UNIVERSITY OF PADOVA

Department of General Psychology

Department of Developmental Psychology and Socialization

Master Degree in Applied Cognitive Psychology

Final dissertation

**The influence of motivation on the unconscious
reprocessing of automatic evaluations: the case of
prejudice**

Supervisor

Prof. Luigi Castelli

Co-supervisor

Prof. Rob Holland

Candidate: Tiziano Suran

Student ID number: 1080809

Academic Year 2014/2015

INDEX

1. Introduction	1
2. Measuring attitudes in the domain of racial prejudice	4
2.1 Implicit measures	5
2.1.1 The implicit association test	5
2.1.2 The evaluative priming	7
2.1.3 Projective measures: the affect misattribution procedure.....	9
3. The interplay between implicit and explicit attitudes	12
3.1 The associative-propositional evaluation model.....	13
3.2 The MODE model.....	16
3.2.1 Motivation to control for prejudice.....	18
3.2.2 Internal and external motivational factors	19
3.3 The iterative-reprocessing model.....	20
4. Conscious and unconscious processing	24
5. Unconscious control of prejudiced reactions	30
6. The present study	30
7. Pilot study.....	34
7.1 Method	34
7.1.1 Participants.....	34
7.1.2 Design	34
7.1.3 Materials and procedure.....	35
7.2 Results	36

7.3 Discussion	38
8. Main study	40
8.1 Method	40
8.1.1 Participants.....	40
8.1.2 Design	41
8.1.3 Materials and procedure.....	41
8.2 Results	44
8.2.1 Predicting explicit attitudes	48
8.3 Discussion	51
9. Conclusion.....	54
REFERENCES	55

APPENDIX

1. Introduction

On February 4, 1999, four police officers of the Street Crime Unit of the New York Police Department were patrolling Wheeler Avenue, a street in the neighbourhood of South Bronx. In front of an apartment, they spotted a Black individual, whom they thought matched the description of a suspect, and decided to stop him. On the other hand, the Black person was all but a criminal; he was Amadou Diallo, a young immigrant from Guinea, working at the time as a peddler in lower Manhattan. When the police, dressed in plainclothes, started approaching him by ordering him to freeze, he mistook them for a group of robbers. However, the policemen did not know this, as all they saw was a suspect Black guy. When, Diallo started to pull out his black wallet from his pocket, the policemen, in that split-second, identified the object as a weapon, opened fire on the 22 year old and killed him (Gladwell, 2007).

After the killing of Diallo, many people were accusing the NYPD of basing their decision to open fire on the prejudice towards Blacks, stereotypically depicted as aggressive and as criminals. But to what extent does prejudice influence our behaviour? As it will be described in the following chapters, the influence of prejudice and stereotypes often occurs outside people's awareness, making it really difficult to avoid their effects. However, some people may be motivated to overcome this unwanted influence, and they often succeed. But is it necessary to be aware of the influence of a prejudiced reaction in order to counter it? This successful countering is often attributed to deliberative processes that require awareness to be put in motion. Once a motivated individual is aware of an automatic reaction, he will be able to counter it. Nevertheless, given the automaticity of such reactions, and the vast amount of processes that occur unconsciously, conscious awareness may not be that crucial when it comes to controlling unwanted prejudiced reactions.

From an evolutionary perspective, quick evaluations may have been in the past a crucial factor for determining survival in dangerous environments. In prehistoric times, deciding quickly whether the thing that was approaching us was another human or a lion potentially looking for a prey was a matter of life and death. Through natural selection, human cognition has gradually adapted to respond quickly to stimuli in order

to determine whether they were harmless or dangerous (Cacioppo & Berntson, 1999). In modern times, even though our environment and what we experience as threats may not be as dangerous as they may have been thousands of years ago, we still use the same evaluative mechanisms in our everyday life.

We constantly evaluate what surrounds us. The most basic evaluation is related to the valence of an object, which determines our positive or negative disposition towards it. This disposition is referred to as an attitude (Eagly & Chaiken, 1993; Wilson, Lindsey, & Schooler, 2000). We may therefore have a positive or a negative attitude towards something, being it a person, a political issue or a consumer product. When we have to decide about the nature of the object in question, our attitudes play a central role in the process. The outcome of this evaluation will then determine our future behaviour towards that object.

It is not necessary to personally know an individual in order to have an attitude towards him. This is due to our reliance on the processes of categorization. In order to make our world simpler, we automatically categorize the objects that surround us (Bruner, 1957). We do the same for people within our social environment, like identifying individuals as being members of a certain group (Tajfel, 1969). We may base our categorization of individuals on different factors, among which gender, age or race. For this reason, we associate every individual we meet with the group we think he belongs to. Therefore, we may not have an attitude about a certain person that we don't know, but we do have attitudes towards the group we associate him with. In this way, the attitude towards a group is generalized to all its members. When such attitude is negative, we define it as prejudice.

Not only it is unnecessary to know an individual to have an attitude towards him, we don't even need to personally know any of the members of his group in order to have an attitude towards them. In fact, attitudes can be built from the simple information we acquire from our social environment (Hilton & von Hippel, 1996), from our parents (Castelli et al., 2007, 2009) and through mass media (Tukachinsky, Mastro, & Yarchi, 2015). Based on the type of information conveyed, we start to build a pre-conceived idea about the characteristics of that group. This knowledge is defined as a stereotype (Hilton & von Hippel, 1996). However, the problem arises when our pre-conceived notions and attitudes towards a group are in conflict with what we know

about the individuals that belong to it from personal contact (and, at least according to my experience, in the great majority of cases they are). We may also be motivated to avoid attributing the stereotype that we associate a group with to its members when we interact with them or when we talk about them because we may think that is wrong. But even if we explicitly reject the validity and endorsement of stereotypes, their activation is directly linked to the process of categorization, which can still influence certain behaviours (Devine, 1989).

Studies suggest that when we automatically engage into evaluative processing, the less time we have to act following the initial automatic evaluation, the more will our behaviour be influenced by it. However, when we do have more time to elaborate our evaluations, the initial automatic response will be re-evaluated in the light of an increasing amount of information. This control was usually associated with conscious awareness of the object that activated the reaction. However, it has been suggested that, just like the activation of attitudes, their re-evaluation also occurs automatically, and the more the re-evaluative cycles, the higher the chance of becoming aware of the evaluation (Cunningham & Zelazo, 2007). Taking inspiration from studies involving unconscious elaboration (e.g., Sklar et al., 2012), the research presented here is aimed at exploring the possibility to re-evaluate automatically activated attitudes according to explicit beliefs without the aid of conscious awareness. Was it the case, it would be also possible to better understand the mechanisms behind the lack of reportability of some of those automatically activated attitudes that in particular cases guide our behaviour contrarily to our will.

In the present thesis, I start by introducing several measures of implicit attitudes in the domain of racial prejudice. Then, I will provide an overview of the models that explain the relationship between implicit and explicit prejudice and its moderators. Subsequently, the concept of unconscious processing is explained, together with its possible application in the domain attitudes. The main research is then presented, describing its piloting, the main hypothesis, the methods and the results. In the last section, the results are further discussed by including suggestions for future studies.

2. Measuring attitudes in the domain of racial prejudice

The measurement of attitudes has been one of the main research topics in the domain of intergroup perception and prejudice. Initially, their assessment was carried out by using simple questionnaires, in which respondents were asked to indicate their disposition towards specific groups, like for example Blacks. However, soon the question emerged about the genuineness of the reported attitudes. One of the reasons for this was the idea that people may not be willing to express an adverse disposition toward Blacks due to social norms that condemn any expression of racism. This statement may nowadays seem obvious, but is important to know that during the first half of the past century, racial segregation was considered to be a normal thing. Blacks had very few rights compared to Whites and racism did not have the negative connotation it has today. It was only in the second half of the 20th century that the situation started to change. With the 1960s Civil Rights Movement the process of racial desegregation was put in motion, Black Americans started acquiring more and more rights, until they were finally considered officially equal to Whites. Subsequent national surveys found that the proportion of White Americans stating that they would vote for a qualified president that was Black had tripled after that transitional period. The same was true for the reported support for racial integration in schools (Schumann, Steeh, & Bobo, 1985). Nevertheless, the lack of reported prejudice did not mean the absence of it. In fact, it was believed that a prejudice that may at that point manifest itself in more subtle ways needed more subtle measures in order to detect it (Gaertner & Dovidio, 1986).

One of the initial efforts towards measuring prejudice more indirectly consisted in the modification of the questionnaires. The Modern Racism Scale (MRS; McConahay, 1986) was introduced with the intent of assessing people's attitudes in such a way that would not make them feel in apprehension that their answers may be considered racist or prejudiced. For example, people may be asked to report how much they agree with certain beliefs that White Americans may hold towards Black Americans like "It is easy to understand the anger of Black people in America", or "Blacks are getting too demanding in their push for equal rights". Even though the

“non-reactive” measure of prejudice that the MRS represents may have been more indirect than simply asking about one’s prejudice, the attitudes expressed in this way could have still been adapted to appear more socially acceptable. In a study, the MRS was administered first in a collective session and then, a couple of months later, it was re-administered individually to the same participants. Furthermore, in the second session, for one randomly selected group of participants the experimenter administering the scale was black and for another was white. Overall, the scores in the second individual session were lower than those in the first collective one. Additionally, the greatest decrease in MRS score was found in the black experimenter condition (Fazio et al., 1995). These findings show that, also on the MRS, people are very well capable of controlling their responses and adjust them according to the situation.

2.1 Implicit measures

Taking inspiration from research involving implicit memory (e.g., Shiffrin & Schneider, 1977), new measures of attitudes were introduced in the ’90. This time, instead of relying on people’s self-self reports, the new measures relied on implicit evaluations. Unlike with the explicit evaluations, in which the evaluation is indeed explicit in people’s response, implicit evaluations require their meaning to be inferred from other indicators, like participants’ response latencies within evaluative tasks. The attitudes measured in this way were therefore defined as implicit attitudes (Greenwald & Banaji, 1995).

2.1.1 The implicit association test

One of the most famous implicit attitude measures is the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998), which measures the strength of the association between categories and valences. What qualifies it as an implicit measure is the fact that the attitudes are inferred from participants’ response times. The IAT consists of five blocks of trials: three practice blocks and two critical blocks. Taking the race IAT as an example, in the first practice block participants are required to categorize pictures of white and black individuals appearing one by one in the middle of the screen based on their racial group membership (Black vs. White). In the second practice block, participants are asked to categorize, in the same way, positive and negative words based

on their valence (Good vs. Bad). The third block is the first critical block. This time, both adjectives and pictures appear one by one on the screen. However, each category label is paired with a valence attribute on the same response key. Therefore, the subject will categorize Black faces and Bad words by pressing one key, and White faces and Positive words by pressing the other key. This critical block would be called congruent, as, at least for white participants, it is expected that its category/attribute pairings are more strongly associated in memory, making it therefore easier and faster to complete the task. The fourth block is the same as the first, and the fifth one is what in this case would be the incongruent critical block. In this block the pairings are changed by swapping the category labels. An example of the four block types can be seen in Figure 1.

Figure 1. Example of the four types of blocks contained in the IAT: practice block with categories (a), practice block with adjectives (b), congruent critical block (c), incongruent critical block (d).

As participants' task is to be as quick and accurate as possible, it is expected that the incongruent block will be more difficult and will take more time to complete as the paired categories and attributes are weakly associated. To avoid attributing the inhibition to the pairings trained in the previous congruent block, the order of the two critical blocks is usually counterbalanced between participants. What has been found by using this measure is that, on average, White participants are faster in the trials of the congruent blocks compared to the ones in the incongruent blocks, showing a stronger association between the concepts paired in the congruent blocks (e.g., Greenwald, McGhee, & Schwartz, 1998; Olson, & Fazio, 2004; Foroni & Semin, 2012).

2.1.2 The evaluative priming

Another implicit measure is the evaluative priming paradigm (Fazio, Jackson, Dunton, & Williams 1995). This test is based on the idea that the perception of an attitude object, which serves as the prime, automatically activates a spreading activation process of the associated concepts related. This activation subsequently influences the response to a stimulus that follows the prime (Neely, 1977). In the evaluative priming task, people are asked to evaluate the valence of certain words that appear on a screen by pressing one of two buttons, one representing "negative" valence and the other "positive". The words, in this case the adjectives, are unambiguous in terms of their valence, therefore are generally regarded as either positive (e.g. attractive, wonderful, likeable) or negative (e.g. offensive, disgusting, annoying). Before the presentation of each word, a face of a White or a Black person briefly appears in the middle of the screen. When measuring implicit racial attitudes, the idea behind this procedure is that seeing the picture of an individual that belongs either to the Whites or to the Blacks group, automatically activates a positive or negative evaluation. The automatic evaluations in this way would make the concepts that are also liked to it in terms of valence more accessible to the subject, and therefore facilitate the categorization. This facilitation, given the brief interval between the onset of the prime and the adjective, is less likely to be influenced by deliberative processes, relying much more on automatic responses. If an individual automatically evaluated Blacks more negatively, he is expected to be quicker in categorizing a negative rather than a positive adjective when

these adjectives are presented immediately after a picture with a face of a Black individual.

