LUNDS UNIVERSITET

DEPARTMENT OF PSYCHOLOGY

Intentional forgetting of stereotypes

Helena Lundström

Magister thesis spring 2013

Supervisor: Mikael Johansson

Abstract

This study investigates the possibility of intentional forgetting of congruent and incongruent stereotypes belonging to either the in-group or an out-group. Previous research has shown that we perceive people differently depending on whether they belong to our in-group or an out-group. Differences have also been shown in how we process information that is congruent or incongruent to a stereotype. This was studied in a directed forgetting test where participants were asked to learn a list of both stereotype-congruent and –incongruent features belonging to persons from both the in- and out-group. The were then asked to forget the first list and learn a new which was identical to the first list but with new features. The main hypothesis that there will be a difference in recall of stereotype-congruent vs. stereotype-incongruent features depending on if they belong to the in- or out-group and if they were in the first or second list. The prediction was also made that the effect of directed forgetting would be mediated by how strong the participants' stereotypic thinking was and their ability to recruit inhibitory control, as an effect of working memory capacity. This was tested by IAT and OSPAN. The hypothesis was partly confirmed and the results showed a significant difference in recall where participants recalled more stereotype-related information concerning the out-group but were also able to intentionally forget them. For the in-group, participants recalled more of the incongruent features but there was no significant difference between the first and second list. Implicit attitudes and working memory capacity did not influence the pattern of results. The present study reveals a difference in how we perceive and are able to forget congruent and incongruent stereotypic information about the in- and out-group.

Keywords: *Stereotypes, Directed forgetting, in-group, out-group, working memory capacity, IAT*

Introduction

Stereotypes acts as a guide that can help us navigate in a social context and have long been known to influence how we perceive and interpret other people's behavior. In some situations they can serve as a good guideline but there are also side effects to stereotypes. Stereotypes is a form of categorization, we tend to generalize some features to other groups of people that can make us prejudiced towards others. When we rely on these stereotypes it can mislead us to interpret a situation or person in the wrong way. For instance, light-skinned tend to interpret ambiguous behavior in a more negative way when it is performed by a black actor than when performed by a white actor (Duncan, 1976). With the help of our stereotypes we often quickly form an opinion about a person we meet. But what happens when this opinion or belief turns out to be wrong? Imagine meeting a person that you perceive as being a bit hostile. The next time you meet that person he acts nicely and it turns out that he just had a really bad day the last time you saw him. Although you now know that this person can be nice, it might take you a long time to form a new opinion about this person. Hence, our opinions about other people can be hard to change, especially when we rely on stereotypical beliefs. In order to reduce our stereotypes we need to understand its underlying mechanisms and this has been explored mainly within the field of social psychology.

Although we often are aware of the stereotypes considering our own in-group, there are differences when encountering a person from the out-group or in-group. The out-group homogeneity effect is the tendency to presume that everyone in the out-group is similar to each other (Gilovich, Keltner & Nisbett, 2006). For instance, we believe that they all look and think alike. When we meet a person from the out-group, we then tend to treat them as a representative of the group. For the in-group, on the other hand, this effect does not emerge and we tend to encounter a member of the in-group on an individual-to-individual basis. There is a simple explanation to this, we normally have more in-group contacts that give us more opportunities to encounter evidence that violates the stereotype, and we learn that our in-group consists of people with different opinions, looks and feelings. The amount of interaction with a social group thus seems to have a large effect on how we encounter and treat a person from the out-group. A lot of out-group contact is therefore thought to be one way of reducing stereotypes about that specific group (Gilovich, Keltner & Nisbett, 2006). Research has shown that by solely imagining intergroup contact with a person who belongs to a negatively stereotyped group can reduce hostility against this group and also change the stereotypical belief (Brambilla, Ravenna & Hewstone, 2011). Thus, it is possible to change the beliefs about an out-group by actively trying to do so.

As noted earlier, when we encounter a person from the out-group and allow ourselves to get influenced by our stereotypes, we will remember the stereotypical traits more than individuating traits, even though the stereotypical traits may not make justice to the person encountered. Luckily, it turns out it works the other way around as well. Our memory plays a big role in how we form and rely on our stereotype beliefs. Dunn & Spellman, 2003, showed that by practicing non-stereotype traits about an out-group member, we can inhibit the recall of a stereotype concerning the same group and thereby forget our stereotypes. Thus, the reduction of stereotypes is a cognitive task related to our memory. Again, research has shown that by actively focusing on a persons individuating traits, often incongruent or irrelevant to the stereotype, rather than relying on stereotypical traits can reduce or make us forget our stereotypes.

Forgetting often happens unintentionally or automatically. This is dependant on our cognitive resources, it would be impossible to remember everything and nonetheless unnecessary. But we can also forget intentionally, where we block or inhibit unwanted memories. This too is an adaptive function. For instance, suppression of personal unwanted memories has been under discussion in psychology for a long time (Freud, 1938). It can also be adaptive in other ways, imagine a jury being told that they have heard incorrect information and false accusations against a prosecuted, they then have to be able to dismiss that information. Also, to appear non-prejudice one might want to intentionally forget their stereotypes, in order to not let them hinder us from getting the right impression of a person or not act upon the stereotypes. Intentional forgetting has commonly been studied through a paradigm called *Directed Forgetting* (DF).

Directed Forgetting

There are two different types of DF, the item-method and list-method. In a typical DF, using the list-method, the participant is told to learn to remember a list of items (list 1). After this study phase, the participant is instructed to either forget or remember the subset of items in the test. A new list of items to be studied (list 2) is given to the participant. After the second study phase, the participant is tested on how many items they remember from both the first and second list. DF-tests reveals that participants remember less items from list 1 than list 2, which indicates that they intentionally forgot most of the items from the first list (MacLeod, 1975). The effects that emerge can be measured in two different ways. Typically, participants in the forget-condition demonstrate a poorer recall of the first list compared to the

remember-condition. This is known as the *costs* of DF. The fact that in the forget-condition, you also typically remember more items from the second list compared to the remember-condition is known as the *benefits* of DF (Sahakyan & Goodmon, 2007). The list-method can be used as described above, with one forget condition and one remember condition. It can also be used as a within-design, where all participants are asked to forget the first list and remember the second.

