conceptual mechanism as STI.

Although no paradigm is entirely free from limitations (Butler & Berry, 2001), the free association paradigm has the potential to provide a cleaner measure of trait inference as it is uncontaminated by the explicit recall and retrieval processes that affect traditional memory measures and it is a conceptually-driven measure. Our findings suggest that the free association task is sensitive to trait inference but its validity requires further testing, the next step being a verification of its insensitivity to backward associations especially when it is utilized as an immediate measure. Moreover, contrary to the first two experiments, in the third experiment, one can argue that in the test phase when the photo is presented as an opportunity to better memorize it, it is conceivable that the participant might try to recall the behavior as a strategy to improve the memory of the photo. This recollection of the behavior can trigger the inference at test. A subliminal priming of the photo would solve this problem by decreasing the probability of explicit sentence recall, a modification that should be considered in future research.

Also, awareness of the study purpose should be controlled for by using a questionnaire or structured interview. Even if filler as used, and the relation between the memorization phase and the free association phase is not obvious, we don't know for sure that the participants are not aware of it.

7.7 REFERENCES

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? the role of word frequency in the neglected decision stage. *Journal of Experimental*

Psychology: Human Perception and Performance, *10*(3), 340–357.

Bassili, J. N., & Smith, M. C. (1986). On the spontaneity of trait attribution: Converging evidence for the role of cognitive strategy. *Journal of Personality and Social Psychology*, *50*(2), 239–245.

Bassili, J. N., Smith, M. C., & MacLeod, C. M. (1989). Auditory and visual word-stem completion: Separating data-driven and conceptually driven processes. *The Quarterly Journal of Experimental Psychology*, *41*(3), 439–453.

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.

Butler, L. T., & Berry, D. C. (2001). Implicit memory: Intention and awareness revisited. *Trends in Cognitive Sciences*, *5*(5), 192–197.

Carlston, D. E., & Skowronski, J. J. (1994). Savings in the relearning of trait information as evidence for spontaneous inference generation. *Journal of Personality and Social Psychology*, *66*(5), 840–856.

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.

Carlston, D. E., Skowronski, J. J., & Sparks, C. (1995). Savings in relearning: Ii. on the formation of behavior-based trait associations and inferences. *Journal of Personality and Social Psychology*, *69*(3), 420–436.

Claeys, W. (1990). On the spontaneity of behaviour categorization and its implications for personality measurement. *European Journal of Personality*, *4*(3), 173–186.

Conrad, C. (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory and Cognition*, *2*(1), 130–138.

Corbett, A. T., & Dosher, B. A. (1978). Instrument inferences in sentence encoding. *Journal of Verbal Learning and Verbal Behavior*, *17*(4), 479–491.

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*(3), 466–472.

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.

Dosher, B. A., & Corbett, A. T. (1982). Instrument inferences and verb schemata. *Memory and Cognition*, *10*(6), 531–539.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Geiselman, R. E., & Bagheri, B. (1985). Repetition effects in directed forgetting: Evidence for retrieval inhibition. *Memory and Cognition*, *13*(1), 57–62.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, *27*(2), 222–248.

Ham, J., & Vonk, R. (2003). Smart and easy: Co-occurring activation of spontaneous trait inferences and spontaneous situational inferences. *Journal of Experimental Social Psychology*, *39*(5), 434–447.

Hamann, S. B., & Squire, L. R. (1996). Level-of-processing effects in word-completion priming: A neuropsychological study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 933–947.

Hourihan, K. L., & MacLeod, C. M. (2007). Capturing conceptual implicit memory: The time it takes to produce an association. *Memory and Cognition*, *35*(6), 1187–1196.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*(5), 513–541.

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

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*(2), 163–182.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766.

Light, L. L., Prull, M. W., & Kennison, R. F. (2000). Divided attention, aging, and priming in exemplar generation and category verification. *Memory and Cognition*, *28*(5), 856–872.

Lupfer, M. B., Clark, L. F., & Hutcherson, H. W. (1990). Impact of context on spontaneous trait

and situational attributions. *Journal of Personality and Social Psychology*, *58*(2), 1239–1249.

MacLeod, C. M. (1989). Directed forgetting affects both direct and indirect tests of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 13–21.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314–324.

McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, memory, and cognition*, *12*(1), 82–91.

McKoon, G., & Ratcliff, R. (1989). Semantic associations and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 326–338.

McKoon, G., & Ratcliff, R. (1990). Dimensions of inference. *Psychology of Learning and Motivation*, *25*, 313–328.

Neely, J. H., Keefe, D. E., & Ross, K. L. (1989). Semantic priming in the lexical decision task: roles of prospective prime-generated expectancies and retrospective semantic matching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(6), 1003–1019.

Nelson, D. I., & Goodmon, L. B. (2002). Experiencing a word can prime its accessibility and its associative connections to related words. *Memory and Cognition*, *30*(3), 380–398.

Newman, L. S. (1991). Why are traits inferred spontaneously? a developmental approach. *Social cognition*, *9*(3), 221–253.

Norris, D. (1986). Word recognition: Context effects without priming. *Cognition*, *22*(2), 93–136.

Orghian, D., Garcia-Marques, L., Uleman, J. S., & Heinke, D. (2015). A connectionist model of spontaneous trait inference and spontaneous trait transference: Do they have the same underlying processes? *Social Cognition*, *33*(1), 20-66.

Orghian, D., Ramos, T., & Garcia-Marques, L. (2016). Activation is not always inference: word-based priming and spontaneous trait inferences. *Manuscript submitted for publication*.

Potts, G. R., Keenan, J. M., & Golding, J. M. (1988). Assessing the occurrence of elaborative inferences: Lexical decision versus naming. *Journal of Memory and Language*, *27*(4), 399–415.

Roediger, H. L., & Blaxton, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory and Cognition*, *15*(5), 379–388.

Roediger, H. L., Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of endel tulving* (pp. 3–41). Hillsdale, NJ: Erlbaum.

Roediger, H. L., Weldon, M. S., Stadler, M. L., & Riegler, G. L. (1992). Direct comparison of two implicit memory tests: Word fragment and word stem completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1251–1269.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical association*, *88*(424), 1273–1283.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *13*, 501–518.

Seidenberg, M. S., Waters, G. S., Sanders, M., & Langer, P. (1984). Pre-and postlexical loci of contextual effects on word recognition. *Memory and Cognition*, *12*(4), 315–328.

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.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review*, *96*(1), 58–83.

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.

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.

Todorov, A., & Uleman, J. S. (2004). The person reference process in spontaneous trait inferences. *Journal of Personality and Social Psychology*, *87*(4), 482–493.

Uleman, J. S. (1999). Spontaneous versus intentional inferences in impression formation. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 141–160).

New York, NY: Guilford.

Uleman, J. S., Hon, A., Roman, R. J., & Moskowitz, G. B. (1996). On-line evidence for spontaneous trait inferences at encoding. *Personality and Social Psychology Bulletin*, *22*(4), 377–394.

Uleman, J. S., & Moskowitz, G. B. (1994). Unintended effects of goals on unintended inferences. *Journal of Personality and Social Psychology*, *66*(3), 490–501.

Uleman, J. S., Moskowitz, G. B., Roman, R. J., & Rhee, E. (1993). Tacit, manifest, and intentional reference: How spontaneous trait inferences refer to persons. *Social Cognition*, *11*(3), 321–351.