What is commonly found is that White participants are on average significantly faster to categorize a positive word as such when preceded by a White face rather than a Black face, and faster to categorize a negative word as such when preceded by a Black rather than a White face. This effect is found when the primes were presented supraliminally (Fazio et al., 1995; Foroni & Semin, 2012) and also when they were presented subliminally (Dovidio et al. 1997; Olson & Fazio, 2002). As a note, the degree of implicit prejudice found in this way was not related to participants' scores on explicit self-report measures.

Even though both the IAT and the evaluative priming are supposed to measure the same constructs, like attitudes toward social groups, there is often a low correlation between the two (Fazio & Olson, 2003). In a study aimed at exploring these differences (Foroni & Semin, 2012), participants had to complete the two implicit measures with race-related stimuli while a group of them had to contract their facial muscles by holding a pen between their teeth. As previous research found that after the presentation of valenced primes, people tend to activate certain facial muscles (Dimberg, Thunberg, & Elmhed, 2000), it was argued that this mechanism may play a role in a priming paradigm such as the evaluative priming. Such mechanism was supposed to serve as a proprioceptive feedback for the individual, facilitating the response to a subsequent target of the same valence. The muscle participants were blocking was the zygomatic muscle (used when smiling), which is automatically contracted after the presentation of positively valenced stimuli and inhibited after negatively valenced ones. Results showed that, while the IAT-effect remained unaltered between the 'pen' and 'no-pen' conditions, the priming effect in the evaluative priming ceased to be significant in the 'pen' group. These findings led to conclude that the IAT and the evaluative priming rely on different mechanisms, suggesting that while the former relies on the associative strength between the category labels, the latter would be based on more immediate affective responses to the specific stimuli (Foroni & Semin, 2012). For this reason, a measure like the evaluative priming may be more suited for the exploration of automatic evaluative processing.

2.1.3 Projective measures: the affect misattribution procedure

At the beginning I made an example of racial stereotype by describing the shooting of Amadou Diallo. What happened during that night can be explained without necessarily involving racial prejudice and stereotypes - which can indeed play an important role. The crime rate in the Bronx is much higher when compared to the rest of New York, and Diallo knew it. For this reason, when Diallo saw four armed men approach him, the first thing that came to his mind was the most accessible one in that context: they wanted to rob him. The same reasoning can be applied for the policemen. When they were about stop Diallo, in their minds there was the idea of capturing a criminal who was potentially carrying a weapon with him. When Diallo pulled out a hardly recognizable black object from his pocket, what the officers saw was exactly what they were thinking about. For the officers, Diallo's wallet immediately became a handgun. The idea of "gun" was so strong that in that brief and potentially life-threatening moment they projected it to the black wallet.

The misattribution of the gun to Diallo's wallet may have been influenced by prejudice. In a task similar to the evaluative priming, participants saw pictures of white and black individuals immediately followed by either a gun or a neutral object. Their task was to categorize the second picture either as a tool or a weapon (Payne, 2001). What the results showed was that participants were much faster at categorizing the guns that followed a black face than when followed by a white one. Moreover, when they were instructed to categorize the object within half a second after its appearance, they made much more errors in mistaking the tools for guns when they were preceded by a black face. This effect, named as the weapon-identification-bias, was found both in black and in white participants. These findings were replicated in several other studies using different methods, all of them showing the same weapon bias in which Blacks are associated with the concept of guns (Greenwald, Oakes, & Hoffman, 2002). Moreover, in some cases the bias was found to be equally strong in Black as in White participants (Correll, Park, Judd, & Wittenbrink, 2002). But if the effect of a face is strong enough to shape the way we interpret a clearly distinguishable object, how would it affect the perception of an ambiguous one? Can this bias be used to assess the degree of one's implicit prejudice?

The idea that people project their internal state to the external world is not new. Tests like the Rorschach inkblot test and the Thematic Apperception Test (TAT) are interpreted as relying on people's projections of their internal state. In the first one, people are asked to describe what they see when looking at inkblots; on the TAT, they are asked to describe an ambiguous scene. Their descriptions are then interpreted in order to assess the subjects' personality.

Starting from the idea of projection of one's internal states to external objects, Payne and colleagues developed a paradigm with the purpose of measuring implicit attitudes. The idea they had was that if people are asked to judge an ambiguous symbol that is preceded by a prime, their judgments will be influenced by the valence of the prime. This new paradigm was named the affect misattribution procedure (AMP; Payne et al., 2005). The structure of the AMP is similar to the one of the affective priming paradigm. In fact, a prime is followed by a target on which the participant has to respond with one of the two response keys (see Figure 2).

Figure 2. Example of stimuli used in the race AMP in temporal succession. From left to right: prime, Chinese pictograph, pattern mask.

In the AMP, however, the target that appears after the prime is an incomprehensible Chinese character. Participants are instructed to simply attend the face primes and judge only the Chinese characters based on their visual pleasantness. For this purpose, they are given two response options, one button for judging “unpleasant” characters and one for the “pleasant” ones. The final score is then calculated by subtracting the proportion of “pleasant” responses to the pictographs that followed White primes from the proportion of those who followed the Black ones.

As a measure for implicit attitudes, the AMP has been used in several domains, like political behaviour (e.g., Carraro, Gawronski & Castelli, 2010), eating behavior (e.g., Hofmann, van Koningsbruggen, Stroebe, Ramanathan, & Aarts, 2010) and prejudice (e.g., Cooley, Payne, & Phillips, 2014; Imhoff & Banse, 2009; Payne et al., 2005). When used to study racial prejudice, it has been found that on average participants tend to categorize as less “pleasant” the Chinese characters that follow Black faces as compared to those that follow the White ones. Interestingly, the priming effects in the AMP are still significantly obtained when participants are warned beforehand about the potential influence of the primes on their judgments (e.g., Payne et al., 2005).

3. The interplay between implicit and explicit attitudes

At this point, it would not be plausible to claim that implicit attitudes represent people's "true" attitudes. At the beginning I ascribed the introduction of implicit measures to the need to overcome the social-desirability bias. However, one would in this case expect to find a strong correlation between implicit and explicit measures when social-desirability is low. Nevertheless, while some studies support the idea that social-desirability mediates the implicit-explicit relationship (see, Nosek, 2005; Riketta, 2006), others found either no effect (e.g., Egloff & Schmukle, 2003; Hofmann et al., 2005), or even the opposite pattern (e.g., Hofmann et al., 2005; Riketta, 2006). Still, the uncontrollable, automatic and (often) unconscious nature of implicit attitudes does not necessarily classify them as "true". There are different reasons why individuals may explicitly report an altered evaluative response. Next to the fear of social disapproval, a person may not report any explicit prejudice because he holds nonprejudiced beliefs. If this individual does have a strong implicit negative attitude that contrasts his beliefs, it will be hard for him to overcome his automatic reaction if he had to choose his behaviour quickly. However whenever there will be enough time to overcome the automatic response, like when responding to a questionnaire or choosing the appropriate words in a conversation with an individual of a different race, our beliefs, or explicit attitudes, will eventually prevail.

The lack of correspondence often found between implicit and explicit attitudes (for a review, Hofmann et al., 2005) lead to research what parts of people's behaviour they predicted, and which of the two types of attitudes prevail in what situation. For example, while explicit attitudes influence more controllable aspects of behaviour, like the choice of words in a conversation with an individual of another race, implicit attitudes were found to influence less controllable aspects of behaviour, like for instance peoples' blinking rate, eye contact or amount of self-touching (Dovidio et al., 1997; Olson & Fazio, 2007). Behind these findings there are several models that try to explain them. The most influential ones are defined as the dual-process models. The main idea behind these models is that people possess two distinct evaluative systems. Under this perspective we can therefore distinguish between implicit and explicit evaluations, or evaluative systems. While the former system is usually described as fast, automatic and often unconscious, the latter is slower, requires deliberation and awareness (Bargh et al.,

1996; Greenwald & Banaji, 1995; Wilson, Lindsey, & Schooler, 2000). It is important to note that the distinction between these two systems is not often precisely defined, in the sense that both evaluative systems have some characteristics in common. The models that I will describe, all valid and effective at explaining a vast range of findings, differ on the degree to which implicit and explicit evaluative systems are distinct from one another, with the third (and last) model even denying the dual view and claiming the existence of a single evaluative system.

3.1 The associative-propositional evaluation model

The associative-propositional evaluation (APE) model, in explaining the differences between implicit and explicit evaluations, distinguishes between two distinct mental processes (Gawronski & Bodenhausen, 2006, 2007, 2011, 2014). Implicit evaluations are described as relying on *associative* processes, whereas the processes that underlie explicit evaluations are *propositional*.

Associative processes work under the principle of activation of mental associations, which are generated through experience coming from the individual's distant and recent past (Castelli et al., 2010). The spread of activation is driven by the similarity of features between the initial stimulus and our mental representations. For example, if the features of the face of an individual resemble those of our mental representation of Blacks, then the concept of Blacks will be activated, together with all the concepts associated to it. The concept of "Blacks" could be therefore linked to all the stereotypes about Blacks. However, not all the associations with Blacks are activated each time. In fact, the associated concepts are activated depending also on the context in which we are in that moment. Encountering a black person in the middle of the day at a running track will activate different associations from those that would be activated if we encountered him at night in a dark street of the Bronx. Therefore, temporal and spatial contiguity also play an important role for associative processes. Importantly, these associations do not need to be considered valid by the individual in order to be activated (Devine, 1969).

When engaging into propositional processing, the reactions elicited by the activated concepts are translated into statements. For example, a negative reaction

evoked by seeing an object could be translated into the statement “I dislike this object”. Propositional processes further assess the subjective validity of the generated statement in the light of other temporary activated propositions related to it and previously validated. For instance, if the initial negative reaction was related to an African American and translated into “I dislike African Americans”, it could be in conflict with other activated propositions, both evaluative (e.g., “Evaluations of disadvantaged groups are wrong”) and non-evaluative (e.g., “African Americans are a disadvantaged group”). The resolution of this inconsistency will be determined by the negation of one of the activated prepositions. For example, if the rejected preposition is the one evoked by the negative reaction (e.g., “I dislike African Americans” transformed into “I like African Americans”) there will be less correspondence between implicit and explicit evaluations. In this case, implicit evaluations will be negative and explicit evaluations will be positive. If the rejected preposition is one of the other two (e.g. “Evaluations of disadvantaged groups are ok”, or, “African Americans are not a disadvantaged group”), the correspondence between implicit and explicit measures will be higher and in this case both negative. This “bottom-up” influence of associative processes can work also in the opposite direction in a “top-down” fashion. The selective retrieval of positive information from memory that would validate a certain unprejudiced proposition can also activate associations, thus having again a greater correspondence - in this case in a positive direction - between implicit and explicit evaluations (Peters & Gawronski, 2011).

Top-down processes that imply the negation of the proposition formed from automatic reactions can also have “ironic effects”. The proposition “women are bad drivers”, when negated, can produce the proposition “women are not bad drivers”. Nevertheless, the same proposition will again contain the concepts of “women” and “bad drivers”, which, regardless of the negation in between, will strengthen the link between the two through associative processes. The same association can occur when watching a pharmaceutical commercial. Although the drug is supposed to eliminate or alleviate a negative condition (like an antibiotic for treating pneumonia) it will be automatically associated with the negative affective response that the condition may evoke. Furthermore, the absence of the notion of negation in associative processes has been found also for simple primes presented subliminally. More precisely, when

smokers were subliminally presented with “no smoking” signs, their approach tendencies towards smoking-related stimuli increased, as if the negation was not processed (Earp et al., 2013; see also, Deutsch, Gawronski, & Strack, 2006).

When conflict between the two processes arises, there are multiple factors that come into play. One important factor consists of the available cognitive resources. Associative processes are characterized by a greater automaticity, which consequently requires less cognitive resources in order to be activated. When these resources are reduced, people’s behaviour will be much more influenced by the reactions elicited by the associative system. In this situation, automatic evaluations will be more likely to determine behaviour. For example, by performing a parallel memory task, the nature of a white individual’s interaction with a black experimenter reflected much more his scores on implicit measures rather than self-reported attitudes (Hofmann, Gschwendner, Castelli, & Schmitt, 2008).

A second factor is the time factor. As associations are transformed into propositions only after they have elicited an initial evaluative response, more time will be needed for propositional processes to perform a validation. Therefore, if people are put under time pressure, their behaviour and explicit evaluative responses will again be more influenced by the effects of associative processes (see, Koole, Dijksterhuis, & van Knippenberg, 2001; Ranganath, Smith, & Nosek, 2008).

The third factor that the APE model considers is people’s motivation. However, the model interprets them as a proposition entering into play during the process of resolution of inconsistencies. Taking the previously made example with negative reactions to African Americans, people’s goal to be (or appear) unprejudiced can be conceptualized as a proposition like “I don’t want to evaluate disadvantaged groups negatively”. However, if, when resolving the inconsistency, the proposition “African Americans are a disadvantaged group” becomes negated (“African Americans are not a disadvantaged group”), then the motivational factor will have a lesser impact.