In the item-method all items are presented in one list, directly after the item has been shown the participant is cued to either forget or remember the item (MacLeod, 1975).

Theories of Directed Forgetting

While the two methods both produces a DF-effect, research has shown that there might be different memory mechanisms for each method. Both methods produces an effect when the participant is tested on their memory through free recall, but when using a recognition test for memory performance the DF-effect has only been observed in the item-method. The effects of the item-method have been debated but the most prominent explanation is *selective rehearsal.* This theory suggests that the participant keeps rehearsing and processing the item until the cue appears, stops processing if instructed to forget the item and continues to rehearse if instructed to remember. When memory performance is then later tested, participants consequently recall the to-be-remembered items better as they were processed in a more elaborative way. This explanation however, does not apply to the list-method, as the cue to forget does not appear until the all of the items has been presented in the first list. Therefore, all items must have been rehearsed and processed to the same extent.

The dominant theory to explain the phenomena of the list-method was proposed by Bjork (1989), and states that the effect is due to *retrieval inhibition.* According to this view, all items are encoded and presumably rehearsed by the participant when presented with the first list. When the forget-instruction appears, the items from the first list are inhibited in order to not interfere with the encoding of the new list. When tested on both list, participants consequently have a lot more trouble remembering the inhibited items. This account explains why no DF-effects have been observed when using recognition in the test phase. The inhibition or blocking of the item is thought to be released when seeing the item and therefore produces no DF-effect. Geiselman, Bjork and Fishman (1983) tested the inhibition account in a series of different experiments. In these experiments, participants were presented with a list of words where they were told to remember every other word and the remaining words were

to be judged on pleasantness, but not remembered. As in a typical DF, half way through the list, participants were told to forget all the words they have been presented to, that the first list was solely for practice. The result revealed that when commanded to forget, it negatively affects participants' later memory for both the words they were meant to judge (but not remember) and for the words meant to-be-remembered. Geiselman, Bjork and Fishman argue that this finding is a support for the inhibition account and that the effect cannot be explained in terms of encoding mechanisms as selective rehearsal. They propose that the forget-instruction actuates a process that inhibits or block routes to the first-list-words in episodic memory.

This theory has been challenged by the *context change account* (Sahakyan and Kelley, 2002). The account introduces the idea that when the forget-instruction appears, participants actively change their state or mood (internal context). When then being tested and trying to retrieve the items, they remain in their state from the second list and which thereby facilitates the retrieval of the items from this list and prevents recall from the first list because they were encoded in another state. This view has been supported by studies when participants remembered more items from the second list when they were induced to be in different states during the first and second list, even when the forget-instruction was absent. This effect did not occur when participants were in a consistent internal context through the whole test.

Both of the above theories propose that the effects of DF occur at retrieval as opposed to the selective rehearsal account. Whereas the *inhibition account* states that retrieval is inhibited and thereby blocked for recall, the *internal context account* proposes that the retrieval is being hindered by the shift of internal context.

MacLeod (1999) also pointed out that because the forget-instruction appears after the items in the first list has already been processed, the DF-effect is most likely to be a result of a post-processing restructuring of the items in memory. Retrieval inhibition is a mechanism that can provide this type of restructuring of memory. It has also been argued that if inhibitory processes serve as an adaptive function, it needs to be under voluntarily control. Otherwise not just irrelevant information would be inhibited but also the information we want to remember. Hence, inhibition can be seen as an intentional and goal-serving mental mechanism (Wegner & Schneider, 1989).

Working memory and Directed Forgetting

A high working memory capacity (WMC) is characterized by an ability actively maintain

accessibility of items in memory and also an ability to handle cue overload. A low WMC on the other hand, is characterized by failing to utilize an appropriate strategy to access cues in order to recruit the right memory (Unsworth, Spillers & Brewer, 2012). WMC is also thought to be related to inhibitory control. Research has shown that individuals with higher WMC are better than those with low WMC at recruiting the inhibition mechanism (Redick, Heitz, & Engle, 2007). Thereby, individuals high in WMC can better dismiss irrelevant or unwanted information and should also perform different in a DF-test than those low in WMC. WMC is commonly tested with *Operation Span Task* (OSPAN), which requires the participant to solve mathematical operations while remembering letters or words at the same time. This task demands the participants to control inhibition of irrelevant information competing for attention in order to actively maintain the relevant information in memory.

Delaney & Sahakyan (2007) found that participants with high WMC remembered fewer items from the first list, showing that they in fact are better at inhibiting unwanted information (as they were told to forget the first list). But no interaction was found between WMC and recall for the second list. Another study revealed a different result, this time the results suggested that high WMC enhances recall on the second list but does not affect forgetting of the first list (Soriano & Bajo, 2007). These inconsistent results, however, might be due to the different testing procedures used in the experiments. A third study mixed the designs of the previous studies. In this study, the experiment was the divided into two blocks for each participant. The lists contained of unrelated words, in the first block the participant was asked to either forget or remember the first list and was instructed to do the opposite in the second block. After the DF, an OSPAN task was used to measure WMC. Through this design, it was demonstrated that high WMC predicts both enhanced recall for the second list and better inhibition (forgetting) of the second list (Aslan, Zellner & Bäuml, 2010). So, research has shown that WMC *does* affect the outcome of DF-tests, although it is still not completely clear *how* it works. It is possible that a high WMC enhances the function to maintain different mental contexts, the relation between DF and WMC does not necessarily have to point out inhibition as the only effect. More research has to be done before making further conclusions.