Uleman, J. S., Newman, L. S., & Moskowitz, G. B. (1996). People as flexible interpreters: Evidence and issues from spontaneous trait inference. *Advances in Experimental Social Psychology*, *28*, 211–279.

Uleman, J. S., Winborne, W. C., Winter, L., & Shechter, D. (1986). Personality differences in spontaneous personality inferences at encoding. *Journal of Personality and Social Psychology*, *51*(2), 396–403.

Van Overwalle, F., Drenth, T., & Marsman, G. (1999). Spontaneous trait inferences: Are they linked to the actor or to the action? *Personality and Social Psychology Bulletin*, *25*(4), 450–462.

West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *8*(5), 385–399.

Whitney, P. (1986). Processing category terms in context: Instantiations as inferences. *Memory and Cognition*, *14*(1), 39–48.

Whitney, P., & Kellas, G. (1984). Processing category terms in context: instantiation and the structure of semantic categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 95–103.

Whitney, P., Waring, D. A., & Zingmark, B. (1992). Task effects on the spontaneous activation of trait concepts. *Social Cognition*, *10*(4), 377–396.

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

Wigboldus, D. H., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484.

Wigboldus, D. H., Sherman, J. W., Franzese, H. L., & van Knippenberg, A. (2004). Capacity and comprehension: Spontaneous stereotyping under cognitive load. *Social Cognition*, *22*(3), 292–309.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology*, *47*(2), 237–252.

Winter, L., Uleman, J. S., & Cunniff, C. (1985). How automatic are social judgments? *Jouranl of Personality and Social Psychology*, *49*(4), 904-917.

Wyer, R. S., & Srull, T. K. (1986). Human cognition in its social context. *Psychological Review*, *93*(3), 322–359.

Zárate, M. A., Uleman, J. S., & Voils, C. I. (2001). Effects of culture and processing goals on the activation and binding of trait concepts. *Social Cognition*, *19*(3), 295-323.

8

FINAL REMARKS

> *But I'll add though that there is something at the bottom of every new human thought, every thought of genius, or even every earnest thought that springs up in any brain, which can never be communicated to others, even if one were to write volumes about it and were explaining one's idea for thirty-five years; there's something left which cannot be induced to emerge from your brain, and remains with you forever.*
>
> Dostoevsky, 1996, p. 368

One of the main challenges humans have is that our inner selves are hidden from our view. Thoughts, personality, emotional states, beliefs, goals, and intentions, are all important constructs to understand others. However, none of them are actually observable. People are forced to find ways to attain these constructs. Dostoevsky's quote is illustrative of this difficulty in communicating our own mental states to others. Fortunately, we have a powerful tool helping us to move a little closer to understand others' selves, it is our ability to infer. We infer a lot of rich information about others' inner selves from their actions. By observing others' behaviors, we infer intentions and personality traits that motivated their actions.

In the present dissertation, we were interested in the way people infer personality traits from others' behaviors and, in particular, in the spontaneous aspect of these inferences, *i.e.*, in spontaneous trait inferences (STI from now on). In the realm of STI two main aspects were explored. The first one concerned the underlying mechanism of STI, by contrasting STI with spontaneous trait transference (STT from now on), a cognitive error that results in the transference of the inferred trait to the wrong person. By exploring the theoretical aspects of STI we encountered a

series of methodological challenges. These challenges gave birth to the second part of this dissertation, which concerns the efficacy of current measures in detecting spontaneous inferences. Moreover, we suggest a new paradigm as a possible way to deal with some of the limitations presented by more traditional paradigms.

In this section we intend to review the main work and ideas advanced in the dissertation and how they contribute to the current state of the art. Moreover, we discuss possible follow-up actions and try to foresee some of the possible paths STI and STT research may take in the future.

Our general approach in this dissertation was to take advantage of a series of methodologies that different areas of psychology can offer (from computation to attention, discourse comprehension and implicit memory) in order to answer the questions we proposed to investigate. Borrowing tools widely used in cognitive psychology to study social psychology, can bring the two fields closer and facilitate conceptual integration. And, as it has been suggested and shown before, such integration can be beneficial to both parties (Devine, Hamilton, & Ostrom, 1994).

8.1 FINDINGS AND THEIR LIMITATIONS

After STT was discovered the urge to understand the process responsible for STI and how it differs from STT compelled the researchers to conduct a series of experiments in order to distinguish STI from STT. The abundant empirical differences found were considered by some as demonstrations that STI was a result of an attributional process whereas STT a result of an associative process (*e.g.*, Carlston & Skowronski, 2005). However, the use of attribution to explain STI is not unanimous (*e.g.*, Bassili & Smith, 1986). To explore this theoretical question we used connectionist modeling.

The use of computational models in general and connectionist models in particular is popular in many fields of psychology. However, the role of computational modeling in the research enterprise is not always clear. It cannot be seen as a substitute of experimentation, it is just another tool that aids us in deriving hypotheses from theories that hold a certain degree of complexity (Smith, 1996). In other words, by using simulations we can, more precisely and reliably, describe a theory and consequently derive empirical implications that can be further tested. Researchers have been using this tool to describe basic processes of vision and memory, but also higher

level phenomena as categorization and decision making (Kruschke, 1992; Weber, Goldstein, & Busemeyer, 1991). Moreover, computational modeling has been recurrently used to discard dichotomous theories or explanations. For example, in the memory literature it was suggested that, in some situations, people learn classification by representing prototypes, that is, by assigning new stimuli to the category whose prototype they most closely match. This led to the suggestion that there might exist two distinct types of memories, episodic (based on exemplars) and generic (based on prototypes). This suggestion was mostly based on a single piece of evidence showing that the ability to classify old exemplars declined over time whereas the ability to classify prototypes did not (prototypes that were not even seen during the original learning; *e.g.*, Donald, Joseph, Don, David, & Steven, 1973). However, Hintzman and Ludlam (1980) showed that, under certain assumptions, in a model where only exemplars, and no prototypes, were being stored was possible to obtain differential forgetting rates for prototypes and old exemplars. Thus, MINERVA model showed that this evidence wasn't conclusive in what concerned the used of abstraction during classification learning and that simply forgetting old exemplars completely accounted for the differential forgetting.

In the current work, instead of using simulations to derive hypothesis, we used it to test the plausibility of a dualist hypothesis, just like in MINERVA (Hintzman & Ludlam, 1980). The suggestion of this dualistic hypothesis is part of a theoretical debate where, based on simples dissociations found empirically, two distinct processes are postulated to explain the existence of STI and STT. This use of computational modeling, can, in these cases, be regarded as formal demonstrations that help researchers to discard evidence that is not diagnostic.

In particular, we showed that, by using a single auto-associative network, STI, STT, and the main empirical differences between the two can be replicated without assuming the existence of an additional (*i.e.*, attributional) process. The model we presented in Chapter 2 is based on simple associations between nodes that are established due to co-activation. The learning of the network is based on an error-reduction mechanism called *delta learning rule* (McClelland & Rumelhart, 1989), where the internal input (similar to expectations about the world) is iteratively compared with the external input (similar to facts in the world). This comparison results in updated links between the nodes in order for the model to better predict the external input events provided by the environment.