3.2 The MODE model

The influence of implicit and explicit attitudes on behaviour is also analyzed by the MODE model proposed by Fazio (1990). The acronym stands for Motivation and Opportunity as Determinants of behaviour, and what it suggests is that the predominance of either implicit or explicit attitudes in directing a person's behaviour are its motivation to overcome the automatic response and the opportunity to do so. The opportunity factor here is intended as the availability of cognitive resources and the time that a person has to enact the required behaviour. The idea is similar to the one proposed by the APE model, however, there is a much stronger importance given to the motivational factor. According to the MODE model, the processes that are involved in translating attitudes into behaviour are two: the spontaneous and the deliberative.

Spontaneous processes can guide behaviour without the individual's intentionality or awareness, and their working principles are similar to the ones of associative processes in the APE model. The automatically activated attitude evoked by an object can bias the individual's perception about it, as it may elicit a positive or a negative automatic evaluation that will influence the way in which new information about the object is processed. However, when an attitude object can be evaluated in different ways, the person's most accessible attitude will be more likely to prevail on others. For example, a rifle can be seen as a killing weapon or as a sport tool. Depending on which of the two categories evokes stronger, more accessible attitude, the individual will more likely categorize the rifle as belonging to that category, which will subsequently influence the way he perceives it and evaluates it. The same is true when dealing with people. As it is possible to categorize an individual in several ways, like on the basis of gender, age and race, the most easily accessible category will more likely prevail on the others (Smith, Fazio, & Cejka, 1996). In this way, the choice of our behaviour can become the result of spontaneous processes.

When an individual is motivated to avoid acting according to his automatically-activated attitudes, he can deliberate on his behaviour and choose to inhibit it, or at least influence the outcome of the automatic evaluation process. Nevertheless, this inhibition will occur only when individuals have the opportunity to do so. In short, the MODE model explains the effect of attitudes on behaviour in the following way: automatically-

activated attitudes can determine behaviour unless the person is motivated to do contrariwise. The motivation to counter the automatic reaction will have an effect on the reaction only if there is the opportunity (time and resources) to do so (see Figure 3).

Figure 3. Scheme of the MODE model. Motivational factors will influence the effect of an automatically activated attitude on behaviour only if opportunity will allow it.

In a first experiment testing the MODE model (Sanbonmatsu and Fazio, 1990), participants were given a description of two separate stores and asked to form a general impression of each one of them. In the descriptions they were given, one store was described in a more positive way regarding its single departments, except for one: the camera department. The other store was described as the complete opposite, all the departments were worse than the ones of the former store, but the camera department was better. After having formed a general impression of the two stores, participants were asked to indicate from which store they would buy a camera if they needed one.

If participants decided to choose the generally better store, their decision would have been interpreted as the effect of the more positive attitude evoked by that store. On the other hand, choosing the generally worse store but with a better camera department would have meant putting effort in overcoming the effect of the more positive attitude evoked by the other store by focusing on the attribute of interest. Motivation was enhanced in some participants by telling them that they would have had to compare their choice with the choice of another participant and publicly explain the reason

behind it. Furthermore, opportunity was manipulated by requesting a group of participants to respond quickly, i.e. by putting on them time pressure. The results showed that those participants who were more motivated and had more time to decide were more likely to choose the generally worse store with a better camera department. However, when motivation or opportunity were low, the choice was more likely to be based on the overall characteristics of the store, leading people to choose the one with the worse camera department. This example shows how, in order for motivation to influence attitude evoked responses, people need to have the opportunity to do it. Moreover, when opportunity is low, the behavioural decision people will take will be based on the general impression of the attitude object (the store), regardless of people's motivation to focus on a certain aspect that counters the general impression (the camera department). The same occurs during social interactions. The behaviour towards a member of a different racial group will be much more influenced by our automatic attitudes towards his group when we don't have the time or cognitive resources to counter their effect, regardless of our motivation not to be prejudiced.

3.2.1 Motivation to control for prejudice

In the case of racial prejudice, it is not easy to manipulate participant's motivation, also because the majority of people are already motivated not to be prejudiced. However, it is possible to measure the extent to which they are motivated to avoid being prejudiced, and also distinguish the different components of this motivation. In their Motivation to Control for Prejudiced Reactions scale (MCPR), Dunton & Fazio (1997) consider two factors to be important in driving people's motivation for being unprejudiced. The first one is the "Concern for Acting Prejudiced", which refers to the degree to which one tries to appear unprejudiced either to oneself or to others, and its subscale includes items like "In today's society is important that one not be perceived as prejudiced". The second factor has been termed "Restraint to Avoid Dispute", measuring the desire to avoid conflict with or regarding Blacks. The Restraint subscale containing items like "If I were participating in a class discussion and a Black student expressed an opinion with which I disagreed, I would be hesitant to express my own viewpoint".

The MCPR scale was used to predict people's self-reported attitudes on the Modern Racism Scale from their score on the racial evaluative priming paradigm. As expected, according to the MODE model, individuals that were highly motivated had an overall lower score on the explicit measure (indicating more positivity towards Blacks) when compared to those with low motivation. Furthermore, a significant correlation between the explicit and implicit measures, measured with the MRS and the evaluative priming respectively, was found in individuals with low motivation. Hence, those with low motivation to be unprejudiced that displayed a stronger implicit prejudice towards Blacks were more likely to also report a greater prejudice on the self-report measure.

When looking at the two components of the MCPR, they seem to relate to different behaviours. Dunton and Fazio (1997) found that individuals that scored high on Restraint corrected their attitudes in two ways. Those who had more negative automatic attitudes toward Blacks, were more likely to correct it by attributing more positive traits to Blacks. On the other hand, individuals with more positive automatic attitudes toward Blacks also made corrections, but in this case by attributing more positive traits to Whites.

The results are different when Concern is involved. Individuals with higher Concern that display a stronger implicit prejudice on the evaluative priming, are more likely to correct their response by making them appear having more positive attitudes towards Blacks. However, when automatic attitudes towards Blacks are already positive, fewer corrections in a pro-White direction are made (Fazio & Olson, 2014).

It is therefore possible to interpret Restraint as having a bi-directional effect, acting in opposition to automatic attitudes towards Blacks regardless of their valence, being more positive or more negative. On the other hand, Concern acts more in a pro-Black direction when automatic attitudes are negative, without having a strong impact when automatic attitudes are already positive.

3.2.2 Internal and external motivational factors

It is possible to decompose motivation to control for prejudice in different ways, other than dividing it in Concern and Restraint. When introducing the idea of Concern, I described it as the extent to which one tries to appear unprejudiced towards oneself and the others. However, wanting to appear unprejudiced because of one's own personal

values can have a different effect when this motivation comes from external factors. In fact, one may not want to appear prejudiced because it is more concerned about the way in which people may judge him, rather than holding unprejudiced beliefs. In such case, people's self-reported attitudes will be more negative in the absence of social pressure, but no changes will occur if the individual is motivated to be unprejudiced because of his beliefs about prejudice.

Following this reasoning, Plant and Devine (1998) developed the Internal and External Motivation to Control for Prejudice scale. Their scale was also divided into two subscales, this time differentiating between people's Internal motivation to respond without prejudice (IMS) and their External motivation to respond without prejudice (EMS).

The scale's predictive validity was tested by putting subjects into two different self-reporting conditions. Half of the participants had to report their endorsement of stereotypic beliefs toward Blacks individually and anonymously. The other half had to report their level of endorsement orally in front of an experimenter. As expected, the results showed an overall lower level of prejudice endorsement in the participants scoring high on the IMS as compared to their weakly motivated counterparts. This low endorsement did not make their reports differ between the two reporting conditions. A significant difference in reports between the two conditions was found in those who were low on Internal motivation but high on External, showing a lower endorsement of stereotypes in the public condition compared to the anonymous one.

3.3 The iterative-reprocessing model

Unlike other dual-process models of attitudes, the iterative-reprocessing model (IR; Cunningham & Zelazo, 2007) proposes a different way of interpreting the relationship between implicit and explicit evaluations, by taking a more neuroscientific approach. According to the IR model, attitudes are processed by a single system, and the path from implicit to explicit attitudes is described as a continuum. Automatic attitudes are always activated in the evaluative process, however, they are quickly reprocessed and transformed in what we would define as explicit attitudes, which are not separate from the implicit ones, but are directly formed from them.

The automaticity of evaluative processing was found in a study in which participants were asked perform an evaluative task and then a non-evaluative task (Cunningham, Raye, & Johnson, 2004). In the evaluative task, participants were required to evaluate the valence of certain stimuli (good vs bad), while in the non-evaluative one they were asked to categorize stimuli on the base of their concreteness (abstract vs concrete). The fMRI scans revealed that the two brain areas involved in the processing of positive and negative stimuli respectively, were active both in the evaluative condition and in the non-evaluative one.

The automaticity of evaluative processes was further tested in the domain of prejudice. During a neuroimaging study (Cunningham et al., 2004), a sample of White participants was shown pictures of White and Black individuals while undergoing functional magnetic resonance imaging (fMRI). The fMRI measures the blood-oxygen-level-dependent (BOLD) signal, which indicates increases or decreases in the activation of specific brain areas based on the detected consumption of oxygen. A further manipulation done to test the automaticity of the evaluative process involved the duration of the presented faces. Some pictures were presented for only 30 milliseconds, with the purpose to make participants unaware of the face while recording their brain activity. The other face pictures were presented for 525 ms, enough to make it possible for participants to see them clearly. After the scanning, all of the participants completed an Implicit Association Test. Furthermore, explicit attitudes were assessed using the classic MRS and, in order to test the effect of people's motivations to control for prejudice, the Internal and External Motivation to Respond Without Prejudice scale was also used.

Although all participants reported low levels of prejudice and high internal motivation to respond without prejudice, on average they showed an automatic negative association to Black faces as compared to the White ones on the IAT, indicating a greater implicit prejudice.

The fMRI results showed an increased activation of the amygdala on Black faces compared to the White ones when the presentation was brief (30 ms). This activation was significantly correlated with their scores on the IAT. As the amygdala is a brain region involved in emotions, usually showing a greater activation related to the perception of negative stimuli rather than positive ones, it plays also a central role in

evaluative processing. In this case, the greater activation of the amygdala was related to the greater negative association evoked by the pictures that was also found by the IAT.

However, when the presentation of the pictures was longer (525 ms), the activation of the amygdala was significantly lower, showing no significant difference between black and white faces. In case of the longer exposition to the pictures, an increased activity was found in the anterior cingulate cortex (ACC) and the lateral prefrontal cortex (PFC) for black faces compared to the white ones. These two areas are usually associated with regulatory processes and executive functions, with the specific involvement of the PFC in ambivalent evaluations. The differences between the short and the long picture durations indicate that the automatic reactions of the amygdala can be regulated, and even overcome, by control processes put in motion by other cortical areas. Interestingly, the activation of the dorsolateral PFC predicted the decrease in amygdala activation between the short and long conditions, thus providing evidence of the interaction between limbic and cortical regions during the evaluative process. Lastly, the degree of discrepancy between the implicit measure (IAT) and the explicit one (MRS) was also correlated with the activation of the PFC (in this case the ventrolateral portion).

From this evidence, the IR model breaks down the evaluative process in a more neuropsychological way. When seeing a stimulus, the input signal is sent through both direct and indirect pathways to the amygdala. From here the automatic response is generated, which represents the automatic attitude. The information also arrives to other cortical regions, either already processed from the amygdala, or through the thalamus. These cortical regions interpret the stimulus and the reaction of the amygdala in the light of relevant contextual information, like for example the motivation to appear unprejudiced. These areas would then regulate the backgrounding and foregrounding of what is considered to be relevant or irrelevant information. The reprocessed information is then sent back to the amygdala, regulating its initial response (see Figure 4). This process of information and response reprocessing is believed to occur in cycles or iterations, with each cycle gradually regulating and correcting the response of the amygdala. From studies involving single-cell recordings and depth electrodes, it is supposed that the iterative reprocessing loops occur with a frequency of 5 Hz (one cycle every ~200 ms). This interpretation would explain the increased activation of the ACC

and the PFC and the deactivation of the amygdala that the previous study found. However, the involvement of cortical areas does not always result in a decrease in amygdala activation, as the amygdala response can be both reduced and also enhanced. Such increased activation is indeed observed when participants are shown pictures of people belonging to stigmatized groups towards which a negative response may be seen as socially acceptable (Krendl et al., 2006).

Figure 4. Scheme of the evaluative process according to the iterative-reprocessing model.