The DF-paradigm has commonly been used to explore how we can intentionally dismiss irrelevant or invalid information. If we through a DF-test deliberately inhibit memories and irrelevant or invalid information, is it then possible that we can intentionally

forget our stereotypes? And if we can intentionally forget our stereotypes, is this being mediated by our WMC?

Stereotypes and Directed Forgetting

More or less everyone has stereotypes concerning different out-groups. Often we are unwilling to express our stereotypical beliefs explicitly; therefore an implicit measure can be more reliable. One common way to test against which groups and how strong these stereotypes are is through the *Implicit Association Test* (IAT), (Greenwald, McGee & Schwartz, 1998). IAT is based on participants' reaction time and measures the difference when between associating 2 targets (e.g. dark- or light-skinned faces) with 2 evaluative attributes (e.g. positive or negative words). It is a sorting task where the participant follows a certain rule to press one key button when the target or attribute belongs to the category on the right and another key button if it belongs to the category on the left. After practice phase, the stimuli is combined where one target and one attribute appears to the right and the other two to the left, thus, they share the same response key (e.g. dark-skinned+positive word, light-skinned+negative word). The combination is then switched (e.g. light-skinned+positive word, dark-skinned+negative word). Hence, this results in two reaction times and comparing the reaction times for the two different combinations assesses the scores. A faster reaction time is thought to measure a higher association. For instance, a faster mean reaction time when using the same response key for negative words and pictures of dark-skinned faces compared to positive words, is assumed to reveal an automatic negative association for dark-skinned. It is expected that the reaction is slower when the targets and attributes are incongruent to the participant.

Although IAT is held as a reliable measurement of stereotypes and is widely used, it has met some criticism. It has been debated whether IAT captures the individuals or cultural evaluations, thus, when being exposed a lot to a certain culture we tend to adopt the cultures beliefs and stereotypes that then reflects the results of the IAT (Greenwald, McGhee & Schwartzman 1998). Another criticism is based on familiarity, suggesting that IAT does not reflect preferences but which target we are more familiar with. For instance, a light-skinned person would then have a faster reaction when combining a light-skinned person with a positive word rather then when combined with a dark-skinned, due to the fact that this person is more frequently surrounded by and therefore more familiar with light-skinned people (Rudman, Greenwald, Mellott & Schwartz 1999).

Considering the fact that we are unwilling to explicitly express stereotypes, this suggests that we do not want to either appear or think of ourselves as being prejudiced. In today's modern society a lot more people are striving for equality amongst different social groups, bringing them to want to encounter people based on their individuation and not by stereotypes. It is therefore an important question if we can inhibit those stereotypes.

Prior work has tested if it is possible to intentionally forget stereotypes via DF. When dealing with stereotypes as opposed to neutral items some factors are worth to consider. DF-tests can be attentionally demanding, due to the mental resources available. One aspect of the effects of DF is the possibility of proactive interference. Proactive interference occurs when already existing information interfere and therefore makes it difficult to learn new information. A fundamental function of memory is the relation between different items and memories. This is an adaptive function, although it can create problems as well. Two items associated with the same trigger word will compete for recall. For instance, simply parking your car everyday in the same parking lot can make it difficult to remember where you parked it today as opposed to yesterday or the week before. Nevertheless, the idea is that frequently recruited memories will gain easier access in memory as opposed to less frequent memories as the connection between those and the trigger word will weaken. Hence, if you parked your car in the same spot everyday, you will have no problem remembering where you parked it.

When encoding the second list in a DF, participants often experience proactive interference, from the items in the first list. Previous research suggests that the inhibition in a DF-test depends on how strong the items in the second list compete in memory with the first list that is meant to be forgotten. When the items in the second list appears to be weaker compared to the first list then no competition appears for the first list, as no inhibition occurs, no effect of DF can be observed (Conway et al., 2000). As stereotypes are a schematic knowledge, easy brought to mind, stereotypes are thought to produce more proactive interference than non-stereotypic features, however this will be discussed further below. As the intensity of proactive interference increases, the attentional demands should as well, which will aggravate the recruitment of inhibitory controls. In a typical DF, the instruction to forget list 1-items impairs later recollection of those. At the same time, they facilitate the recollection of list 2-items, since the forget-instruction reduces the memory load. This effect emerges given that the participant can disregard the list 1-items. But what happens when the first list is more memorable (as stereotype-congruent information is thought to be) than the second list?

Macrae et. al, 1997, tested the possibility to intentionally forget stereotypes in a series of experiments. They predicted that if a certain stereotype was activated, then a particular pattern of the DF-effects would appear if the items meant to be forgotten was stereotype-congruent in relation to the activated stereotype. Since the list to be forgotten would produce proactive interference and thereby more difficult to forget, it was proposed that it would dimmish the memorial advantage for the second list.

In the first experiment, half of the participants were primed with an article to activate the stereotype considering a child abuser. The first list in the DF contained 10 stereotype-congruent features for a child abuser. Memorizing this list was thought to be easier for the primed participants. After the first list, half of the primed and half of the unprimed participants were told to forget the list; nothing was said to the other half of the participants. The second list also contained 10 stereotype-congruent features for a child abuser, however these features were irrelevant in relation to the article. The results revealed that when asked to freely recall the features a DF-effect appeared. Hence, participants asked to forget the first list had more trouble recalling that list and had a better recollection of the second list. Also, participants primed with the article remembered significantly fewer features from the second list compared to those who were not primed. When the material in the first list appeared stereotype-congruent, the instruction to forget failed to facilitate recollection of the second list, an effect that is normally observed in a DF-test. Considering these results, it was concluded that intentional forgetting of stereotype-congruent information is a highly demanding task, more effortful than forgetting features that is not connected to a certain stereotype.