In the first simulation, a larger STI effect than STT was obtained, replicating the difference reported in the literature (Goren & Todorov, 2009, experiment 1). This difference in the magnitude of the two effects is usually interpreted as being a result of differences in the processes underlying the two phenomena, even though this is merely a quantitative difference. In computational terms, what motivated the difference between the two effects was the assumption that the actor is more salient in the context of their own behavior than is the non-actor (a communicator, for example). Being more or less salient means more or less attention is paid to it and, in the model's language, means the input for the person node receives more or less activation. This differential activation results in stronger or weaker connections that are created between the person node and the node representing the trait. Thus, more attention is paid to the person, stronger is the link created between that person and the trait inferred from the behavior. For the network to be able to "infer" traits from behavioral descriptions, the model is also presented with a phase where associations between nodes representing behaviors and nodes representing traits are being learned.

When the actor and the non-actor are presented simultaneously with the behavior, STT reduces, and this is interpreted as the attributional process disrupting the associative one (*e.g.*, Crawford, Skowronski, & Stiff, 2007). However, an alternative explanation is that the differential in the relevance is more obvious within this setting and as such the differential in attention is larger. When the relevant actor is presented next to the irrelevant person, the actor might seem even more relevant than in cases where he/she is presented alone. Thus, by implementing this differential in attention via larger activation of the actor node than of the non-actor node, in the second simulation, a significant reduction in the STT effect was obtained.

In the third simulation, we tried to model an experiment where the subjects, instead of being instructed to memorize the material or to familiarize with it, were instructed to detect whether the actor or the communicator was lying about the message they were conveying (Crawford, Skowronski, Stiff, & Scherer, 2007). It is argued that when a concurrent inferential task (like detecting if someone is lying) is performed while the behavioral material is being encoded, this concurrent task will disrupt the attributional process and, thus, the STI will decrease to STT's level. However, in this experimental setting the photos of the people are more important than the behaviors they are communicating and critically, they are equally important in both conditions because, regardless of their role (as a communicator of someone else's action or their own), the

task is to detect whether the person in the photo is lying. Computationally, in this simulation, we assumed that equal amount of attention was paid to both photos and, of course, this eliminated the differences between STI and STT because the way the model creates these differences is based mainly on differences in the activation of the nodes representing the persons.

Finally, in the fourth simulation, we computationally replicated a larger halo effect in STI than in STT (Skowronski, Carlston, Mae, & Crawford, 1998; Carlston & Skowronski, 2005). This was done by combining the attention based assumption and an extra step where the model learned that certain traits are associated with other valence congruent traits.

Moreover, to test the robustness of the model it was shown that the results can be generalized to a large range of parameters values. By applying more general principles of theory testing (Roberts & Pashler, 2000), we also discuss the predictive boundaries of our model. That is, besides showing which data the model can fit, we discuss and show what kind of data (by suggesting hypothetical experiments) could not be fitted by the model and thus, would falsify it.

The simulations can be seen as a formal demonstration that the empirical evidence that we have at the moment does not suffice to consider the existence of another process beyond simple association. It is not that we advocate for a single process view of STI, what we would like to plead is that stronger and different evidence (*e.g.*, double dissociations) is necessary for dualistic claims. We would also like to suggest that the use of simple computational frameworks like this one can be useful to test dualistic views in other areas and debates of social psychology.

One limitation of our argument is the fact that it is based in Occma's razor (also known as the parsimony principle). But, of course, there is no decisive reason to believe that the simplest model is the true one, and it can also be seen as diminishing the complexity associates with our cognition. Thus, ultimately, this Chapter is not providing us with any definitive answers regarding the process underlying STI. While this formalization suggests that a dualist view is weakly supported by data, it does not discard it. Moreover, it also does not say that the single process view is the true model. In sum, the modest contribution of this model is to demonstrate that the empirical evidence that is marshaled to back the claim for different processes underlying STI and STT is not compelling and still insufficient to provide support for this claim.

Importantly, in the model, we assumed attentional differences (motivated by the relevance of the actor) in order to replicate the empirical data. This attentional assumption is, however, debatable,

specially because some research shows that attention might not have a relevant role in what concerns the differences between STI and STT (Crawford, Skowronski, Stiff, & Leonards, 2008).

Social stimuli, just like any stimuli, need to be attended in order to be more deeply processed. And, like in any type of perception, perceptive signals are available for further selection and interpretation. Attending to these signals is a requirement for our brain to process and interpret it in terms of meaningful representations. The attention and the interpretation are not completely separate steps, specially when the attention is driven by top-down goals. We know, for example, that our brain uses different areas during goal-directed (prefrontal neurons) and stimulus-driven attention (parietal areas; Corbetta, Patel, & Shulman, 2008). When it comes to people perception, it seems specially relevant to consider the influence of top-down or goal-directed attention since variables like expectancy or social categories might come into play (Bodenhausen & Hugenberg, 2009). This distinction between more stimuli-driven or more goal-driven attention is important in STI and STT because the role of the person (actor or communicator) is a crucial piece of information that is expected to affect, in a top-down manner, the way the person is attended to. Moreover, we know that attention is composed of at least three separate processes: disengagement of attention from an initial target, directing of attention to a new target and reengagement of attention (Posner, Walker, Friedrich, & Rafal, 1984), sometime called the disengage-move-engage model.

Thus, in the third Chapter we tried to use paradigms and theoretical knowledge from more general attentional literature to investigate a different facet of the STI and STT debate. We used the principles from disengage-move-engage model and what is known about the inhibition of return phenomenon in order to employ the spatial cueing paradigm. In addition, we measured eye movements as a proxy for the attention allocation towards the elements presented in typical STI and STT trials.

In order to explore the role of attention in STI and STT, we employed a modified spatial cueing paradigm in experiment 1 and eye-tracking devices in experiments 2 and 3. In the first experiment, subjects were presented with behavioral descriptions that were followed by photos of actors or photos of people that were said to be randomly paired with the behaviors. The presentation of the photos was very short (200 or 400 ms) and so, the information about the role of the person (actor or non-actor) and the behavior was provided before the photo with enough time to be properly processed. A top-down orientation of attention is expected to occur depending on the

relevance/role (actor or non-actor) of the person in respect to the behavior. After the presentation of the face, an arrow appeared on the screen and the participant had to indicate as quickly as possible the direction of the arrow. This qualifies as a spatial cueing task because the face was cueing (validly if the location of the two were the same or invalidly if the two were on different locations) the arrow. A cueing effect was detected, that is, the response to the arrow was faster in the cases where the face appeared on the same location as the arrow (valid cue) than in the cases where the location did not coincide (invalid cue). When the face was presented for a longer time (400 ms) and it was invalidly cueing the arrow, the latency to indicate the direction of the arrow was longer if the face was of an actor than when it was irrelevant. We interpreted this result as the actor's face holding more the participant's attention than the irrelevant's person face. In other words, when the subject is encoding information about an actor, it takes longer to disengage from the actor face and attend the other side of the screen where the arrow is presented.

In the second and third experiments, the false recognition paradigm was applied while the participants' eye fixations were being monitored. The critical monitoring occurred during the learning phase where the behavioral sentences and the photos of the actor or the irrelevant person (that was said to be randomly paired with the behavior in experiment 2 and to be the communicator of someone else's behavior in experiment 3) were presented. In both experiments we found a larger number of fixations on the actor's face than on the irrelevant person's face. This result supports the hypothesis that more attention is paid to the actor than to someone that is not the actor of the behavior.