As in the IR model the evaluative process is represented as a continuum, the question remains about the impact of awareness on the outcome of the evaluation. Both the evaluative process and the iterative reprocessing were found to operate automatically and outside of people's awareness. Since iterative reprocessing can be seen as the counterpart of the APE model's resolution of inconsistencies, it would be plausible to claim that propositional processes do not need awareness in order to be activated. From this perspective, awareness does not look to be essential in the evaluative process and in distinguishing between implicit and more explicit evaluations. But the IR model argues that, while the initial iterations in the evaluative process may be automatic and unconscious, evaluations that are based on an increasing amount of iterations should be more reflective and conscious. Nevertheless, the amount of information that it is possible to elaborate without the need of awareness, i.e. amount of iterative cycles that can occur unconsciously, is still a topic of research.

4. Conscious and unconscious processing

The awareness of a stimulus may not always be necessary for it to have an effect on our judgments and behaviours. I first mentioned the idea of processing stimuli outside awareness in relation to the evaluative priming paradigm with subliminally presented prime images. Finding the same facilitation effects of Black and White faces also when they were presented subliminally, suggests that at least associative processes do not need conscious awareness of the stimuli in order to be activated (Dovidio et al. 1997). On the other hand, if we wanted to put in motion propositional processes, we may need an increased amount of effort that would result in a greater elaboration of the stimuli, an elaboration that is usually thought to require conscious awareness. But before being able to discuss the involvement of conscious awareness in processing stimuli, it is necessary to define consciousness and how it emerges.

Starting from neuropsychological findings on visual awareness, it has been theorized that consciousness emerges as a result of a widespread recurrent neural processing (Lamme, 2006). This means that, in order to perceive a stimulus consciously, not only do visual areas need to be active and interacting with one-another, but the neural activation has to extend to frontoparietal areas of the brain. Only in this case will a visual stimulus be reportable.

In a study on inattention blindness, subjects were shown a stream of black and white letters on a background made of stripes oriented all in the same direction. Participants were asked to press a button whenever a white letter was presented. However, on half of the trials with white letters, nine squares made of the same stripes of the background, but oriented in a different direction appeared. The squares, being characterized only by the different orientation of the stripes, were not noticed by half of the participants, as their attention was taken away by the task. While performing the trials, their brain activity was also recorded. In those participants that did not report perceiving the squares, their specific neural activation during those trials was recurrent, but also limited to the visual areas of the brain. However, when the activation was more widespread, and not limited to visual areas, participants reported being able to perceive the squares (Scholte et al., 2006).

On the base of conscious accessibility and processing of the stimuli, there are several differences between inattention blindness and classic priming. In priming,

given the brief prime duration, the primes can't be consciously perceived even when looking at them. From a neural perspective, the signal may be too weak to generate any recurrent processing within and outside visual areas. However, during inattention blindness, the stimuli are strong enough to be perceived, but the subjects' attention is drawn away from them. Nevertheless, when people focus their attention on them, they can report seeing them. Therefore, stimulus strength can play a role in making a stimulus become conscious, but it will be conscious only when attention will be allocated on it.

It has been proposed that the processing of a visual stimulus can be divided into three categories: subliminal, preconscious and conscious (Dehaene et al. 2006). What distinguishes these states are the stimulus' strength and the person's attention to it. *Subliminal processing* arises when a stimulus is weak, but there are some differences in processing it based on the allocation of attention. If attention is occupied elsewhere, brain activation will be weak, no recurrent processing will be engaged, and, even though there may be some priming, the signal will quickly decay and there will be no reportability. If, on the other hand, attention is present, there will still be no recurrent processing (accompanied by the lack of reportability) but the signal will be strong enough to enable semantic priming. When the stimulus is strong but attention is absent, we have *preconscious processing*. This would be the case of inattention blindness. In preconscious processing, recurrent processing within occipito-temporal regions of the brain will be engaged, but it will be confined to those areas. The priming will occur on multiple levels, and, although the conscious experience of the stimulus will not be reported, it will be possible to do it once attention is directed towards the stimulus.

It is possible to argue that preconscious processing stays between subliminal processing and conscious processing: the stimulus is still subliminal, but can eventually become conscious when it is given attention. It is only during *conscious processing*, when both the stimulus is strong and the person attends it, that the neural activation and recurrent processing extend to fronto-parietal areas, making the person aware of the stimulus. The activation can in this case be maintained at will, until attention is shifted to another stimulus. However, this other stimulus to be attended could have already been processed by preconscious processing. In the inattention blindness example, subjects were consciously processing the letter that appeared in the middle of the screen,

but some of them managed to shift their attention to the surrounding squares, which their preconscious was already processing. Nevertheless, the degree of stimulus complexity that can be elaborated and the amount of its elaboration through unconscious processing, is still a topic of research which I also want to explore.

As associative processes are often regarded as not being capable of more elaborated thought, this limitation has also been attributed to the elaboration of a stimulus outside one's awareness, that is, unconscious processing. For this reason, unconscious processing was thought to be related exclusively to associations eliciting automatic reactions, without the capacity of analyzing more complex information. This idea was also validated by previously discussed studies showing the absence of the elaboration of negation while subliminally presented with, for example, "no smoking" cues (Earp et al., 2013).

Other studies also concluded the inability for unconscious processes to elaborate multiple-word expressions presented subliminally (Greenwald & Liu, 1985). For example, it was found that the subliminal presentation of the expression "enemy fails", which may be considered as a positive sentence, would prime participants according to the valence of the single negative words. Nevertheless, as previously argued, the brief presentation of the multiple-word expressions may not have served as a fair comparison between conscious and unconscious processes. However, a potential drawback of this study may have been the presentation of single negative words making up positive expressions. In a study related to unconscious processing, participants were told that on each trial some words would appear on the screen, followed by a string of Xs, and their task was to say whether they thought that the subliminally presented words were positive or negative (Dijksterhuis & Aarts, 2003). It was found that negative words presented subliminally were more often correctly categorized than positive words, and even though no participant reported seeing the words, the categorization was above chance. This lead to conclude two things: the first is that the valence of a stimulus is elaborated before other aspects, which would be in agreement with the IR model previously described, and the second that negative stimuli need a lesser input strength to be elaborated when compared to positive ones. For this reason, the lack of positive priming with the subliminal presentation of the expression "enemy fails", may be accounted not to the lack of unconscious elaborative capacity, but rather to the easier

elaboration of negative words combined with the overall briefness of the stimuli. Therefore, even though the multiple-word expression may have had a positive valence, there may have not been enough time to process it.

Some researchers came to question the attribution of exclusively associative processes to unconscious elaboration (see, Bargh & Morsella, 2008). It has been suggested that conscious processing may be even more limited than the unconscious one. In fact, conscious processing cannot elaborate subliminal primes at all. Furthermore, it may be unfair to compare conscious and unconscious processing on the basis of primes of short duration. For this reason, the exploration of unconscious processing may be limited by the briefness of the stimuli used so far to study it. Because of their briefness, subliminal primes are far weaker than supraliminal primes, and the signal may not be processed as it would if the presentation lasted as long as with supraliminal priming.

In order to test the elaborative capability of the unconscious, it is at least necessary to give it enough time to process the primes. However, there is the problem of presenting stimuli for a longer time without subjects being aware of them. A potential solution comes from research on perception involving binocular rivalry and flash suppression. Binocular rivalry emerges when two pictures are presented one on each eye. As a result, what people perceive is an alternation between the two images (Blake & Logothetis, 2002). This alternation, whose frequency is unpredictable, makes it hard for binocular rivalry to be used as a priming technique on its own. A variant of binocular rivalry is the flash suppression, in which one image is presented to one eye first, and then another image is flashed to the other eye while the first is still present. For example, the left eye may initially be shown a face while the right one is shown a blank screen. Subsequently, the right eye is also shown a picture, for example, of a car. This sudden presentation makes people perceive only the second image, while the first one, still present, is suppressed from consciousness for a short amount of time (Wolfe, 1984). Nevertheless, while one image can be made invisible for a short period following the appearance of the second one, it is still initially consciously perceived. None of the two techniques on their own are optimal to study unconscious processing and suppress images from awareness for a longer duration. However, this result is achievable when they are combined.

A new method of perceptual suppression resulting from the combination of binocular rivalry and flash suppression was introduced by Tsuchiya & Koch (2005), and was named continuous flash suppression (CFS). In CFS, by using a set of mirrors, one eye is presented with high-contrast coloured patterns (often described as Mondrian-like¹) that change at a rate of approximately 10 Hz (once every 100 ms). At the same time, on the other eye a low contrast image is presented. This image remains suppressed from awareness while attention is occupied by the flashing patterns. In Tsuchiya and Koch's study, the low contrast image (a Gabor patch) was successfully kept outside awareness for a mean of 56 seconds, lasting even up to 3 minutes. The application of CFS has made it possible to further explore the capacities of unconscious elaboration, without the drawback of having a too brief, and therefore weak, stimulus presented subliminally.

The unconscious elaboration of multiple-word expressions that previously failed to show the expected priming results was tested again using CFS. In one study, participants were shown the Mondrian patterns one eye, while on the other one a multiple-word expression whose contrast increased with time was shown. In the middle of both images there was a fixation cross (perceived as a single one) and participant's task was to say, as soon as they started perceiving it, whether a string of letters (the expression) appeared above or below the fixation cross (Sklar et al., 2012). In this case, the expressions that were either positive or negative, were made of relatively neutral words. An example can be the expression "black eye", made of two neutral words that together convey a negative affective meaning. As found in the study with single subliminal negative and positive words, negative expressions were perceived significantly sooner than the positive ones. These findings finally showed the potential of unconscious elaboration to go beyond the processing of single words, extending its capacities to multiple-word expressions that require a more abstract level of analysis. In the same study, people that were subliminally primed with three-digits mathematical equations (e.g., $9 - 4 - 3 =$) by using CFS-masking. After the priming, lasting at least 1.000 ms, participants were presented a number that they had to pronounce. Results showed that when the target numbers were the solution to the subliminally presented

¹ Pieter Cornelis "Piet" Mondrian was a Dutch painter, exponent of the "De Stijl" art movement, also known as neoplasticism.

equation, participants were faster at pronouncing them. These results contributed expanding once the capabilities attributed to unconscious processing. The capabilities were this time related to sequential manipulations of abstract symbols - numbers - according to acquired rules - mathematical operators.

The findings about unconscious processing using CFS-masking started to put into question the unique attribution of more complex elaborative capacity to conscious processing, extending the research to a more social domain. A study combining CFS-masking and fMRI found that, when compared to pictures of houses, fearful faces presented for 2.5 seconds elicited both an increased amygdala response and an increased activation of the parietal cortex (Troiani, Price, & Schultz, 2012). The latter region, associated with processes related to directing attention, has never before been found to be active after the presentation of subliminal social stimuli, indicating that these stimuli can indeed reach cortical areas without accessing awareness.

The research described so far shows how it is possible to successfully unconsciously process several classes of stimuli, ranging from multiple-word expressions, to mathematical equations, and emotional faces. Nevertheless, there was no study testing the unconscious evaluative processing in the domain of racial prejudice.

5. Unconscious control of prejudiced reactions

The IR model theorizes that the response of the amygdala induced by a Black face on a White participant with high implicit prejudice, is overridden according to his motivation not to be prejudiced by the automatic engagement of frontal cortical areas (Cunningham & Zelazo, 2007). However, being the activation of similar cortical areas also associated with conscious processing (Lamme, 2003; Dehaene et al. 2006), it puts into question the extent of the automaticity of iterative reprocessing when awareness of the stimulus is absent. The activation of parietal cortical areas found after fearful faces (Troiani, Price, & Schultz, 2012), suggests that a subliminally presented face may be elaborated beyond the initial affective response. However, to the best of my knowledge, there has not been any research so far trying to find out whether unconscious processing can activate, apart from an initial affective response, also regulatory processes that are capable of influencing an evaluative response according to factors normally associated with awareness. The question can be put in these terms: if a highly implicitly prejudiced individual, who is also explicitly highly motivated not to be prejudiced, is presented with the face of a Black person, will his initially negative evaluative response become more positive with time even if he is not aware of seeing the face?

6. The present study

According to the findings of previous research related to implicit prejudice with subliminal priming, the effects of the subliminal presentation of the face of an individual belonging to a different racial group are similar to the ones found when, on the same paradigm, the presentation is supraliminal (e.g., Dovidio et al. 1997; Olson & Fazio, 2002). This indicates that a weak stimulus is sufficient to elicit a certain reaction. However, unlike when answering a questionnaire, there is not much time to elaborate a subliminal stimulus. According to the IR model, the briefness of such prime does not allow for a more complex elaboration and an increased number of evaluative cycles before the stimulus decays. Nevertheless, a longer stimulus so far would put in motion an increased amount of iterations, but only when awareness of the stimulus was also present. However, it is yet unknown whether longer iterative reprocessing would also occur and influence the evaluative process in the absence of awareness of a stimulus, or

if awareness is really necessary to engage into iterative reprocessing that would lead to a more positive evaluation.