If suppressing stereotype information is this demanding, what happens when the perceivers mental resources is reduced when trying to do so? Will the ability to suppress unwanted stereotypical information be undermined? As mentioned above, the activation of stereotypes is an automatic process. It has been argued that the activation may depend on the availability of attentional resources. Since the activation occurs automatically it should be less dependant on cognitive resources than individuating information. However, there is another possible explanation. Because of the unwillingness to show or get influenced by the stereotypes, perceivers often try to inhibit or correct the stereotype. This process is as effortful and cognitive demanding as individuation (Sherman et al., 1998). This assumption will be discussed further below.

Another study by Macrae et. al, 1997, was conducted, this study was identical to the previous, only this time, half of the participants were asked to count the vowels in the features during the second list. This condition was added in order to deplete participants' cognitive resource. Results showed, not surprisingly, that when the perceivers' mental capacity is depleted their ability to dismiss the unwanted stereotypical information was reduced. Hence, our ability to intentionally suppress stereotypes is mediated by the availability of mental resources.

Forgetting stereotypically congruent and incongruent features

As mentioned above, we often rely on stereotype schemes when encountering a person from an out-group and we tend to process the information in a shallow way. Previous research suggests that when we encounter a person from a *new* social group, we typically pay more attention to the person's behavior in order to create a coherent representation of the new social category. The more we are exposed to the new social group, the more we rely on our already existing knowledge of the group, and the information turns expectancy-congruent. When the information is familiar it demands less processing and waste of cognitive resources (Sherman, 1996). As a result, when less processing is needed when encountering a familiar situation or person, the encoding of the unfamiliar can increase if we are willing. Stereotypes thus seem to be efficient.

According to the *encoding flexibility of stereotypes*, stereotypes can facilitate the encoding of information both congruent and incongruent in relation to the stereotype. The encoding of the stereotype-congruent information is facilitated by the already existing stereotype-scheme, however the scheme also allows a more poorly encoding. This in turn, leaves more cognitive resources for encoding the stereotype-incongruent information, especially when the cognitive capacity is restrained. This however, does not suggest that the incongruent information will be fully understood and reduce the stereotype, but solely more processed. Sherman et al. (1998) tested this model by letting their participants form an impression of a person described with 30 features. The features were either congruent or incongruent to the stereotype of the person. The amount of time the participant spent on reading each feature was recorded. Half of the participants were placed in a cognitive load-condition where they were asked to remember an eight-digit number at the same time as forming the impression. The results showed that when the cognitive capacity was high, the participants spent equal time reading the stereotype-congruent and –incongruent words. Although, when the cognitive capacity was constrained, participants spent more time reading

the inconsistent words in relation to the stereotype, hence, spent more resources on these features. The results were interpreted as when cognitive capacity is low, we rely less on stereotypes. Further, when we have plenty of cognitive resources, the encoding of stereotype-congruent and –incongruent information is roughly the same, but the congruent information is typically more memorable as it follows our already existing scheme about the social group.

If stereotype-congruent and –incongruent information is processed differently, how will this affect the encoding and forgetting in a DF-test? Based on the previously presented research, the effects of DF might be different depending on whether the lists contains of stereotype-congruent or –incongruent items and which of these items are meant to be intentionally forgotten. The concept of stereotypes is fluent and therefore can be problematic to test. When using an IAT, what is being measured is solely a preference for one target group over another. Hence, there are difficulties in explicit testing of stereotypes where we assume that some features are connected to the specific target group. Some features can be strongly linked to the target group, but even if they are a common belief, we cannot be sure that everyone agrees with it. For example, it might be a common belief that everyone with a bad memory is old or that all good basketball players are black. But even if those features are strongly linked to the target group, they also can apply to almost every other target group. Stereotypes can also change, as mentioned above; repeated inter-group contact with members of an out-group can have that effect. As one of the criticisms against IAT claims, the stereotypes we measure might be a cultural belief (Coutant et al., 2011). When explicitly testing stereotypes, we have to assume that a specific feature connected to a target group is a general belief that everyone more or less agrees with. When it comes to testing stereotypes in a DF-test, as noted earlier, the effects of proactive interference should be considered.

Araya, Akrami & Ekehammar (2003), investigated the forgetting of stereotypically congruent and incongruent information in a DF-test. Participants were subliminally primed with either a photograph of a dark-skinned referred to as an immigrant (out-group) or a light-skinned referred to as a Swede (in-group). They were then presented with a DF-test, where the first list contained stereotypically congruent and negative words for an immigrant. Half of the group was then told to forget the first list and the other half did not get such an instruction, although, both groups received a second list. The second list contained stereotypically congruent and positive words for a Swede. When asked to recall both lists, the participants primed with an immigrant recalled less stereotypically congruent words when they were

asked to forget the first list than those who did not receive such an instruction. Those who were primed with a member from the in-group, Swede, and so was presented with incongruent information during the first list forgot the same amount of words regardless of if they were instructed to forget or not. This suggests that the incongruent information was processed elaborately. Further, they found no difference in recall of list 1 and 2 in the forget-condition. The effect of DF did not occur when the information that was meant to-be-forgotten was incongruent in relation to the previously primed stereotype. Also, this suggests that the proactive interference was the same when asked to forget regardless of priming condition. This finding is coherent with prior research that proposes that incongruent information is being more elaboratively processed and therefore as memorable as congruent information.

The present study

The present study aims to contribute to the existing literature about directed forgetting by investigating the forgetting of stereotypes, a field which is still relatively unexplored. The field of directed forgetting will be brought closer to social psychology and enable a better understanding of both directed forgetting and the mechanisms behind stereotyping.
The main intention is to investigate the possibility of intentionally forgetting stereotypes and if this type of memory will act as a semantic memory, commonly used in DF-tests. The present study will investigate if there is a difference in intentionally forgetting and then relearn both stereotypically congruent and incongruent information. It will also explore if there is a difference in forgetting this information when it considers an out-group or in-group. Although prior research has investigated this before, to my knowledge, it has solely been done through priming the participants in order to activate a stereotype on beforehand. This study will examine what happens when no stereotype has intentionally been activated. As a more natural situation, participants will be presented with persons from both the in- and out-group connected with both stereotypical congruent and incongruent features.