These studies present, however, several limitations, that make our conclusions less straightforward. First of all, in the first experiment there is no measure of STI and STT whatsoever. Although STI and STT are, according to the research reviewed in Chapter 1, supposed to occur in a unsolicited manner, routinely, and under a wide range of contexts and situations, the fact is that we cannot provide any evidence of these inferences in the present paradigm. That means that even though we find differences in the attention paid to the actor in comparison to the irrelevant person, we do not have evidence that the participants actually infer traits from the behavioral descriptions and whether the traits become associated with the people in the photos. Moreover, when we further measured STI and STT, by using a false recognition paradigm (the last two experiments in the Chapter) we found no differences between the two phenomena. This might

mean that the attentional differences are not necessarily translated into differences in the way the trait is being linked to the person. However, we did find a correlation between the number of fixations on the actor's face (and the number of transitions between the face of the actor and the behavior) and the STI effect. These same correlations were not significant for the STT. This might be an indication that attention has a role in STI, or, in other words, that the way people direct their attention affects the link between the actor and the trait. However, the way the attention is allocated does not relate to the STT effect.

This set of studies show that attention is differently engaged in STI and STT, and specially that more attention correlates with more binding between the actor and the trait. But, we still don't know what is the role of attention and what is the cause of this different engagement.

The lack of difference between the magnitude of the two effects in the false recognition paradigm, reported in Chapter 3, might be due to several reasons. It might be due to the lack of statistical power, due to the lack of sensitivity of the paradigm itself to the differences, or due to a combination of both. The paradigm is based on the comparison between match and mismatch conditions. In the mismatch condition the trait presented at the test was not implied in the sentence that was presented with that photo during the learning phase (it is a trait implied in a sentence seen with a different person). Critically, we know that STI is sensitive to halo effect, that is, a generalization to other non-implied traits that are consistent with the one implied (Carlston & Skowronski, 2005; Crawford, Skowronski, & Stiff, 2007). Thus, for the mismatch condition to be a good control, the mismatched trait paired with the actor in the test phase should not overlap in meaning or valence with the trait actually inferred from the sentence and associated with that person. If they are related, then we have a control that is contaminated by the inferred trait from the sentence. One way of overcoming this limitation is by assuring that the two meanings do not overlap, by using, for example, antonyms of the implied traits in the mismatch condition at test.

In experiment 1 from Chapter 4 we used antonyms in the mismatch condition. This time we found a significant difference between STI and STT effects, even though the effect was modest. Moreover, in experiment 2 we applied the forced choice recognition paradigm to test whether the trait is inferred from the behavior differently in STI and STT. In this experiment, participants started with the typical learning phase from the false recognition paradigm (that is, memorizing photos of people and behaviors), but in the second phase they performed two different tests. One

test consisted of presenting two very similar sentences that only varied in the presence of the to-be-inferred trait, and the participants' task was to choose the one they saw in the learning phase. Participants were equally good in distinguishing the presented from the non-presented sentences in both STI and STT, meaning the trait was inferred at similar extent in both conditions. In the second test participants were presented with new and old faces and their task was to say, for each of them, if they had seen the person in the learning phase. In this test participants revealed more familiarity with the actor faces than they did with the faces of non-actors. This result is in agreement with what we found in the experiments from Chapter 3, where more attention was paid to the actors' faces than to the irrelevant persons' faces. The results in both tests are also consistent with the assumptions that we made in the model presented in Chapter 2. The first assumption was that the actor's face is more activated that the face of a non-actor person, and the second assumption was that activating the trait from the behavior was an independent step that was not affected by the relevance of the person.

The main limitation in this Chapter is the fact that the forced choice recognition test was conducted in a different experiment instead of being integrated into the false recognition paradigm presented in the first experiment. Showing that the trait is equally inferred in STI and STT in the same experimental setting where we show differences in the magnitudes of the two effects, would have been a much stronger evidence. Moreover, having the memory test for the faces in the same experiment as the measurement of STI and STT effects would have allowed us to find if the two are correlated. In other words, it would have allowed us to investigate if a better memory for the face relates with a larger binding between the face and the trait.

Summing up, from these studies, we can conclude that first, the way the trait is inferred from the behavior is similar in STI and STT. Second, the difference resides in the way the visual representation of the actor is attended to and consequently processed. And finally, it is likely that the attention affects the binding between the actor and trait, but more research is needed to explore this relationship.

In the second part of this dissertation, two methodological aspects are discussed. These two aspects were inspired by the discourse and comprehension literature and by implicit memory research. Again, we are trying to use more general psychological principles and knowledge to solve problems encountered in this specific field.

The first aspect concerns the word-based priming, a limitation present in many paradigms used to study STI and neglected by most of the researchers. Word-based priming is an alternative way to activate a trait that is not via considering the meaning of the behavior, but through specific words in the sentence. The phenomenon of semantic priming has a long history in psychological science (Anderson, 1983; Collins & Loftus, 1975; Neely, 1977) and, in particular, the word-based priming has been a frequent concern in discourse and text comprehension literature (McNamara & Magliano, 2009). Disregarding word-based priming and its effects on more complex processes can lead to serious confounds. Each time a word is encountered, related concepts are automatically primed and researchers argue that this automatic activations should not be confused with knowledge-based inferences or elaborations (*e.g.*, Kintsch, 1993, 1998). Strategies were developed in order to distinguish the two and reach text-based inferences. A consequence of consulting the text comprehension literature to inform our investigation on STI is that some of the problems we encounter in social inference were already solved in text comprehension literature. In terms of scientific economy, using discoveries and conclusions from one field to inform solution in a neighboring field is a way of recycling knowledge that can save us time, money and effort. Thus, a suggested way of controlling for word-based activation (Keenan, Potts, Golding, & Jennings, 1990) is by using rearranged sentences as controls (sentences that have the same words as the trait-implying ones but with the words rearranged in such a way that the meaning changes and the traits are not implied in the sentences anymore). We discussed this problem and the correspondent proposed solution in detail in Chapter 5 and, in Chapter 6, we explore how it affects immediate and delayed measurements of STI. After creating a set of trait-implying sentences and their rearranged/control counterparts, a lexical decision task was conducted in order to distinguish between two types of sentences. The two types are: those pairs of trait-implying and rearranged sentences where the trait-implying counterpart leads to more activation of the trait than leads the rearranged (meaning that there is a real inference occurring alongside any word-based activation) - strong pairs, and those pairs where the activation of the trait is similar in both sentences (meaning that there is no extra activation coming from trait-implying sentences when compared with the control versions) - weak pairs. After making this distinction, these two groups of material were used in delayed, immediate and explicit measures. The results suggest that immediate measures, such as probe recognition paradigm, are sensitive to word-based prim-

ing. In other words, these two groups of material performed differently in the probe recognition task. In the weak group, the trait-implying and the rearranged sentences did not show differences in reaction times, meaning they were not capable of detecting STI. In the strong pairs, the responses were slower for the trait-implying condition, meaning that for this group of material participants were able to detect STI. However, in a delayed measure, like the false recognition paradigm, both groups of material allowed for STI detection, meaning that the weak group (the one that is more affected by the word-based priming) allowed for STI detection as well. Thus, the word-based activation is a major problem when trait inferences are measures via online tests, but less of a problem when the measurement is delayed.