Researchers found that the unconscious elaboration of stereotype-related information leads to a different outcome than the conscious one. More precisely, they claim that when people engage in unconscious thinking (compared to when they engage into conscious thought), they tend to rely less on the stereotypical information given about an individual when having to make a judgment. This phenomenon was found in a person memory paradigm (Bos & Dijksterhuis, 2011). Participants were given stereotype-congruent and stereotype-incongruent information of an individual initially described as belonging to a different racial group. The two groups were then asked to report their impressions of the individual and to recollect the information describing it. But before giving the judgment, one group of participants had some time to think about the impression that individual left them, another group was distracted by an unrelated word search puzzle. The group that was distracted by another task, and, according to the authors, engaging into unconscious thought, judged the person in a less stereotypical manner and remembered more stereotypical-incongruent information than the other conscious-thought group. It may be that any influence of an initial unwanted automatic attitude towards the described individual was better overridden while attention was occupied elsewhere. However, it is hard to make this claim as participants' motivation to avoid being prejudiced was not considered. Additionally, as all the participants were aware of the target individual to evaluate, we cannot exclude that the reprocessing that may have occurred in the unconscious thought group has been triggered consciously.

If the engagement into a greater iterative reprocessing could occur without the person being aware of the stimulus eliciting the automatic reaction, motivational factors may have the opportunity to influence the evaluative process like the MODE model describes. I therefore hypothesize that by prolonging the duration of a subliminal stimulus, motivational factors will influence its evaluation thanks to the automaticity of iterative reprocessing. If a participant is motivated not to be prejudiced, any initial negative response mediated by the amygdala after seeing a Black individual, should get reprocessed through the engagement of an increasing amount of brain areas, among which the prefrontal cortex. This process should therefore reduce the negativity of the evaluation the more time the brain has to reprocess it. These processes are not displayed

when employing short subliminal stimuli. With brief subliminal presentations, the response to the stimulus may decay by the time it starts to be reprocessed. To avoid such problem, the continuous flash suppression paradigm may become useful.

As I am interested in studying the dynamics of evaluative processing, I need a measure that allows to (a) measure the evaluative response elicited by certain primes, and (b) explore the temporal dynamics of such response. A study related to the domain of self-regulation of eating behaviour suggested the paradigm to use for such purpose. The aim of the study was to test the temporal dynamics of the evaluative processing of highly palatable foods when participants were motivated to avoid them (Hofmann et al., 2010). In the study, participants were asked to complete an AMP containing as primes, among other control stimuli, highly palatable food pictures. They were then split into normal eaters and dieters according to their scores on a questionnaire related to eating restrictions, indicating the degree of their motivation to avoid highly palatable foods. Two variables were manipulated in the AMP, one between subjects and one within. As the within subjects variable, the interval between the prime and the target – the inter-stimulus-interval (ISI) - was manipulated. On some trials, the ISI was brief (100 ms), while on others it was long (1000 ms). Regarding the between subjects variable, while a group of participants directly completed the AMP, another group completed a lexical decision task² before the AMP. This lexical decision task contained words related to highly palatable foods, and it served as a preexposure to food condition. The dependent variable was the percentage of pleasant responses to food stimuli, compared between normal eaters and dieters.

Results showed that, without the pre-exposure, dieters tended to have a significantly reduced overall percentage of pleasant responses to foods when compared to normal eaters in the brief ISI condition (Hofmann et al., 2010). Even though in dieters the proportion of pleasant responses was lower after longer ISI, the difference was not significant. However, after the preexposure, while the proportion of pleasant responses to food primes was almost equal in the brief-interval trials, the proportion of pleasant responses to long ISI trials in dieters increased to the point of becoming even greater than the one of normal eaters.

² In a lexical decision task, participants are asked to categorize strings of letters by saying whether they are words or non-words.

The devaluation found in the non-preexposure condition was attributed to the effect of the dieting goal, while the sudden switch to an increased amount of pleasant responses was attributed to the inhibition of the dieting goal caused by the preexposure, which activated the goal of enjoyment of foods (Hofmann et al., 2010). In the latter case, as described by the IR model, the initial response of the amygdala to highly palatable foods was equal in both dieters and normal eaters. It is only at longer intervals that the reprocessing included into the evaluation other motivational factors. As in dieters the motivation to approach foods was boosted by the preexposure condition, it influenced the evaluative (re)processing in a pro-foods direction. For the purposes of the present research, these results showed the capability for the AMP to successfully capture the temporal dynamics of evaluative processes.

It would be possible to apply a similar paradigm to the domain of racial prejudice, but there are some important things to consider first. In the previous study, the exploration of the temporal dynamics of evaluative processing was done by manipulating the interval between the prime and the target. As found by Payne et al. while studying the AMP, by increasing the interval between the prime and the target, the initial response can gradually start to decay.

Additionally, the significant differences between the two time conditions after the preexposure can be attributed to a preservation of the signal thanks to the strong motivation to approach it because of the hedonic pleasure associated with such objects, however, such motivation and the associated reward are not present in the domain of racial attitudes.

For the abovementioned reasons, rather than manipulating the ISI in the AMP, the manipulation of the duration of the primes may be better strategy to adopt in order to study the temporal dynamics of evaluative processes in the domain of racial prejudice. By preventing the decay of the stimulus, it may be possible to better capture the expected change in the evaluative process across time according to people's motivation.

7. Pilot study

In order to prepare for the main experiment, the aim of the pilot was to replicate the findings of Payne and colleagues (2005) by testing the race AMP in Nijmegen. As done by Payne et al., to assess the moderation of motivation on the relationship between implicit and an explicit prejudice, I included the Internal and external motivation to control for prejudice scale (Plant & Devine, 1998) assessing participant's motivation not to be prejudiced, and a feeling thermometer as an explicit measure of attitudes towards Blacks and Whites.

I expected to find participant's motivation to moderate the relationship between the explicit and the implicit measure. This means that, while highly motivated individuals would not display any correlation between the feeling thermometer and the AMP, the correlation would be significant when motivation would be low. This correlation would indicate that the attitudes found on the implicit measure would have also been expressed explicitly. Therefore, individuals with more negative implicit attitudes towards Blacks would also report more negative attitudes on the feeling thermometer, and individuals with more positive attitudes towards Blacks would also report more favourable feelings towards Blacks.

7.1 Method

7.1.1 Participants

A total of forty-six students (33 females, $M_{\text{age}} = 22.4$) from the Radboud University completed the pilot. They all received 5€ for their participation both to this study and a second unrelated one that was always done afterwards. As one participant claimed that she was able to understand what part of the Chinese characters meant, her data were excluded from the final analysis, leaving 45 participants (32 females, $M_{\text{age}} = 21.8$).

7.1.2 Design

The independent variable, consisting of three types of prime (Black, neutral, White) was manipulated within participants. The dependent variable was the proportion of “pleasant” judgments given to the pictographs in each prime condition of the AMP.

7.1.3 Materials and procedure

After being welcomed by the experimenter (me), participants were lead to the cubicles and verbally introduced to the task. They were told that they would participate to a “symbol judgment study” that examined “how people make simple and quick judgments” (cit., Payne et al., 2005, p. 280). They were also told that on each trial pairs of pictures would have appeared on the screen, the first one being a face, the second being a Chinese character. Their task was to ignore the face and quickly judge the character based on its visual pleasantness by using two response keys on the keyboard. If the character was less visually pleasing than an average Chinese character, they had to press the key labelled “unpleasant”, while if it was more pleasing than average, they had to press the key labelled “pleasant”.

Pictures of emotionally neutral faces of white and Black males from the Chicago Face Database (Ma, Correll, & Wittenbrink, in press) were selected as primes. All the faces from the database were previously rated on several dimensions. The ones included in the pilot were chosen based on their representativeness of their racial group (the faces had over 90% of respondents judging them as Blacks and Whites) in order to reduce any possible ambiguity. Furthermore, the chosen faces were all equated on the basis of their attractiveness and perceived age, which had to be the closest possible to their mean. The mean attractiveness of the chosen Black and White faces ($M_s = 3.53$ and 3.56 , respectively) was not significantly different between the two groups, $F(1,23) = .02$, *ns*. The same goes for the perceived ages of Black and White faces ($M_s = 26.63$ and 27.17 , respectively), $F(1,23) = .07$, *ns*. The final pool consisted of 12 White and 12 Black faces. These 24 pictures were then edited by cutting out the visible parts of their gray t-shirts, leaving only the head and part of the neck. Additionally, the picture of a gray square was created, serving as a neutral control prime.

Each trial began with a fixation cross in the middle of the screen (500 ms), followed by one of the primes (75 ms), a blank screen (125 ms), the target pictograph (100 ms) and finally a mask of black and white dots. The mask lasted until participants gave their response. The prime was randomly selected and paired with a unique Chinese character on each trial, for a total of 72 trials.

The target Chinese pictographs were taken from the original set used by Payne et al. (2005). During a pre-piloting session with 8 participants, I found a negative

correlation between the “pleasant” judgments of the Chinese characters and the number of lines they consisted of. Chinese characters were regarded as being more visually pleasing the less number of lines they were made of. Since this correlation was much weaker when considering only the characters consisting to 7 or more lines, I decided not to include characters with 6 lines and less. The idea behind this choice was that having overall more complex characters would have increased their ambiguity, leading to an easier misattribution of the reaction elicited by the primes.

After the AMP, participants were asked to complete a questionnaire consisting of the 10 items of the Internal and external motivation to control for prejudice scale (Plant & Devine, 1998) and the feeling thermometer. For the latter, as done by Payne et al. (2005), participants were asked to express their feelings towards Blacks and Whites³ on a 9-point scale, anchored by the labels 1 (“cold and unfavourable”) and 9 (“warm and favourable”). All together, the task lasted approximately 7 minutes.

7.2 Results

To analyze the effect of the primes on the judgments of the characters, I conducted a single factor (prime: Black, neutral, White) repeated measures ANOVA. The results showed no significant effect of the primes, $F(2, 88) = 2.18, p = .12$. Nevertheless, as found by Payne et al., the effect was strongly driven by the fact that participants liked much more the characters that appeared after neutral primes when compared to those that were preceded by faces - the effect got closer to significance as a quadratic trend, $F(1, 44) = 3.24, p = .08$, but not as a linear trend, $F(1, 44) = .74, ns$. When the neutral primes were removed from the analysis because of less theoretical interest, the results were far from significant, $F(1, 44) = .74, ns$.

Even though on the feeling thermometer participants reported more favourable feelings towards Whites than towards Blacks ($M_s = 6.64$ and 6.60 respectively), the difference was not significant, $F(1, 44) = .04, ns$.

Before testing the relationship between the score on the explicit and the implicit measure and the possible moderation of motivation, each participant’s AMP score was calculated by subtracting the proportion of pleasant responses after White faces from the

³ Participants were also asked to report their feelings towards Asians, Moroccans and Hispanics, but these groups were included only as a filler and were not considered during the analysis.

proportion of pleasant responses after Black faces. In this way, a higher AMP score would indicate a greater positivity towards Blacks. The same was done for the explicit measure by using the judgments for Whites and Blacks on the feeling thermometer, thus creating a Self-report score.

There was a significant positive correlation between the Self-report and the AMP score ($r = .53, p < .01$). To test whether there was a relationship between the type of trials and the feelings toward the two groups, partial correlations were made. Therefore, responses to neutral and Black trials were controlled when White trials were analyzed; when Black trials were analyzed, the trials controlled were the White and the neutral ones. The results showed that positivity on Black trials was marginally significantly positively correlated to positive feelings toward Blacks ($r = .29, p = .06$), but not significantly correlated to feelings toward Whites, even though the correlation went into a negative direction ($r = -.21, ns$). On the other hand, positivity on White trials was positively correlated to feelings toward Whites, but this partial correlation did not reach significance ($r = .25, p = .10$). Finally, positivity on White trials was significantly negatively correlated to feelings toward Blacks ($r = -.45, p < .01$).

Regarding the motivational factor, I expected the participants that were less motivated to control for prejudice to show a greater correlation between their AMP scores and Self-reports. I initially checked whether there was a correlation between the AMP and participant's motivation on the two scales, as the AMP score should not be influenced by motivational factors. As expected the correlation between the AMP and the internal and external motivation subscales was not significant (respectively, $r = .17, ns$, and, $r = -.14, ns$).

The moderation of motivation on the relationship between the AMP and the Self-report was tested via multiple regression. The dependent variable was always the Self-report, while the predictors were the AMP score and the two motivational scales. The two predictors were entered in the model first, followed by their interaction. The results showed that both the interaction between the AMP score and Internal motivation was not significant, $\beta = .79, t(41) = .62, ns$, as well as the interaction between the AMP and External motivation, $\beta = -.78, t(41) = -1.59, ns$. However, the casewise diagnostics revealed two potential outliers with standardized residuals of 3.65 and 2.53. When excluding these two cases, as can be seen from Table 1, the internal motivation effect

became significant, $\beta = .31$, $t(39) = 2.81$, $p < .01$, but the internal motivation x AMP and external motivation x AMP interactions did not, $\beta = 1.48$, $t(39) = 1.55$, *ns*.