If it is possible to intentionally forget stereotypes during directed forgetting, the effect should be mediated by at least two factors. The first factor considers the strength of the stereotypes. Depending on whether the stereotype is strong or not there should be a difference in processing and consequently the inhibition of the stereotype. If the participant has a strong stereotype considering the social group, then the features will appear either congruent or incongruent. Concluded by previous research, the effects of directed forgetting acts differently when a stereotype has been activated. The other factor to consider is working

memory. Thus, the amount of working memory capacity has shown to be related to inhibition control and affects the outcome of a DF-test. The ability to inhibit irrelevant stereotypes as well as incongruent information should therefore also be affected.

The hypothesis of the study is the following: 1) Fewer features from the first list of the DF-test will be recalled than from the second list. 2) The stereotype-congruent features will be better recalled than the incongruent features for the out-group and the opposite should appear for the in-group. 3) There will be a difference in recall of stereotype-congruent vs. stereotype-incongruent features depending on if they belong to the in- or out-group and if they were in the first or second list 4) The amount of stereotypic thinking, meausered by IAT, will mediate the recall of stereotype features, where those high on IAT will remember more stereotype features than those low on IAT. 5) The ability to intentionally forget the first list will be mediated by working memory capacity.

Method

Participants

A total of 30 native Swedish participants were recruited amongst friends and family, with a mean age of 30.47 (*SD* = 8.44). The participants all gave their informed consent and took part of the experiment voluntarily and without any reward.

Materials

IAT. The stimuli in the IAT consisted of 16 photographs in color. Half of the pictures depicted northern-European males with light skin and the other half represented middle-eastern males with darker skin. The entire face was shown against a black background. The evaluative attributes consisted of 8 positive and 8 negative Swedish words. The instructions and design of the test was a replication of the tests available at the Project Implicit website (http://implicit.harvard.edu).

Directed Forgetting. The stimuli for the directed forgetting task were 80 features divided into two lists. Half of the features were congruent with the stereotype for a darker skinned person belonging to the participants´ out-group and referred to as an Arab. The other half were congruent with the stereotype for a light-skinned (for the participants´), in-group person referred to as a Swede. A pretest was conducted to determine the features for each stereotype. The 5 participants were given a list of 200 features and told to rate how well each feature would fit in to each stereotype. The participants were told to categorize the features based on what they thought most people would agree upon and not their own beliefs.

4 photographs, showing the entire face against a black background, were used. Half of the pictures depicted an Arab and the other half depicted a Swede. In the first list of the test a typical Swedish or Arabic name was connected to each picture as well as 5 stereotype congruent and 5 stereotype incongruent features. In the second list the same pictures and names were used, this time with 5 new stereotype congruent and 5 stereotype incongruent features. The participants were tested with a cued recall that consisted of a paper with the 4 pictures and names.

OSPAN. A traditional OSPAN was used, based on simple math problems and letter spans. The instructions and design of the OSPAN was identical to the tests available at http://psychology.gatech.edu/renglelab/Eprime2.html.

Design

The experiment was a within-design consisting of 3 different blocks, IAT, DF and OSPAN. The OSPAN and IAT were identical for all participants. For the DF-test the stimuli in list 1 and 2 were randomized in order to neutralize primacy and recency effects among the participants. It was also counterbalanced so that for half of the participants the complete lists were switched. The OSPAN and IAT acted as covariates for the DF.

Procedure

For each participant the experiment consisted of three blocks. The first block started with a display informing the participant that this was a reaction test. The first part was the IAT that consisted of seven phases where the participants were told to categorize target stimuli into categories as fast as possible. In the first phase, the participant had to place the target stimuli (pictures of either Arabs or Swedes) into the correct category (Arabs to the left – Swedes to the right) by pressing the right key button. This was done for 20 trials. The next phase was similar to the first phase, although this time the participant was asked to place words such as ”Happiness” or ”Pain” into the categories (”Bad” to the left, ”Good” to the right). The third phase combined the categories in the first two phases. ”Bad” and ”Arab” to the left and ”Good” and ”Swede” to the right where the participant placed both pictures and words into the categories for 20 trials. The fourth phase was equal to the third, but with 40 trials. The fifth phase was equal to the first phase but the categories swapped sides. In the sixth phase the categories was again combined during 20 trials, this time with ”Swede” and ”Bad” to the left and ”Arab” and ”Good” to the right. The seventh phase was identical with the one before, but with 40 trials.

The directed forgetting block started with a message telling the participant that the next test considered memory performance. In the first list, the participant was the presented with 4 slides, each displayed for 2 minutes. Every slide consisted of a name and photographs of a face of either an out-group person (Arab) or an in-group person (Swede), 2 of each. 10 features was connected to each person, 5 stereotype congruent and 5 stereotype incongruent. The stereotypes were organized in a manner where the congruent features for the out-group acted as incongruent for the in-group (but with a synonym) and the other way around. The participant was told to memorize every person and their features. They were also told that they were going to be tested on their memory later. In order to help the participants encoding, they also had to rate, on a scale from 1-5, how well they thought each feature described the person. This was done on a separate sheet. They were also told that they were allowed to use different methods, such as saying the features out loud, write them down (as long as they did not used it in the later test phase) or try to create a character based on the features to help them remember. After the first list a message explained that the participant had to forget all the features, they were not going to be tested on the first list but that they had to learn a new list consisting of the same persons but with new features. The second list was identical to the first except for the new features; they also had to rate the new features. The pictures with their features were randomized for each participant.