In Chapter 5, we present control material in order to overcome the word-based activation problem. However, this material needs to be tested with much larger samples. Moreover, our results suggest that the rearranged sentences are more difficult to comprehend, meaning more research is needed to select those pairs that are similar in comprehension. Finally, as suggested in the Chapter, it is not enough to create and use rearranged sentences, it is also necessary to assure that they activate the trait significantly less than the implying sentences. Thus, this material should be tested with an activation measure in order to select those pairs with larger disparities in the activation of the trait. For example, in Chapter 6, we applied the lexical decision task to part of the sentences presented in Chapter 5, and that allowed us to discriminate between strong and weak material regarding the inference detection.

The purpose of the experiments presented in Chapter 6 was to study the time course of word-based and text-based activation. The main limitation of the Chapter relates with the fact that the trait-implying sentences are compared with rearranged sentences, but they were never compared with more traditional controls like neutral sentences. Comparing trait-implying sentences with rearranged and then trait-implying with neutral would allow us to better describe and quantify the confound present in the literature. A second important limitation is the way the time course of the word-based activation is being investigated. Paradigms that measure the inference at different moments in time are used in order to investigate the effect of word-based priming at different points in time. However, we cannot assure that differences in the effects are due to the moment the measurement is done, because, beside changing the timing of the measurement, we also change the paradigm. A possible solution would be to use the same paradigm and vary the time

of the measurement. While this would help us to understand the time course of the activation, this new version of the paradigm will not correspond to any existing paradigm.

Finally, in Chapter 7, we discuss two more methodological aspects regarding the paradigms applied in studying STI. The first regards the problem of contamination with explicit recall that affects memory measures and the second regards the distinction between data and conceptually-driven measures. We also present a new paradigm that offers critical advantages to STI investigations.

In this Chapter we not only present a new paradigm adapted from memory literature, we use the insights from this literature to analyze the restrains of the old paradigms. A common concern that researchers studying implicit memory have is the contamination problem. Contamination is said to occur when intentional and conscious recollection affects performance in implicit tasks. Some authors believe that it is almost impossible to escape this contamination (Butler & Berry, 2001). This insight was informative for a thorough analysis of the measures that are used in STI. Some authors (Keenan et al., 1990) working in text comprehension suggested that activation measures, that put much less emphasis on the recall of the information and more on detecting the activation of the inference, are less contaminated by explicit recall. However, an additional insight from the memory literature revealed us the importance of the type of processing that the measures rely on. Thus, a task can be data-driven or conceptually-driven (Roediger & Blaxton, 1987), that is, it can rely more on processing the perceptive features of the stimuli or rely more on the general meaning of the stimuli. This is important because one of these two types of tasks is more adequate to detect conceptually-driven mechanisms as the STI are.

Memory measures are contaminated by explicit retrieval processes, which means that the inference detected, instead of being a reflection of an on-line mechanism, might very well be a reflection of mechanisms taking place at retrieval. Activation measures, on the other hand, have a different problem related to the type of processes they rely on. Measures like lexical decision and word stem completion are mostly data-driven, meaning the performance rely on shallow processing and superficial features of the stimuli. We argue that conceptually-driven tasks, because relying on top-down processing, are more appropriate to measure STI which is a conceptually-driven mechanism. In particular, we borrowed the modified free association task developed by Hourihan and Macleod (2007) and introduced as an uncontaminated conceptual and implicit

memory task. In our first experiment of this Chapter, we apply this new paradigm in a immediate fashion, that is, participants were presented with the sentence (trait-implying or rearranged) and immediately afterwards were presented with the trait. They were instructed to say the first word that came to their minds when reading the prompts. Participants were faster in giving an associate to the trait when it followed a sentence that implied that trait than when it followed a sentence that had the same words rearranged to not imply the trait.

However, there is a limitation that all immediate tests present when applied to inference making. If the inference needs more time to develop, an immediate test won't be able to detect it. For that reason, in experiment 2, we measured the inference after all the sentences were learned. In this experiment, beside the modified free association task, we also asked participants to perform two more tasks, a naming task where they had to read the word (trait) as fast as possible and the modified Stroop task where the instruction was to name the color ink the word (trait) was written in. In the free association task we replicated the result (shorter reaction times in the trait-implying condition than in the control). With the same design (in the learning phase) and same material, in the naming task, we did not detect any trait inference, that is, there was no facilitation for the trait-implying condition. With the modified Stroop task, as expected, we found more interference in the trait-implying condition than in the control. The main advantage of the free association task, in comparison to the Stroop task, is that the effect with free association task depends less on the specific word/trait that was primed. If the participant infers a trait that is slightly different (a synonym, for instance) from the trait presented at test, this inference might not be detected via Stroop task but might still be detected by the free association task because it depends on the activation of the entire network of the trait and not only on the activation of that specific trait.

In the last experiment of this Chapter, we tested the modified free association task by applying it to the STI and STT debate. In particular, we wanted to test the paradigm's sensitivity to the difference between STI and STT. The results revealed a dissociation. For the STI trials, when the actor was presented with a match trait at test versus a mismatch (a trait implied in a sentence paired with a different face), a facilitation was detected in generating an associate, whereas the opposite patterns was detected for STT (shorted reaction times for mismatch). We did not expect a reversed effect for STT but a possible explanation for it is discussed. Specifically, we suggest

that in STT condition an inhibitory mechanist might be taking place in order to prevent the association of the trait with the irrelevant person.

This new paradigm requires further testing. In particular, the next important step is to investigate how MFAP is affected by backward associations. The backward priming is an interesting aspect of trait inferences because it relates to a theoretical dichotomy between inductive (from behavior to trait) and deductive (from trait to behavior) inferences. Maass, Colombo, Colombo and Sherman (2001) introduced the term induction-deduction asymmetry (IDA) after showing that people draw more behavior-to-trait inferences than trait-to-behavior inferences. Note that backward association is exactly the generation of a behavior from a given trait; it is a deductive inference. The rationale behind this result is the fact that a single trait finds expression in many different behaviors whereas a behavior is usually diagnostic of one or few traits. In other words, a trait is linked to many behaviors while a behavior is linked to fewer traits. And assuming that activation is inversely related to the number of existing relations/links, the activation of a concept/node results in a "fan effect", that is, each link receives only part of the entire activation. The fan effect is suggested to be larger for the deductive inference because the trait is linked to many behaviors and thus, the connection of the trait to each of those behaviors is weaker. This finding suggests that what we have been calling backward association might be weaker than the forward association in the context of trait inference, but it is still necessary to verify if results similar to the ones obtained in this Chapter can be explained purely based on backward associations. Moreover, in the third experiment, when the photo is presented for the second time, the participant might try to recall the behavior as a strategy to improve their memory of the photo. This recollection can trigger the inference at test. A subliminal priming of the photo would solve this problem by decreasing the likelihood of explicit sentence recall, a modification that should be considered in future research. And, finally, even if fillers are used and the relation between the memorization phase and the free association phase is not obvious, we don't know for sure that the participants are not aware of it. Thus, in order to verify whether the participants are aware of the purpose of the experiment, a questionnaire or structured interview should be conducted after the test phase.