Predictors	B	SE	<i>t</i> (39)	<i>p</i> <
Internal motivation	.31	.11	2.81	.01
AMP	-1.07	.97	-1.11	.28
Internal x AMP	1.48	.95	1.55	.13
External motivation	-.24	.11	-2.12	.04
AMP	.92	.40	2.31	.03
External x AMP	-.50	.4	-1.24	.22

Table 1. Results of the regression predicting Self-report racial attitudes from the AMP score and the two motivation scales.

7.3 Discussion

Even with (9) more participants and twice as many trials than Payne and colleagues in their original study, the AMP was not able to detect any significant differences in racial evaluations. This could be due to the fact that in my sample (taken from the Nijmegen student population), there may be a much lower degree of racial bias when compared to the sample recruited in the US by Payne et al. For this reason, with a greater sample it may be possible to find a significant difference, as the prime effect may be too weak in order to be detected with 45 participants.

Regarding the motivational factors, even though External motivation was not significant, the Internal motivation was significantly correlated with self-reported attitudes, indicating that the more people were internally motivated, the more they reported positive feelings towards Blacks. However, unlike expected, Internal and External motivation did not significantly moderate the relationship between the Self-report and the AMP. The expectation was to find a different correlation between Self-report scores and AMP scores when comparing individuals with low and high Internal motivation. More specifically, only the AMP scores of the individuals with low motivation were expected to significantly correlate with their Self-reports, but this correlation was expected not to be significant in those highly motivated.

When debriefing the participants, I noted that many reported that the Chinese characters were often too brief to be perceived appropriately, and that they therefore had to respond “randomly”. It may be that the non significant effect of the primes was due to the impossibility to see the target and therefore misattribute the reaction elicited by the primes. In a recent paper, Payne & Lundberg (2014) discussed this issue, suggesting that, even though students may not have problems with very brief stimuli, when testing on the general population, with older, less familiar with computerized tasks or less educated participants, the duration of the targets can be extended up to 300 ms without influencing the priming effects.

8. Main study

The main study had a similar structure to the pilot, and consisted of an updated version of the race-AMP, the Internal and external motivation to control for prejudice scale (Plant & Devine, 1998) and a feeling thermometer. Based on the results of the pilot study, one of the main changes consisted in increasing the duration of the pictographs from 100 ms to 200 ms. The reason for this change was to make it easier for participants to see the targets in order to better capture the affect misattribution, as suggested by Payne & Lundberg (2014). Since I am interested in studying the dynamics of the process of correction of prejudice, the duration of the primes was manipulated. I expected to find a significant increase in the positivity of the evaluative response towards Blacks between the primes of short and long duration in highly motivated individuals compared to the lowly motivated ones. Participant's awareness of the primes was also manipulated in order to test the automaticity of such process, expecting to find the same difference in the evaluative judgments also when awareness of the primes was absent.

According to the MODE model, I also expected to find a significant interaction between the implicit measure and the amount of motivation when used to predict explicit attitudes, meaning that, a greater implicit negativity towards Blacks would correspond to a more negative explicit attitude when motivation to control for prejudice is low. On the other hand, a greater implicit negativity would be reflected by a greater explicit positivity towards Blacks only in highly motivated individuals.

8.1 Method

8.1.1 *Participants*

A total of eighty-four students (64 females, $M_{\text{age}} = 22.13$) from Radboud University completed the pilot. They all received 5€ for their participation both to this study and a second unrelated one that was always done afterwards. Five participants were excluded from the analysis: three of them were excluded as they reported being able to see some of the suppressed faces, and one participant was excluded because of being capable to understand some of the Chinese characters used in the AMP. The fifth

one was excluded because it did not belong to the category of Whites or Blacks⁴. Their exclusion did not significantly change the results of the analyses. The total amount of participants in the final analysis was 79 (32 females, $M_{\text{age}} = 22.11$).

8.1.2 Design

The experiment consisted in a 3 (Black vs. White vs. neutral prime types) x 2 (short vs. long prime duration) x 2 (supraliminal vs. subliminal prime presentation) design. The prime type and prime duration variables were manipulated within subjects, while the awareness of the primes was manipulated between subjects. The dependent variable was the proportion of “pleasant” judgments given to the pictographs in the AMP.

8.1.3 Materials and procedure

The instructions given to participants in the “Symbol judgment study” were the same as those from the pilot study, with a few additions. Firstly, all subjects were warned about the potential influence of the primes on the evaluation of the pictographs, and they were then asked to try not to be influenced and judge exclusively what the characters evoked. This warning was also tested in one condition of one of the experiments described in the original article introducing the AMP, without having a significant effect on the outcome. The same warning was also used in the same article when testing the effect of Black and White faces, finding a significant effect of the primes (Payne et al., 2005).

Secondly, participants in the subliminal prime condition were initially introduced to the mirror stereoscope. Once they entered the cubicle with the stereoscope, they saw the device and the screen with two identical images. They were told that the stereoscope would have merged the two pictures together by making them become one, and the purpose of this was to make that image extremely clear and easy to see. By the end of the experiment, none of them had any idea about the true purpose of the device.

Before trying the mirrors, their ocular dominance was assessed through the Miles test (Miles, 1930). Participants were asked to form a small circle by bringing their

⁴ The subject was Indian, and displayed a very dark skin tone that may have played a role in associating itself more with Blacks than with Whites.

hands together and extending them. Afterwards, they were asked to look through the opening at a distant object (a sign on a door) with both eyes open. Finally, by alternating closing the eyes, they were asked to report which eye, after closing it, made it impossible to see the object through the hole. That eye is considered to be the dominant one. In order to try whether the mirrors were correctly calibrated, participants were asked to place their chins on a chin rest, look through the mirrors and describe what they saw and its clearness. According to their description, the mirrors were adjusted until the image was described as clear.

The images consisted of a square frame, three crosses (one above, one below and one in the centre) and the words “Press ‘S’ or ‘5’” written inside the square (“Press” was above the centre fixation cross, and “‘S’ or ‘5’” was below it). In both awareness conditions, the frame and the external fixation crosses were displayed on the screen during the whole task in order to achieve a more stable convergence of the of the two images, as done by other researchers (e.g., Anderson et al., 2012). The reason for having the crosses and the frame also in the supraliminal condition was to resemble the subliminal condition as much as possible.

The cubicle containing the mirror stereoscope had two computer screens. One was positioned in front on the stereoscope and served for completing the AMP. The second one was positioned on the side and rotated in such way that it was not possible to see it while the head was on the chin-rest during the AMP. Given the difficulty of displaying the instructions and responding to the questionnaire through the stereoscope, participants were instructed to use the screen in front of the mirrors only during the main task. The setting of the cubicles can be seen in Figure 5.

The prime pictures were the same ones used in the pilot study, but their contrast was reduced in order to obtain a stronger suppression during the CFS in the subliminal condition. The same lowered contrast primes were used also in the supraliminal condition. Sixteen Mondrian patterns for the CFS were also created by randomly placing circles of different colours on the image.

In both conditions, each type of prime was presented for 16 times in each duration condition (16 time for 100 ms, and 16 times for 1500 ms). The prime was

randomly selected and paired with a unique Chinese character⁵ on each trial, for a total of 96 trials. Participants always kept the same distance from the screen in both conditions by placing their head on the chin-rest. The distance was approximately 55 cm. To maintain the same viewing distance and stimulus size in both conditions, participants were asked to place their head on a chin-rest during the judgment task. In this way, each image inside the frame occupied approximately 5x5 degrees of visual angle.

Figure 5. Experimental setup of the cubicles of the supraliminal (left) and of the subliminal (right) priming conditions.

In the subliminal condition, participants read the instructions on the rotated screen, after which they were asked to place the head on the chin-rest as they did during the calibration. In that moment, the picture shown during the calibration was again presented, and they had to press one of the two indicated buttons to start the task ('S' or '5'). The two buttons were the same ones that had to be used to judge the characters. During each trial, both eyes were initially presented with the fixation cross in the centre of the frame (500 ms). Afterwards, the dominant eye was presented with a Mondrian pattern with the same fixation cross in the centre while the prime appeared to the non-dominant eye (100 ms or 1500 ms). While the prime was presented, the dominant one continued viewing the patterns, which changing at a rate of 10 Hz (one every 100 ms; as in, Tsuchiya & Koch, 2005), with the fixation cross always present in the centre. In both eyes, the primes and patterns were followed by an empty frame (100 ms), the target

⁵ Not having found any correlation between participants' responses and the number of lines that the characters were made of in the pilot, Chinese characters with less than 7 lines were included in the main study.

Chinese character (200 ms), and a black and white noise mask until a response was given. A scheme of the succession of the stimuli can be seen in Figure 6.

In the supraliminal condition, each trial began with a fixation cross in the centre of the frame (500 ms), followed by one of the primes (100 ms or 1500 ms), a blank screen (100 ms), the target pictograph (200 ms) and finally the black and white mask until participants gave a response.

After the AMP, participants were asked to complete the internal and external motivation to control for prejudice scale (Plant & Devine, 1998) and the feeling thermometer used in the pilot. The whole task lasted approximately 15 minutes.

Figure 6. Scheme of the temporal succession of the stimuli presented on each eye in the subliminal condition.

8.2 Results

Given the positive skewness of 6.62 ($SE = .03$) in the distribution of response latencies, trials in which response latency was greater than 3 standard deviations from the mean, and lower than 150 ms (as done by, Hofmann et al., 2010), were not included in the analysis (4.3% of the total number of trials). The resulting proportion of pleasant responses to the three types of prime were .54 ($SD = .12$) for Blacks, .54 ($SD = .13$) for the neutral primes, and .52 ($SD = .13$) for the White ones. On the feeling thermometer, participants reported on average more favourable feelings towards Whites ($M = 6.56$,

$SD = 1.39$) than towards Blacks ($M = 6.30$, $SD = 1.37$), and the mean score on Internal motivation (6.89 , $SD = 1.41$) was higher than the one on External motivation. (4.47 , $SD = 1.52$).

I initially analyzed the effect of the primes on the judgments of the characters and the effect of prime duration and awareness by conducting a mixed-design ANOVA, with awareness of the primes as a between-subjects variable. The result would have revealed if there was any difference between the proportion of pleasant responses to different primes, and if this difference changed between the two prime durations and between the two priming conditions (subliminal vs. supraliminal). The results showed no significant main effect of the primes, $F(2, 154) = 2.30$, $p = .10$. Unlike the pilot, the quadratic trend was far from showing a significant effect, $F(1, 77) = .99$, $p = .32$, compared to the linear one, which got closer to significance, $F(1, 77) = 3.49$, $p = .066$. This indicates that the difference between the Black and White primes was greater than the difference between the neutral one and any of the other two types of prime. Unexpectedly, as can be seen in from the graph in Figure 7, the proportion of pleasant responses was marginally significantly higher after Black compared to White primes.

Figure 7. Proportion of "pleasant" responses in the AMP after Black, White, and neutral primes. Error bars indicate one standard error.

The between-subjects condition was also significant, $F(1,77) = 5.44$, $p = 0.022$. Participants reported less pleasant responses to the pictographs in the subliminal primes condition. All the other effects and interactions were not significant: prime type x

awareness, $F(2, 154) = .21$, *ns*, prime duration, $F(1, 77) = .47$, *ns*, prime duration x awareness, $F(1, 77) = .00$, *ns*, prime type x prime duration, $F(2, 154) = 1.03$, *ns*.

The next preliminary analysis consisted in testing whether participants were able to control their responses according to their Internal and External motivations to control for prejudice. For this purpose, the AMP score was calculated by subtracting the proportion of pleasant responses after White faces from the proportion of pleasant responses after Black faces. A higher AMP score would indicate a greater positive reaction to Black compared to White primes. This score was then correlated with the two motivation subscales scores. The results of this analysis showed no correlation between the AMP and Internal motivation ($r = -.02$, $p = .83$), nor between AMP and External motivation ($r = -.04$, $p = .76$), indicating that indeed participants were not controlling their responses on the AMP, at least not according to their motivations.

For the main analysis, in order to test the change in the evaluation with time according to motivational factors in the two awareness conditions, I ran a mixed-design ANCOVA with prime (Black vs. White) and prime duration (short vs. long) as within-subjects factors, prime awareness (supraliminal vs. subliminal) as a between-subjects factor, and Internal and External motivation as continuous predictors. The analysis showed only a marginally significant effect of duration, $F(1,71) = 3.119$, $p = .082$, all the other effects or interactions were not significant (all $F_s < 1$).

Not having found any of the hypothesized interactions, I decided to further explore the data to see if any pattern in the expected direction emerged. I wanted to check whether motivation had the expected effect in increasing the proportion of pleasant responses from short to long Black prime trials, without having any effect on the White primes trials. I conducted the same analysis as the main one, this time considering the proportion of “pleasant” responses after White and Black primes separately. This analysis showed no significant effects for White primes. However, when looking at the Black primes only, two interactions were marginally significant: duration x awareness, $F(1,73) = 3.14$, $p = .081$, and duration x awareness x External motivation $F(1,73) = 2.99$, $p = .088$.