Participants were then tested on their memory through a cued recall. They were given a response sheet showing the pictures and names of the persons in the test and asked to recall all the features in list 1 and 2 and connect those to the right person. They were given 10 minutes to complete the task.

The last block was the OSPAN task, divided into two sections. The first was a practice section and the second was the actual experiment. In the practice section, letters appeared one at a time for 800 ms and had to be recalled in the same order. The participant clicking boxes next to the appropriate letters tested the recall and a feedback was given. The next practice part consisted of math problems. A math operation appeared on the screen; once the participant knew the answer they made a click with the mouse. A number was shown on the screen and the participant clicked the box ”true” or ”false” depending on the number being the right solution to the math operation. Feedback was given after each operation. The last practice part consisted of both letter recall and solving math operations at the same time. The math operation first appeared on the screen and after solving it the letter they had to recall appeared. The experimental part of the test was equal to the last practice session, consisting of 3 sets of each set-size, ranging from 3-7 which makes a total of 75 math problems and 75

letters to be recalled. The participants was told to keep their math accuracy above 85 %, a percentage of their accuracy was displayed during recall in the upper hand right corner. Participants were then debriefed and thanked for their participation.

Results

The scores in the DF-test were calculated by adding up the correctly recollected features combined to the right person and divided by the total number of features connected to each person. The recall scores were then submitted to a 2 Group (in- and out-group) X 2 List (list 1 and 2) X 2 Congruency (congruent or incongruent stereotype) repeated measures ANOVA. A significant interaction effect in recollection of stereotype-congruent and incongruent vs. in- and out-group was found, $F(1, 29) = 20.96$, $p < .001$, shown in Figure 1. Due to the test design, each list in the DF-test consisted of forty features, a high number of items to remember and to get a correct answer, the feature had to be connected to the right person. This resulted in an overall low recall where some participants had no recall for some variables and a non-normal distribution. To further investigate the differences in mean scores as a function of congruency and group, it was done using a nonparametric test, Wilcoxon Signed Rank Test, in order to not violate the assumption of a normal distribution. The results showed a significant difference in recollection with greater recall for the stereotype-congruent-out-group ($Md = .15$) than for the stereotype-congruent-in-group ($Md = .10$), $z = -2.092$, $p < .05$, with a small effect size ($r = .27$). For the stereotype-incongruent-out-group ($Md = .10$) participants showed less recall than for the stereotype-incongruent-in-group ($Md = .15$), $z = -3.669$, $p < .000$, with a medium effect size ($r = .47$).

To investigate the differences between recall of congruent and incongruent features within the in- or out-group, the same procedure was repeated. The result indicates a lower recall of stereotype-incongruent features ($Md = .10$) when belonging to the out-group than of stereotype-congruent features ($Md = .15$) of the same group, $z = -3.083$, $p < .05$, with a medium effect size ($r = .39$). Although a non-significant difference of recall, the results still suggests the opposite for the in-group, where the recall rate of stereotype-incongruent features ($Md = .15$) is greater than recall of the stereotype-congruent ($Md = .10$) features, $z = -1.747$, $p = .081$.

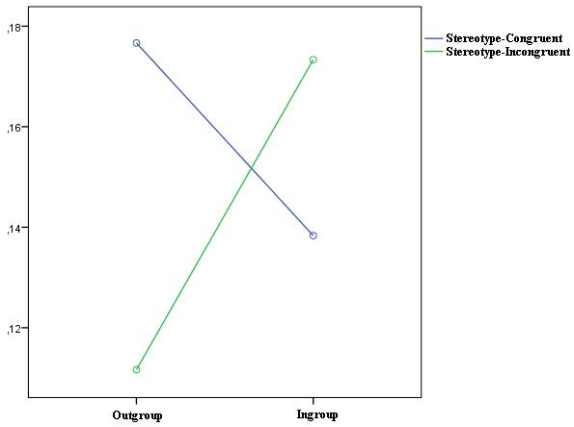

Figure 1. The differences in recall of stereotype-congruent and –incongruent vs. in- and out-group.

The analysis also revealed a significant three way interaction on the participants recall performance, $F(1, 29) = 4.33$, $p < .05$, see Table 1 for means. The resulting plots are shown in Figure 2.

Table 1. *Recall performance in DF.*

Group	**List 1**	**List 2**
Out-group		
Congruent	.097	.257
Incongruent	.083	.140
In-group		
Congruent	.073	.203
Incongruent	.097	.250

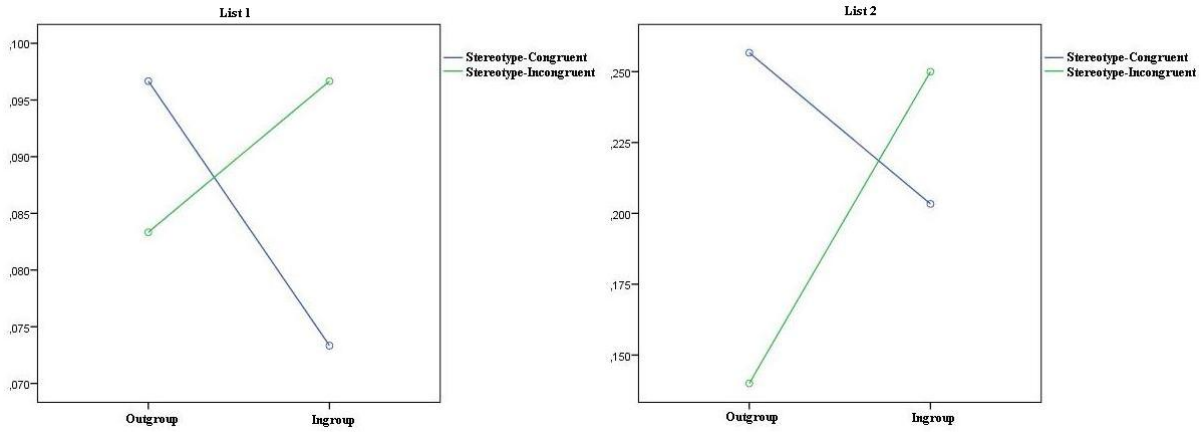

Figure 2. The recall performance in DF as a function of group, list and congruency where the second list act as a baseline and the first list as an effect of a forget instruction.