To sum up the methodological facet of STIs, we would like to add that the word-based activation versus text-based activation debate and the immediate versus delayed measures debate might be more related with each other than it seems at first.

While it is conceivable to think that word-based activations are not real inferences, they might have an important role in the encoding of inferences. McKoon and Ratcliff claim that semantic associative information mediates the strength of an inference (McKoon & Ratcliff, 1989). They demonstrated empirically that strong associates present in the inferential context may contribute to the construction of the inference itself. Thus, the associative context of an inference might actually support the encoding of that inference. Relating this to our discussion regarding the role of word-based activation, eliminating strong associates from the sentences might have a drastic effect on the inference and not because associations were mistakenly taken as an inference but because its encoding depends on the activation of these semantic networks. It can be seem as a contextual facilitator. No inference happens in a vacuum, there is always a context that supports the inference, and that context can sustain the inference more or less. As McKoon and Ratcliff claim, inferences are sustained by more elementary features representing the meaning of the text. The time course of these features's availability, their specificity, and their strength define the strength of the encoded inference. This leads to a gradual view of inferences, meaning the inference is not encoded in a none-or-all fashion, but it can be encoded partially depending on the availability, strength and specificity of the involved features. It follows that, in order to map the information included in the mental representation of the event and its context, different kinds of retrieval conditions are necessary.

Thus, this gradual view of inferences requires the consideration of different types of measures. Traditionally, online measures are preferred because the main objective has been to show that inferences are part of the comprehension process. However, if we assume that the time course and the availability of the features mentioned by McKoon and Ratclif vary, then we need to measure inference at different moments in time. Some features are quickly retrieved from memory (*e.g.*, due to strong associates in the sentence) others require more computation (*e.g.*, more diffuse features with range of variations in their meanings). In STIs, alongside the verbal information conveyed by the sentence, other relevant information present in the inferential context are affecting the strength and the timing of the inference encoding. Part of that context is the photo of the

person, their role, the instruction of the task and the processing goals by it activated. All this information together with the sentence and the associative network backing up its comprehension are critical features leading to spontaneous trait inferences.

8.2 FOLLOW-UPS

8.2.1 *Inference and Dispositional Inference*

Although our research suggests that the empirical evidence in favor of qualitative differences between STI and STT is weak, we deem that there is some more fundamental difference between STI and STT. To explore this difference we need, first, to better understand what a dispositional inference is, and how it differs from mere trait activation. Inferring the trait from the behavior is one thing (assuming it is text-based activation), inferring a disposition about the actor of that behavior is a very different thing. In particular, we think that the trait changes the representation of that actor, meaning the trait is integrated into the overall representation of the person that enacted the behavior. On the other hand, the kind of association taking place in STT should not impact the representation of the person.

An interesting question is whether this distinction regarding the integration of the trait information into the representation of the person can be achieved within a connectionist model, that is, based on associations between nodes. The key answer to this question is the type of network that is being employed. In a localist connectionist model as the one we presented in Chapter 2, each concept (*e.g.*, person, trait) is represented by a single node that has a fixed symbolic meaning (Thorpe, 1998). The different nodes in a network are very similar with each other, the only aspect that varies being the activation state of the node (that can go from 0 to 1, for instance) and the connections between the nodes. And, because the meaning of a node is static, is not possible to simulate changes in the content of a representation. Whereas in distributed connectionist systems (or modularly distributed), which are usually considered biologically more plausible, a semantic entity is represented by a unique and distributed pattern of activity of microsemantics/nodes (that is, by multiple units). Each node in a distributed representation model has no meaning by itself. The internal structure of a representation is expected to be relatively similar to the struc-

ture of the representations to which this examplar is semantically similar in relevant aspects, and quite different from the structure of semantically different concepts (Browne & Sun, 2001). Importantly, that happens because different representations/concepts can share nodes (Page, 2000; Conrey & Smith, 2007). In a distributed model, a representation can be modified through activating or deactivating nodes such that nodes that are part of the pattern representing the trait can be incorporated into the representation of the actor. Moreover, the amount of modification of a certain pattern depends on the context that activates it, and the relevance or the salience of the person can be an aspect defining the amount of integration. Thus, incorporating new traits into a representation means that a slightly modified version of the same pattern will be created in order to better mimic the context.

A different way that would allow us to simulate qualitative differences between STI and STT in an associative network is by having labeled associations like the ones initially popularized by Collins and Quillian (1969) and Collins and Loftus (1975). According to these authors there are different kinds of links representing different kind of relationships between nodes (*e.g.*, "canary IS a bird" and "canary CAN fly"). This is, at some extent, similar to the distinction between the type of links suggested by some proponents of the dualistic view (Carlston & Skowronski, 2005). In this view, STIs are described as being based on labeled links (*e.g.*, "Mary IS smart") whereas the STTs being based on unlabeled links. However, in social cognition, labeled models are not very popular (Conrey & Smith, 2007) and the reason might be the fact that in this field the representation has a much more relevant role than relationships between representations. Moreover, it seems implausible to us that there could be so many types of links as type of relationships between concepts.

But, regardless of the category of the model we choose to simulate dispositional trait inferences, besides being able to distinguish between mere inferences and dispositional inferences and being a good description of STI and STT, for the model to be considered an upgrade and a gain for the literature, it must be able to explain all the previously found results. In other words, all the evidence we discussed regarding the differences between STI and STT should be reproducible within this model. Although these experiments, that were meant to support a very specific theoretical view, were shown to be anemic for such a purpose, they "have a life on their own" meaning they should be taken into account if a new model is to be proposed, as soon as they are part of

an error-correcting context as we believe the psychological science is (Garcia-Marques & Ferreira, 2011, p. 197). Moreover, the differences we found in attention and the way the behavioral information versus visual representation is processed, together with the inhibition that seems to occur in the STT must also be reproducible within this new model. Finally, the model should be falsifiable in terms of the possible and plausible results that can stem from it and, specially, those that cannot be expected from the model.

8.2.2 *Dissociations between STI and STT*

In this dissertation we also showed that people pay more attention to actors than to irrelevant persons presented with the behavior. Moreover, the attention correlates with the STI effect, suggesting that more fixations on the actor's face and more transitions between the actor and the behavior is translated into a stronger STI effect. Important clarifications regarding this relationship are necessary. We don't know whether it is the relevance of the actor that leads to more attention, that, in turn, results in larger associations between the actor and the trait, or if it is the relevance that triggers deeper processing and integrative goals, that results in two distinct and independent consequences, more attention being paid to the actor on one hand, *and* a stronger link between the actor and the trait on the other hand (the two things being correlated because of a common variable that is affecting both). If the relationship between attention and the STI effect is mediated by a goal triggered by the relevance, then manipulating attention paid to the faces (via, for example, the attractiveness of the faces or their presentation time) without affecting the actual relevance of the actors, should not affect STI whereas it should affect STT.