As the interaction between the different factors may be stronger in one of the two awareness groups, and also because being the awareness factor the only between-subjects variable present the interactions above, I decided to run the further analyses for

the two awareness groups separately. Furthermore, in order to reveal a change between the two durations according to Internal and External motivation, I calculated a new score that would have summarized the difference in the proportion of pleasant responses between the two durations, hence eliminating the duration factor. To create this score, I subtracted the proportion of pleasant responses after short Black primes from the proportion of pleasant responses after long Black primes. An increase in this scores, which for brevity I denominated ΔBlacks , would indicate an increase in the proportion of pleasant responses from the short to the long Blacks prime trials. I then did the same for White primes, creating a ΔWhites score.

Figure 8. The correlations between the ΔBlacks score and Internal (left) and External (right) motivation in the supraliminal priming group.

At this point, it is possible to check whether changes in the ΔBlacks score in the two groups were driven by participant's motivations through multiple linear regression by predicting the score from the two motivational subscales. I ran the regression with the two motivations as predictors for the two awareness conditions separately. For the

supraliminal priming group, the analysis revealed a marginally significant effect of Internal motivation, which, as can be seen from Figure 8, correlated positively with the Δ Blacks score, $\beta = .34$, $t(37) = 1.97$, $p = .056$, and a significant effect of External motivation, also positively correlated with the Δ Blacks score, $\beta = .36$, $t(37) = 2.03$, $p < .05$. In the subliminal priming group, no effects were significant. Additionally, no significant effects were found in any group when using the Δ Whites score as the dependent variable.

8.2.1 Predicting explicit attitudes

I ran an additional analysis which, as done by Payne et al. (2005), was aimed at exploring how motivation moderated the relationship between implicit and explicit attitudes. In order to replicate the findings of Payne et al., I expected to find a significant interaction between the AMP score and Internal motivation when predicting people's explicit attitudes. These attitudes were represented by the Self-report score, which was created by subtracting their responses towards Whites from the ones towards Blacks on the feeling thermometer. A higher Self-report would indicate more favourable feelings towards Blacks.

If I wanted to include the duration factor into my model, I could not run the same regression as Payne et al. did because of multicollinearity issues caused by having two similar predictors (the AMP score calculated on short primes and the one calculated on the long ones). To avoid this issue and still include the duration factor, I inserted as dependent variable the AMP score instead of the Self-report. This change would not have influenced my analysis, as now an interaction between motivation and Self-report in predicting the AMP would have represented the same as an interaction between motivation and AMP in predicting the Self-report. I then ran a repeated measures ANCOVA, with awareness as a between-subjects factor, and Self-report, Internal and External motivation as continuous predictors. From this analysis, three effects emerged: a marginally significant main effect of Self-report, $F(1,67) = 2.99$, $p = .088$, a significant interaction between awareness and Self-report, $F(1,67) = 7.95$, $p < .01$, and a significant awareness x Self-report x Internal motivation three-way interaction, $F(1,67) = 7.42$, $p < .01$.

In order to untangle the three-way interaction, I decided to analyze separately the subliminal and the supraliminal priming groups with the same remaining variables, being awareness the only between subjects factor. Additionally, in order to be able to analyze the eventual interactions, I decided to switch to the regression analysis used also by Payne et al. (2005). This time, Self-report was placed as the dependent variable, and not having found any significant effect of prime duration, I put the overall AMP score as a predictor next to Internal motivation⁶. For the supraliminal priming group, I found a marginally significant effect of the AMP, $\beta = -1.31$, $t(36) = -1.99$, $p = .054$, and a marginally significant AMP x Internal motivation interaction, $\beta = 1.27$, $t(36) = 1.84$, $p = .073$. In the subliminal priming group, I found a significant effect of Internal motivation, $\beta = .54$, $t(35) = 3.35$, $p = .002$, and a marginally significant AMP x Internal motivation interaction, $\beta = -1.82$, $t(36) = -1.71$, $p = .097$.

In order to disentangle the two interactions, I then ran a simple slopes analysis to test the relationship between the Self-report and the AMP at -1 *SD* and +1 *SD* of Internal motivation for each of the two awareness groups. In the supraliminal group, at -1 *SD* of Internal motivation, the correlation between the AMP score and the Self-report was marginally significantly negative, $\beta = -2.53$, $t(36) = -1.92$, $p = .06$. At +1 *SD* of Internal motivation, the correlation was not significant, $\beta = -.09$, $t(36) = -.9$, *ns*.

In the subliminal group, at -1 *SD* of Internal motivation, the correlation between the AMP score and the Self-report was positive but not significant, $\beta = 3.55$, $t(35) = 1.66$, $p = .11$. At +1 *SD* of Internal motivation, the correlation was also not significant, $\beta = -.1$, $t(35) = -.52$, *ns*.

Since it was not considered in the regression analyses, prime duration may have played a role in the significant (or non-significant) results found. In fact the interaction between the AMP and Internal motivation could have been much weaker after short or after long primes, without in this way revealing the expected significant correlation between the AMP and the Self-report at -1 *SD* of Internal motivation. With this doubt in mind, I ran a linear regression analysis on predicting the Self-report from Internal

⁶ This process took away the prime duration factor, which may have explained part of the variance. For this reason, I also ran the repeated measures ANCOVA with the AMP score as the DV and Self-report and Internal motivation as predictors, showing a marginally significant interaction between the two both in the supraliminal, $F(1,36) = 3.65$, $p = .064$, and the subliminal priming groups, $F(1,35) = 4.02$, $p = .053$.

motivation and the two AMP (short and long primes) scores separately in the two different priming type groups. All the scores were standardized before the analysis.

Starting with the supraliminal short-primes AMP condition, there was only a marginally significant effect of Internal motivation, $\beta = .25$, $t(36) = 1.90$, $p = .066$. With the long-primes AMP, Internal motivation was marginally significant, $\beta = .30$, $t(36) = 1.86$, $p = .071$, the AMP was significant, $\beta = -1.91$, $t(36) = -2.20$, $p < .05$, and the Internal x AMP Interaction was also significant, $\beta = 1.82$, $t(36) = 2.149$, $p < .05$. The simple slopes analysis of the interaction revealed a significant negative relationship between the Self-report and the long-primes AMP at -1 SD of Internal motivation, $\beta = -3.00$, $t(36) = -2.18$, $p < .05$. At $+1$ SD of Internal motivation, the relationship was not significant, $\beta = -.07$, $t(36) = -.66$, *ns*.

Figure 9. The correlation between the Self-report score and the AMP score calculated on subliminal short-prime trials (a) and long-prime trials (b), and subliminal short- (c) and long-prime trials (d) plotted at -1 (green line) and $+1$ SD (red line) of Internal motivation.

In the subliminal priming group with the short-primers AMP, the results showed only a significant main effect of Internal motivation, $\beta = .49$, $t(35) = 2.97$, $p < .01$. When considering the long-primers AMP as a predictor, there was again a main effect of Internal motivation, $\beta = .61$, $t(35) = 3.8$, $p = .001$, this time accompanied by an effect of the long-primers AMP, $\beta = 2.01$, $t(35) = 2.17$, $p = .04$, and a significant Internal motivation x long-primers AMP interaction, $\beta = -2.70$, $t(35) = -2.15$, $p = .04$. The simple slopes analysis of the interaction, as can be seen on Figure 9, revealed no correlation between the long-primers AMP and Self-report at +1 *SD* of Internal motivation, $\beta = -.06$, $t(35) = -.42$, *ns*. Nevertheless, at -1 *SD* of Internal motivation, the positive correlation between the long-primers AMP and Self-report was significant, $\beta = 4.08$, $t(35) = 2.17$, $p = .04$.

8.3 Discussion

I hypothesized that people correct their initial automatic reactions to faces of individuals belonging to a different race according to their motivations if they are given more time to process them. I also hypothesized that this process occurs unconsciously when people are not aware of being presented the prime images. After the analysis of the results, none of the hypothesized effects were significant. I did not find any significant change in the difference between the proportion of pleasant responses to Blacks compared to White prime trials according to participant's Internal and External motivations to control for prejudice.

I then checked if the relationship between the implicit (AMP) and the explicit measure (Self-report) was moderated by the two types of motivation. Here I found that Internal motivation moderated the implicit – explicit relationship differently in the two prime awareness conditions. In the subliminal priming group, the results were in line with the findings of Payne et al. (2005) and with the predictions of the MODE model (Fazio, 1990). According to the MODE model, when a person is highly motivated not to show any prejudice, there will be no correlation between its implicit and explicit attitudes towards individuals of a different race. Therefore, regardless of the negativity of its automatic reactions, when there will be the opportunity (time and resources) to control that response, the evaluative outcome will be positive. On the other hand, when a person is not motivated to control for prejudice, its implicit attitudes will correlate

with the explicit ones. This is what emerged in the subliminal priming group when considering participants' Internal motivation to control for prejudice. In those with high motivation, the correlation between the implicit and the explicit measures was not significant, whereas in those with low motivation, the same correlation was positive and significant.

In the supraliminal group, however, I found the opposite correlation between the implicit and the explicit measures when Internal motivation was low. What it suggests is that those participants with a low motivation to control for prejudice, the more they were implicitly prejudice the more they reported positive feelings towards Blacks compared to Whites, and vice-versa (the lower the implicit prejudice, the more negative the self-reported feelings towards Blacks). Not having an explanation for the unexpected pattern that emerged, I leave it to future research to explore it further.

Furthermore, it is important to take into account that in the studied sample there was no implicit prejudice towards Blacks to be corrected in the first place. The proportion of pleasant responses on Black trials was even marginally significantly more positive than the one on White trials. In this case, finding an increase in the proportion of pleasant responses on Blacks primes trials between the two prime duration conditions according to motivational factors, would have meant finding an increase in the already present positivity towards Blacks compared to Whites.

The Dutch student population of which the sample is part of may indeed be different from the American one for what concerns prejudice towards Blacks. It would be interesting to test the same paradigm on that same population in which implicit prejudice has already been assessed and which was used in both the studies of Payne et al. (2005) and the ones of Cunningham et al. (2004) that inspired the present research.

For the abovementioned reason, it may be that the correctional effect of motivation was too weak to be detected. For exploratory purposes, I ran a more detailed analysis to see if, although not significant, a pattern in the expected direction emerged from the data in the different prime awareness conditions and for Black and White prime trials separately. What I found was a positive correlation between the increase in the proportion of pleasant responses from short Black prime trials to the long ones and the two types of motivation to control for prejudice in the supraliminal priming group. No correlations emerged for White primes. This suggests that there may have been an

increase in the positivity of the automatic evaluation towards Blacks according to participants' motivations, whereas automatic evaluations of Whites were unaffected. Even though the pattern goes in the hypothesized direction, the effect was not strong enough for us to conclude that the increase in positivity towards Blacks was significantly greater than the one towards Whites.

When looking at the subliminal priming group, this exploration of the data did not reveal any significant effects. But before concluding that there is no unconscious shaping of the evaluative response according to peoples' motivations, it is still unknown whether the primes were successfully elaborated when presented subliminally. The briefness of the suppressed image in the short primes condition may not have lead to a successful elaboration of it, and therefore, even if we found an increase in positivity short Black primes and the long Black primes, that may not have been because of the correction of a negative automatic attitude, but rather the fact that only the long primes were strong enough to elicit a reaction.

Not having individually adjusted the amount of decreased contrast of the primes for each participant, while for some of them (excluded from the analysis) the signal was strong enough to be consciously detected, for others it may have been too weak to be unconsciously perceived, without revealing any effect. To avoid this potential issue and have more chances to obtain a subliminal perception and trigger an unconscious elaboration, future studies should consider adjusting the amount of contrast of the primes for each participant. In this way, each participant can be primed with a picture whose contrast is low enough for him not consciously perceive it, but not too low to completely prevent its further elaboration.

One last factor that has to be taken into account when comparing the analysis results in the subliminal group to the ones in the supraliminal group, is the significantly lower overall proportions of pleasant responses in the former. This finding suggests that the subliminal priming condition may have induced a negative affect, which may have in turn influenced participant's activation of implicit attitudes (e.g., Clore & Huntsinger, 2009).

9. Conclusion

In the present thesis, I tried to give an answer to the question about the possibility of an unconscious correction of people's automatic evaluations according to their motivations in the domain of racial prejudice. A new interesting method was introduced, combining a technique of perceptual suppression with a projective measure of attitudes. According to the iterative reprocessing model (Cunningham & Zelazo, 2007), I expected an increased processing time to transform the negative automatic reactions towards faces of Black individuals (compared to Whites) by making them become more positive if a person was motivated not to be prejudiced. Given the automaticity of the initial reaction and the hypothesized automaticity of the underlying dynamic correctional process regardless of the awareness of the stimuli, I did not expect the unawareness of the primes obtained via subliminal presentation to disrupt it.

In the main study, I looked at how the evaluation of primes consisting Black faces changed from a brief exposition to a longer one. I expected to find a correlation between this change in the evaluation and participants' motivation to control for prejudice that was driven by internal and external factors both with the supraliminal and subliminal presentation of the primes.