As Table 1 demonstrates, there is a difference in means when the stereotypes are congruent or incongruent for the out-group between the first and second list. The same pattern appears for the in-group. To follow up the three-way interaction, a simple interaction, List X Congruency, for the outgroup revealed a significant difference, $F(1,29) = 6.07$, $p < .05$. Wilcoxon Signed Rank Test was used to further investigate the means. A significant difference in means was found between recollection of stereotype-congruent features in the first ($Md = .10$) and second ($Md = .20$) list for the out-group, $z = -3.47$, $p < .05$, with a medium effect size ($r = .45$). Although there was a difference in means considering recollection of incongruent stereotypes for the out-group in the first ($Md = .05$) and second ($Md = .10$) list, it was not significant, $z = -1.40$, $p = .16$.

A simple interaction List X Congruency was performed for the in-group which revealed a non-significant result, $F(1,29) = .30$, $p < 1$, ns. There was a difference in recollection between the first ($Md = .00$) and second ($Md = .20$) list considering the congruent stereotypes, $z = -2.81$, $p < .05$, with a medium effect size ($r = .36$). The same pattern appeared for the incongruent stereotypes where participants remembered significantly less in the first list ($Md = .10$) compared to the second list ($Md = .20$), $z = -3.43$, $p < .05$, with a medium effect size ($r = .44$).

The pattern for the out-group, where the recall performance was better for the congruent compared to the incongruent features, was reduced after the forget-instruction. The pattern for the in-group, where the incongruent features were better recalled, did not significantly differ after the forget-instruction.

IAT and OSPAN

To assess the participants' results of the OSPAN, a partial-credit load scoring was used, as described by Conway et al., 2005. To investigate if WMC could predict recollection in the DF-test, participants were divided into two groups, high WMC (M = .88, SD = .07) and low WMC (M = .65, SD = .08) with a median-split. A repeated measure ANOVA with the OSPAN scores as a between-subjects factor revealed no interaction involving OSPAN, $F(2, 28) < 1$, *ns*. Hence, WMC did not affect recollection in the DF-test.

The IAT-scores were attained by applying the algorithm explained by Greenwald, Nosek & Banaji, 2003. The same procedure was done for the IAT results as for the OSPAN, a high group (M = .14, SD = .04) and a low group (M = .05, SD = .04) with no significant result, $F(2, 28) < 1$, ns.

Discussion

The main purpose of this study was to investigate if there was a difference in inhibition of stereotypes among an in- and out-group. This assumption is based on previous research where it has been shown that depending on the congruency of the stereotype and the familiarity of the social group is processed differently. The first hypothesis stated that there would be a difference in recollection of the first and second list, which both consisted of stereotype-congruent and –incongruent features, belonging to both the in- and out-group. A significant result was attained, although, the result should not appear surprising, as the second list was learned closer in time to the memory performance test. The second list can be considered as a baseline for the test, what are interesting are the results of the first list compared to the second. The results attained when analyzing the first and second list confirms the second hypothesis, which stated that there should be a difference in recollection of the in- and out-group features depending on congruency.

As shown in the results, in the second list, participants remembered few incongruent but a lot of the congruent stereotypes for the out-group. The opposite occurred for the in-group where they remembered few of the congruent but a lot of the incongruent stereotypes. This supports previous research stating that we perceive members of the in-group differently than the out-group (Duncan, 1976). When encountering an out-group member, we rely on our stereotypes, treating them as a homogeneous group and as their features conforms the stereotype this is what we tend to remember. For the in-group on the other hand, we remember less congruent stereotypes because we are aware of the fact that our own social

group is not homogeneous, resulting in a more elaborative processing of individuating features. Hence, what we do remember are the features that violates our belief about the in-group, the individuating features that makes those persons stand out from our in-group which are the incongruent features. For example, a common stereotype about the out-group Arabs is that they all are religious, when describing an Arab as religious, this is very easy to remember. When describing a member of the in-group Swede with the same word, it will also be easy to remember, yet not for the same reason. An in-group member will remember this because it violates the expectancy and as we do not rely on stereotypes considering our own group, this incongruent feature will be processed more elaboratively.

Consider then the first list, where participants were asked to forget all the features. Although the same pattern occurs as in the second list, it was shown that participants remembered significantly less stereotype-congruent features for the out-group in the first list than in the second list.

Participants remembered less incongruent features for the out-group in the first list than in the second list, but there was no significant difference. Overall, the DF-test was very demanding. As mentioned earlier, each list contained of forty features, ten of those were incongruent features connected to an out-group person. A possible explanation for the non-significant result is that the demanding task left no cognitive resources to elaboratively process the incongruent features for the out-group, as there was a very low recall rate for those in both lists that resulted in a non-significant difference. This might seem contradictive to the model *encoding flexibility of stereotypes* mentioned above (Sherman et al., 1998). The model states that when cognitive capacity is constrained, we tend focus more on stereotype-incongruent features. If the DF-test in this study was cognitively demanding, how come participants remembered more of the stereotype-congruent features? Even if the task was cognitively demanding, participants in this study could use all of their cognitive resources to focus on the persons they were presented to and the features connected to them. Studies that have concluded that we tend to rely on stereotypes when cognitive capacity is constrained have forced participants to divide their attention. For instance, counting vowels at the same time as trying to form an opinion about a person with both stereotype-congruent and –incongruent features (Macrae et. al, 1997). The test in the present study was cognitively demanding, yet the cognitive capacity was not constrained. The results suggest that when participants encountered a member of the out-group, they relied on their stereotypes.