In Chapter 7 we introduced a new paradigm to study STI. And by using this paradigm, in the last experiment of the Chapter we found a dissociation between STI and STT. We suggested that an inhibitory process might be taking place in STT. The next step should be to further replicate the finding and test the inhibitory hypothesis. If an inhibitory process underlies STT, that would suggest that if there is something else going on beyond associations it is not in the STI (as suggested by the attributional explanation), but in STT and it would work in the opposite direction of an associative process (*i.e.*, to avoid the link between the trait and the irrelevant person). However, this inhibition might not be enough to eliminate the effect and that is why we still observe

STT in more traditional paradigms. It is also known, for example, that recognition tasks are not ideal (when compared with recall, for example) to detect inhibition because by presenting the inhibited concept for recognition, the inhibition might be released (MacLeod, 1998). This same inhibition might be easier to detect in the modified free association task because even though release from inhibition might occur for the trait itself, the effect of inhibition on the network might still be detectable.

Moreover, STI are sensitive to other informations about the actor, such the social group of the actor (Wigboldus, Dijksterhuis, & Van Knippenberg, 2003). A relevant question is whether the information about the communicator (his profession, for example) interferes with the inference of the trait from the behavior being communicated. We have shown that the way the trait is inferred from the sentence is not affected by the relevance, and so we don't expect the difference to reside in the extraction of the trait from the behavior, but in the integration of the different pieces of information. Exploring this aspect will talk directly to our main interest, that is, whether the trait is integrated differently in the actor's representation and communicator's representation. We expect an interaction between information about the actor and the trait inferred about this person, however no such interaction is expected between the inferred trait and the information about the communicator.

A complementary way of distinguishing the two phenomena is by investigating the differences between STI and STT in terms of brain activations and EEG components. We know that P300 waveforms are observed when a behavior violates a previously inferred trait, especially when a negative behavior violates a positive expectation (Van Duynslaeger, Sterken, Van Overwalle, & Verstraeten, 2008). This can be applied to STI and STT debate, such that in STI trials the two pieces of information (the positive and the negative behaviors) will regard the same actor, whereas in the STT trials, if the first behavior is about the communicator himself then for the second behavior this same person would be merely a communicator. We would expect to find a stronger inconsistency effect in STI, because in the STT the second information won't be integrated into the previously created representation of that person.

Finally, for future investigation we are interested in the role of context in STI and STT effects. If STI is sustained by the processing of the actor and its representation, then the context should not play an important role. In other words, if the context at retrieval does not reinstate the context

at encoding, the STI should not be affected. The opposite would happen in STT, since our hypothesis is that it depends more on the context and the stimuli temporal and spatial contiguity. Context is just an example, but other relevant perceptive versus conceptual manipulations should differently affect STI and STT.

As a final note, we consider crucial that more evidence is gathered about dissociations between STI and STT, in particular double dissociations, so more solid conclusions can be drawn about the distinction in the processes underlying STI and STT.

8.2.3 *Time course of STI*

A topic we mentioned a couple of times across the Chapters is the time course of inferences. As opposed to the text comprehension literature, where the time course of the inferential process was already investigated for some types of inferences (*e.g.*, predictive events, Calvo & Castillo, 1998), the time course of STI is unknown. Knowing the time STI needs to develop is crucial because it will inform our selection of the paradigms to be used in future research and tell us which ones are more effective in detecting STI and when. And differently from the text comprehension research, in STI there is an additional step that might also be dependent on time. More that just detecting the moment that the trait is activated during the reading of the verbal information, we need to know the time necessary for the trait to become connected to the actor and the time it takes for it to be integrated into the representation of the actor. Studying the time course of an inference might even shed some light on the difference between STI and STT, since simple associations as opposed to more elaborated inferences, might need less time.

Thus, in order to explore this aspect, the interval between the processing of the material and the measurement of the inference has to be manipulated and instead of using only immediate versus delayed measures we need to quantify the exact time needed for an inference to be complete.

8.2.4 *Hierarchical structures of traits and STI*

An additional aspect we would like to add to the Final Remarks regards the hierarchical structure of traits. We know, via work developed by Hampson and colleagues (1986), that broader constructs/traits enable one to predict more divers behaviors at a moderate level of accuracy, whereas narrow traits enable us to predict with high accuracy within a limited range of behaviors. These authors claim that traits are organized in hierarchical structures with levels representing superordinate traits, basic traits, subordinate traits, and concrete behaviors (see figure 27 for an illustration). Our hypothesis is that people, when dealing with strangers, as in most of the spontaneous trait inference studies, sacrifice accuracy in order to predict a higher range of behaviors. However, we expect different patterns of activation of the hierarchical structure in STI (where the perceiver can actually benefit from activating additional information), when compared with STT (where the "motivation" to activate more traits in the hierarchical structure is low). In functional terms, when an actor it being perceived, it can be advantageous to create a rich representation of the person. And one way of doing it, is not just by categorizing the behavior and the person in concrete traits ("charitable" because he donated money to a charity institution), but also activating broader and higher level traits in the hierarchy (*e.g.*, "helpful" and "good"). The activation of more general traits might lead to a broader range of predictions about future behaviors of the actor (other helpful and good behaviors will be expected).

Hampson and colleagues (1986) suggested a couple of different ways to investigate the hierarchy of traits. One way of defining the level in the hierarchy of a certain trait is by counting the number of different behaviors people can generate for each trait. Because higher-level traits (*e.g.*, "good") are more abstract they can be described by more behaviors. A different way of allocating the trait in the hierarchy is by measuring the prototipicality and diagnosticity for each behavior in relation to the trait. It is known that superordinate categories have less diagnostic members and also fewer highly prototypical exemplars. Finally, a third type of task they suggest is instructing the participants to select the most meaningful statement from pairs of two: “to be talkative is a way of being extroverted” versus “to be extroverted is a way of being talkative”. These tasks can allow us to investigate the hierarchical structure of traits and more importantly to explore how is this structure activated by different context, such intentional inferences, STI and/or STT.

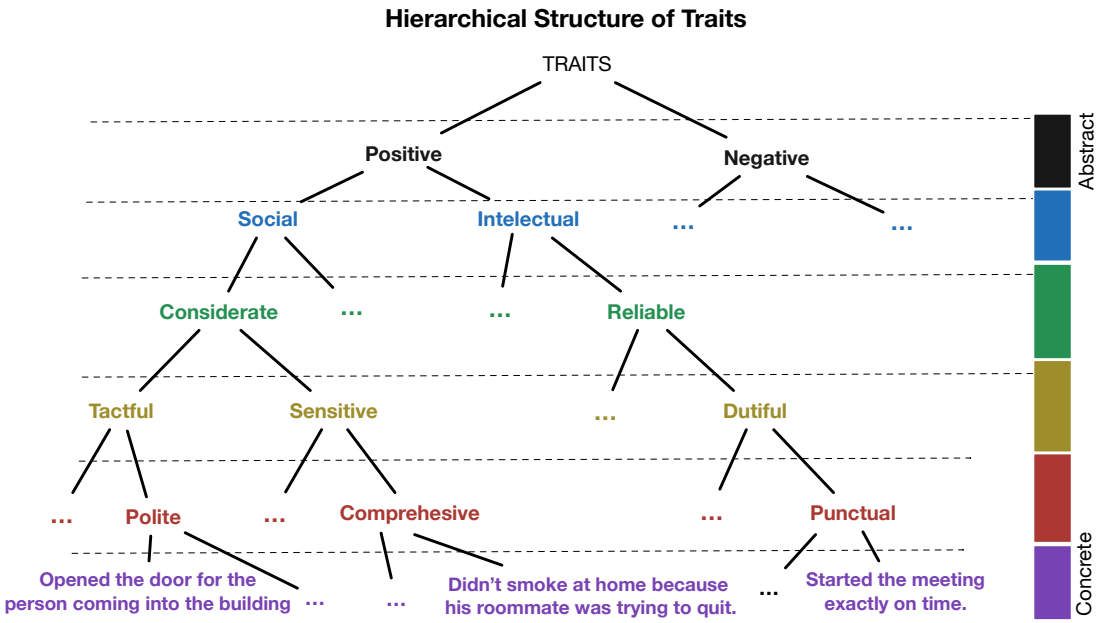

Figure 27: Illustration of the hierarchical structure of traits.