In the end, the hypothesized effect of motivation did not emerge. However, before drawing our conclusions on the null effect of motivation in this context, it is important to take into account the overall lack of implicit and explicit prejudice towards Blacks in the studied sample. Without a negative automatic evaluation, there was not much left to correct.

Finally, an exploratory analysis of the data revealed the expected effects of Internal and External motivation on the increase in the proportion of pleasant responses on longer Black primes in the subliminal priming group. This effect of motivation may have not been strong enough to appear in the main analysis because of the lack of prejudice of my sample. If this was the reason behind the too weak effect, for the future research testing it, a sample taken from a population having a strong implicit prejudice may represent the key to its unveiling.

REFERENCES

- Anderson, E., Siegel, E., White, D., & Barrett, L. F. (2012). Out of sight but not out of mind: Unseen affective faces influence evaluations and social impressions. *Emotion, 12*(6), 1210–1221.
- Bargh, J. A., Chen, M. & Burrows, L. (1996). Automaticity of social behavior: Direct effect of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology, 71*, 230-244.
- Blake, R. & Logothetis, N.K. (2002). Visual competition. *Nat. Rev. Neurosci. 3*, 13–21.
- Bos, M. W., & Dijksterhuis, A. (2011). Unconscious thought works bottom-up and conscious thought works top-down when forming an impression. *Social Cognition, 29* (6), 727-736.
- Bruner, J. S. (1957). *Going beyond the information given*. New York: Norton.
- Cacioppo, J. T., & Berntson, G.G. (1999). The affect system: Architecture and operating characteristics. *Current Directions in Psychological Science, 8*, 133-137.
- Carraro, L., Gawronski, B., & Castelli, L. (2010). Losing on all fronts: The effects of negative versus positive person-based campaigns on implicit and explicit evaluations of political candidates. *British Journal of Social Psychology, 49*, 453–470. doi: 10.1348/014466609X468042
- Castelli, L., Carraro, L., Gawronski, B., & Gava, K. (2010). On the determinants of implicit evaluations: When the present weighs more than the past. *Journal of Experimental Social Psychology, 46*, 186-191.

- Castelli, L., Carraro, L., Tomelleri, S., Amari A. (2007). White children's alignment to the perceived racial attitudes of the parents: Closer to the mother than the father. *British Journal of Developmental Psychology*; 25:353–357.
- Castelli, L, Zogmaister, C., Tomelleri, S. (2009). The transmission of racial attitudes within the family. *Developmental Psychology*; 45:586–591. [PubMed: 19271841]
- Clore, G. L., & Huntsinger, J. R. (2009). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, 11, 393–399.
- Cooley, E., Payne, B. K., & Phillips, K.J. (2014). Implicit bias and the illusion of conscious ill will. *Social Psychological And Personality Science*, 5(4), 500-507. doi:10.1177/1948550613506123
- Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using race to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, 83, 1314–1329.
- Cunningham, W. A., Johnson, M. K., Raye, C. L., Gatenby, J. C., Gore, J. C., & Banaji, M. R. (2004). Separable neural components in the processing of black and white faces. *Psychol. Sci.* 15, 806–813
- Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *J. Cogn. Neurosci.* 16, 1717–1729
- Cunningham, W., & Zelazo, P. D. (2007). Attitudes and evaluation: A social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, 11, 97-104.
- Cunningham, W. A., Zelazo, P. D., Packer, D. J., & Van Bavel, J. J. (2007). The iterative reprocessing model: A multilevel framework for attitudes and evaluation. *Social Cognition*, 25, 736–760.

- Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., and Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn. Sci.* 10, 204–211.
- Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, 91, 385–405.
- Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science*, 14, 14-18.
- Dijksterhuis, A. J. & Nordgren, L. F. (2006). A theory of unconscious thought. *Perspectives on Psychological Science*, 1 (2), 95-109. doi: 10.1111/j.1745-6916.2006.00007.x
- Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, 11(1), 86–89.
- Dixon, T. (2008). Crime news and racialized beliefs: Understanding the relationship between local news viewing and perceptions of African Americans and crime. *Journal of Communication*, 58, 106–125. doi: 10.1111/j.1460-2466.2007.00376.x.
- Dovidio, J. F., Kawakami, K., Johnson, C., Johnson, B., & Howard, A. (1997). On the nature of prejudice: Automatic and controlled processes. *Journal of Experimental Social Psychology*, 33, 510–540.
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers, Fort Worth, TX.
- Earp, B. D., Dill, B., Harris, J., Ackerman, J. M., & Bargh, J. A. (2013). No sign of quitting: Incidental exposure to no-smoking signs ironically boosts cigarette-

approach tendencies in smokers. *Journal of Applied Social Psychology*, 43, 2158-2162.

Egloff, B., & Schmukle, S. C. (2003). Does social desirability moderate the relationship between implicit and explicit anxiety measures? *Personality and Individual Differences*, 35, 1697–1706.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75-109). New York: Academic Press.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology*, 54, 297–327.

Fazio, R. H., & Olson, M. A. (2014). The MODE model: Attitude-behavior processes as a function of motivation and opportunity. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual process theories of the social mind* (pp. 155-171). New York: Guilford Press.

Fazio, R. H., & Dunton, B. C. (1997). Categorization by race: The impact of automatic and controlled components of racial prejudice. *Journal of Experimental Social Psychology*, 33, 451–470.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69, 1013–1027.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, 50, 229–238.

- Foroni, F. & Semin, G. R. (2012). Not all Implicit Measures of Attitudes are created equal: Evidence from an embodiment perspective. *Journal of Experimental Social Psychology*, 48, 424-427.
- Gaertner, S. L., & Dovidio, J. F. (1986). The aversive form of racism. In J. F. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 61–89). Orlando, FL: Academic Press.
- Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology*, 50, 141-150.
- Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, 132, 692–731.
- Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE Model. *Social Cognition*, 25, 687–717.
- Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology*, 44, 59–127.
- Gawronski, B., & Bodenhausen, G. V. (2014). The associative–propositional evaluation model: Operating principles and operating conditions of evaluation. In J.W. Sherman, B. Gawronski & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 188–203). New York: Guilford Press.
- Gawronski, B., & Payne, B. K. (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York: Guilford Press.

- Gladwell, M. (2007). *Blink*. Back Bay Books. ISBN 978-0-316-01066-5.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 192, 4–27.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Greenwald, A. G., Liu, T. J. (1985). Limited unconscious processing of meaning. *Bulletin of the Meetings of the Psychonomic Society* 23:292–313.
- Greenwald, A. G., Oakes, M. A., & Hoffman, H. G. (2002). Targets of discrimination: Effects of race on responses to weapon holders. *Journal of Experimental Social Psychology*, 39, 399–405.
- Hilton, J. L., & von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology* 47:237–271.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measure. *Personality and Social Psychology Bulletin*, 31, 1369–1385.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measure. *Personality and Social Psychology Bulletin*, 31, 1369–1385.
- Hofmann, W., Gschwendner, T., Castelli, L., & Schmitt, M. (2008). Implicit and explicit attitudes and interracial interaction: The moderating role of situationally available control resources. *Group Processes and Intergroup Relations*, 11, 69-87.

- Hofmann, W., van Koningsbruggen, G. M., Stroebe, W., Ramanathan, S., & Aarts, H. (2010). As pleasure unfolds: Hedonic responses to tempting food. *Psychological Science*, 21, 1863–1870. doi: 10.1177/0956797610389186
- Imhoff, R., & Banse, R. (2009). Ongoing victim suffering increases prejudice: The case of secondary antisemitism. *Psychological Science*, 20, 1443-1447.
- Koole, S. L., Dijksterhuis, A., & van Knippenberg, A. (2001). What's in a name: implicit self-esteem and the automatic self. *Journal of Personality and Social Psychology*, 80, 669-685.
- Krendl, A. C., Macrae, C. N., Kelley, W. M., Fugelsang, J. F., Heatherton, T. F., (2006). The good, the bad, and the ugly: An fMRI investigation of the functional anatomic correlates of stigma. *Social Neuroscience*;1(1):5–15.
- Lamme, V. A. (2003) Why visual attention and awareness are different. *Trends Cogn. Sci.* 7, 12–18
- Ma, D. S., Correll, J., & Wittenbrink, B. (in press). The Chicago Face Database: A Free Stimulus Set of Faces and Norming Data. *Behavior Research Methods*.
- Miles, W. R. (1930). Ocular dominance in human adults. *The Journal of General Psychology*, 3(3), 412–430.
- McConahay, J. B. (1986). Modern racism, ambivalence, and the Modern Racism Scale. In J. E. Dovidio & S.L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 91-126). San Diego, CA: Academic Press.
- McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in America? It depends on who is asking and what is asked. *Journal of Conflict Resolution*, 25, 563-579.

- 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, 226–254.
- Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, 134, 565–584.
- Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition*, 20, 89–104.
- Olson, M. A. & Fazio, R. H. (2007). Discordant evaluations of Blacks affect nonverbal behaviour. *Personality and Social Psychology Bulletin* 33(9): 1214-1224.
- Payne, B. K. (2001). Prejudice and perception: The role of automatic and controlled processes in misperceiving a weapon. *Journal of Personality and Social Psychology*, 81, 181-192.
- Payne, B. K. (2006). Weapon bias: Split second decisions and unintended stereotyping. *Current Directions in Psychological Science*, 15, 287-291.
- Payne, B. K. & Cameron, C. D. (2014). Dual Process Theory from a Process Dissociation Perspective. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 188-203). New York: Guilford Press.
- Payne, B. K., Lundberg, K. (2014). The Affect Misattribution Procedure: Ten Years of Evidence on Reliability, Validity, and Mechanisms. *Social and Personality Psychology Compass* 8/12 (2014): 672–686, 10.1111/spc3.12148
- Peters, K. R., & Gawronski, B. (2011). Mutual influences between the implicit and explicit self-concepts: The role of memory activation and motivated reasoning. *Journal of Experimental Social Psychology*, 47, 436–442.

- Ranganath, K. A., Smith, C. T., & Nosek, B. A. (2008). Distinguishing automatic and controlled components of attitudes from direct and indirect measurement. *Journal of Experimental Social Psychology*, 44, 386-396.
- Riketta, M. (2006). Gender and socially desirable responding as moderators of the correlation between implicit and explicit self-esteem. *Current Research in Social Psychology*, 11, 14-28.
- Sanbonmatsu, D. M., & Fazio, R. H. (1990). The role of attitudes in memory-based decision making. *Journal of Personality & Social Psychology*, 59, 614-622.
- Scholte, H. S., Witteveen, S. C., Spekreijse, H., & Lamme, V. A. F. (2006). The influence of inattention on the neural correlates on scene segmentation. *Brain Research*, 1076, 106-115.
- Schumann, H., Steeh, C., & Bobo, L. (1985). *Racial Attitudes in America*. Cambridge, MA: Harvard University Press.
- Shiffrin, R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127-190.
- Sklar, A. Y., Levy, N., Goldstein, A., Mandel, R., Maril, A., & Hassin, R. R. (2012). Reading and doing arithmetic nonconsciously. *Proceedings of the National Academy of Sciences*, 109(48), 19614-19619.
- Smith, E. R., Fazio, R. H., & Cejka, M. A. (1996). Accessible attitudes influence categorization of multiply categorizable objects. *Journal of Personality and Social Psychology*, 71, 888-898.
- Tajfel, H. (1969). Cognitive aspects of prejudice. *Journal of Social Issues*, 25, 79-98.

- Troiani, V., Price, E. T., & Schultz, R. T. (2012). Unseen fearful faces promote amygdala guidance of attention. *Social Cognitive and Affective Neuroscience*. doi: 10.1093/scan/nss116
- Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, 8(8), 1096–1101. doi: 10.1038/nn1500
- Tukachinsky, R., Mastro, D., & Yarchi, M. (2015). Documenting portrayals of race/ethnicity on primetime television over a 20-year span and their association with national-level racial/ethnic attitudes. *Journal of Social Issues*, 71(1), 17–38.
- Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101–126.
- Wolfe, J. M. (1984). Reversing ocular dominance and suppression in a single flash. *Vision Res.* 24, 471–478.

APPENDIX

File names of the pictures from the Chicago Face Database used in the studies

Blacks	Whites
CFD-BM-002-013-N	CFD-WM-003-002-N
CFD-BM-005-003-N	CFD-WM-004-010-N
CFD-BM-010-003-N	CFD-WM-011-002-N
CFD-BM-019-002-N	CFD-WM-012-001-N
CFD-BM-021-021-N	CFD-WM-013-001-N
CFD-BM-022-022-N	CFD-WM-014-002-N
CFD-BM-026-002-N	CFD-WM-015-002-N
CFD-BM-032-024-N	CFD-WM-016-001-N
CFD-BM-037-033-N	CFD-WM-024-015-N
CFD-BM-039-029-N	CFD-WM-029-023-N
CFD-BM-041-035-N	CFD-WM-031-003-N
CFD-BM-045-004-N	CFD-WM-033-025-N