8.3 CONCLUSION

We would like to finish this dissertation with a story, and if we did not convince you by now that trait inference matters, we hope this story convinces you of that.

People have been fascinated for a long time by the possibility of having wild animals, from lions or ostriches to mice or lizards, as pets. Domesticating a wild animal is, however, very difficult, if not impossible, to achieve, unless you are a soviet scientist, named Dmitry Belyaev.

Domestication is a evolutionary process that resulted from selective breeding that occurs over a long period of time. Belyaev was determined to trace the evolutionary pathway of domesticated animals, and to reproduce it experimentally. For long decades his team bred silver foxes under the cover that they were breeding foxes to improve the fur for coats (genetic research was banned in the USSR during Stalin's era). Critically, this team of researchers used trait inferences to accomplish their revolutionary project. An initial group of foxes was selected by the researchers based on their behavior and tolerance towards humans, that is, they were evaluated based on *inferred traits* such docility and tameability. This trait evaluation method continued for many generations of foxes during which the same criteria of selecting the most tame individuals to be parents in the following generation was applied. Surprisingly, after fifty generations of continu-

ous selection, evolution was reproduced. The foxes from the generation corresponding to year 2005-2006 started to understand human cues, they were friendly, they whimpered and licked researcher just like puppies did (Trut, Oskina, & Kharlamova, 2009). This is a good example of how useful the tool of trait inference can be in science as it is in our everyday life.

8.4 REFERENCES

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295.

Bassili, J. N., & Smith, M. C. (1986). On the spontaneity of trait attribution: Converging evidence for the role of cognitive strategy. *Journal of Personality and Social Psychology*, *50*(2), 239–245.

Bodenhausen, G. V., & Hugenberg, K. (2009). Attention, perception, and social cognition. In F. Strack & J. Forster (Eds.), *Social cognition: The basis of human interaction* (pp. 1–22). Philadelphia, PA: Psychology Press.

Browne, A., & Sun, R. (2001). Connectionist inference models. *Neural Networks*, *14*(10), 1331–1355.

Butler, L. T., & Berry, D. C. (2001). Implicit memory: Intention and awareness revisited. *Trends in Cognitive Sciences*, *5*(5), 192–197.

Calvo, M. G., & Castillo, M. D. (1998). Predictive inferences take time to develop. *Psychological Research*, *61*(4), 249–260.

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.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 240–247.

Conrey, F. R., & Smith, E. R. (2007). Attitude representation: Attitudes as patterns in a distributed, connectionist representational system. *Social Cognition*, *25*(5), 718–735.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, *58*(3), 306–324.

Crawford, M. T., Skowronski, J. J., & Stiff, C. (2007). Limiting the spread of spontaneous trait transference. *Journal of Experimental Social Psychology*, *43*(3), 466–472.

Crawford, M. T., Skowronski, J. J., Stiff, C., & Leonards, U. (2008). Seeing, but not thinking: Limiting the spread of spontaneous trait transference ii. *Journal of Experimental Social Psychology*, *44*(3), 840–847.

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.

Devine, P. G., Hamilton, D. L. E., & Ostrom, T. M. (1994). *Social cognition: Impact on social psychology.* San Diego, CA: Academic Press.

Donald, H., Joseph, C., Don, C., David, G., & Steven, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, *101*(1), 116–122.

Dostoevsky, F. (1996). The idiot. Hertfordshire, UK: Wordsworth Editions Limited (Original work was publiched in 1869).

Garcia-Marques, L., & Ferreira, M. B. (2011). Friends and foes of theory construction in psychological science vague dichotomies, unified theories of cognition, and the new experimentalism. *Perspectives on Psychological Science*, *6*(2), 192–201.

Goren, A., & Todorov, A. (2009). Two faces are better than one: Eliminating false trait associations with faces. *Social Cognition*, *27*(2), 222–248.

Hampson, S. E., John, O. P., & Goldberg, L. R. (1986). Category breadth and hierarchical structure in personality: studies of asymmetries in judgments of trait implications. *Journal of Personality and Social Psychology*, *51*(1), 37–54.

Hintzman, D. L., & Ludlam, G. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory and Cognition*, *8*(4), 378–382.

Hourihan, K. L., & MacLeod, C. M. (2007). Capturing conceptual implicit memory: The time it takes to produce an association. *Memory and Cognition*, *35*(6), 1187–1196.

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

Kintsch, W. (1993). Information accretion and reduction in text processing: Inferences. *Discourse Processes*, *16*(1-2), 193–202.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Pressniversity press.

Kruschke, J. K. (1992). Alcove: an exemplar-based connectionist model of category learning. *Psychological Review*, *99*(1), 22–44.

Maass, A., Colombo, A., Colombo, A., & Sherman, S. J. (2001). Inferring traits from behaviors versus behaviors from traits: The induction–deduction asymmetry. *Journal of Personality and Social Psychology*, *81*(3), 391–404.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Intendisciplinary approaches.* (pp. 1–57). Mahwah, NJ: Lawrence Erlbaum Associates.

McClelland, J. L., & Rumelhart, D. E. (1989). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. MIT press.

McKoon, G., & Ratcliff, R. (1989). Semantic associations and elaborative inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 326–338.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, *51*, 297–384.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of experimental psychology: general*, *106*(3), 226–254.

Page, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences*, *23*(04), 443–467.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *The Journal of Neuroscience*, *4*(7), 1863–1874.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? a comment on theory testing.

Psychological Review, *107*(2), 358–367.

Roediger, H. L., & Blaxton, T. A. (1987). Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. *Memory and Cognition*, *15*(5), 379–388.

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.

Smith, E. R. (1996). What do connectionism and social psychology offer each other? *Journal of Personality and Social Psychology*, *70*(5), 893–912.

Thorpe, S. (1998). Localized versus distributed representations. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 549–552). Cambridge, MA.

Trut, L., Oskina, I., & Kharlamova, A. (2009). Animal evolution during domestication: the domesticated fox as a model. *Bioessays*, *31*(3), 349–360.

Van Duynslaeger, M., Sterken, C., Van Overwalle, F., & Verstraeten, E. (2008). Eeg components of spontaneous trait inferences. *Social Neuroscience*, *3*(2), 164–177.

Weber, E. U., Goldstein, W. M., & Busemeyer, J. R. (1991). *Beyond strategies: implications of memory representation and memory processes for models of judgment and decision making.* Lawrence Erlbaum Associates, Inc.

Wigboldus, D. H., Dijksterhuis, A., & Van Knippenberg, A. (2003). When stereotypes get in the way: Stereotypes obstruct stereotype-inconsistent trait inferences. *Journal of Personality and Social Psychology*, *84*(3), 470–484.