For the in-group features, participants remembered less incongruent features in the first list compared to the second list. They also remembered significantly less stereotype-congruent than incongruent features in both lists. As mentioned earlier, when we encounter a member of the in-group we do not rely on our stereotypes, but since the person is a member of our own group and we tend to favor our in-group, if the person is connected to features not compatible for ourselves this information stands out for this person. As we see the out-group as *them*, their congruent features are not what are congruent for *us*, the in-group. As explained earlier, in this test, the stereotype-congruent features for the out-group was used as stereotype-incongruent for the in-group. Thus, for the in-group an expectancy of violation occurs. The information that is not congruent for our in-group "pops out" and therefore becomes more memorable. For example, in 2011 a terror attack took place in Norway, where Anders Breivik killed several people because of their beliefs and opinions. This attack got a huge amount of attention all over the world, obviously because it was a horrible action but also because it violated our expectancies of our in-group. Terror attacks are something *they* do, not something *we* do. This attack to us should then be more memorable then if it was executed anywhere else in the world by a person belonging to a social group we already consider is capable of such an act.

A problem that can occur when using DF as a test is proactive interference. As discussed above, this can be an issue when the participant cannot disregard the first list items as they are strongly linked to the stereotype. In this test it was controlled for by counterbalancing the lists to make sure that one list did not contain stronger stereotypes than the other, after half of the participants were tested, the first and second list changed in order. As shown in the results, participants was able to disregard the first list, hence, no proactive interference seems to have emerged.

One factor that was thought affect the DF-test was WMC, measured by an OSPAN. The hypothesis stated that those who attained a high score on the OSPAN task would easier be able to inhibit the first list, a lower recall, compared to those with a low score. Previous research has shown that WMC does have an effect on DF-tests, although this study was not able to show any significance. As the results shows, the first list was forgotten meaning participants were able to suppress irrelevant information during that part of the test independent of their ability to sort out irrelevant information in the OSPAN task. This might be due to several differences in design between this and former studies. For one thing,

previous studies have used a DF-test containing words with no load and the words have not been connected to a person. The nature of those tests has been closer to a typical OSPAN. Another factor could be the difficulty of the present study. One of the previous studies have been able to show a difference in enhancement of the second list but not a lower recall in the first list. Hence, the WMC in that study did not show a greater inhibition but a greater enhancement. Due to the design, the present study contained a lot more words than a typical DF-test and participants also had to connect the items to a specific person. Participants with high WMC in this study might have showed an enhancement of recall for the second list but were not able to connected to the right person and therefore it was not presented in the test as a correct answer. Another possible explanation is that the present study was very demanding and the OSPAN was the last part. Participants who were very focused during the DF-test and were able to forget the first list might have performed poorly on the OSPAN task due to fatigue. This might have happened the other way around as well, where participants who did not put in a great amount of effort during the DF-test were not as tired during the OSPAN task and therefore performed better. The relation of WMC and intentionally forgetting should be further investigated.

As the non-significant result between WMC and performance on the DF-test can be due to the design of the test, the results of this DF-test can be considered as a support for the *retrieval inhibition account* (Bjork, 1989). The features learnt in the first list, both congruent and incongruent, was most likely intentionally forgotten as an effect of inhibition when given the instruction to forget as the participants recalled less from the first list. To further investigate this, the results should be compared to a group performing the same test but without an instruction to forget.

The features was most likely rehearsed by the participants as they were viewing the features both on screen and also on a paper where they were supposed to rank how well they thought each feature fitted the person. Each slide of every person was also seen for two minutes, a substantial amount of time to rehearse. The *context change account* (Sahakyan and Kelley, 2002) cannot explain the results as it states that a difference of internal context should be due to the DF-effect. In the present study, participants were exposed to the four (from both the in- and out-group) same persons in the two lists and there should be no difference in internal context, yet the DF-effect emerged.

The participants were divided into two groups, high and low, based on their results from the IAT. Participants in the high group were those who had a stronger preference to Swedes compared to Arabs, whereas the low group showed a weaker preference for Swedes. The hypothesis stated that those in the high group should have remembered more of the stereotype-congruent features for the out-group and stereotype-incongruent features for the in-group compared to the low group. In the analysis, this result was not attained. The IAT was added to the test based on the assumption that a stronger preference for the in-group would also measure a sense of prejudice and a stronger activation of the stereotype for the out-group. As mentioned before, IAT has met some criticism. What might have been measured in this IAT are a cultural belief rather than a personal belief and also a result affected by familiarity. Even if we are aware of the stereotype concerning the out-group because of the cultural we live in and thus express them in an IAT, this does not have to be our personal beliefs. Also, the IAT might produce a result show a result based on a preference but a reflection of familiarity. As we spend more time with the in-group this group also comes off as the group we prefer. Participants in this study was clearly aware of the stereotypes for the out-groups as they found them a lot more easy to remember, yet knowing the stereotype does not equal believing in the stereotype.

Social application

The results of this study demonstrate a difference in recollection of features depending on whether the person belongs to the in- or out-group. It also demonstrates a possibility to intentionally forget stereotypes. Although, one should be careful about drawing conclusions of reduction of stereotypes by trying to forget them. Even if we are temporarily able to inhibit or block the routes to them, they are still available in memory. Note also, that it is not my intention to alter the attitudes of the participants. Stereotypes are a solid form of categorization and can be adaptive in many ways. What is more of interest is the difference in how we encounter members of out-groups compared to our in-group. When attaining this knowledge and at the same time realize that it is possible to inhibit our stereotypes, we can be more aware of how we act and encounter people. For instance, at job interviews or in trials, where we are supposed to encounter people in an equal way. Not to mention, in every day life.

Future research

Many interesting questions have been answered in this study, but still, there are questions worth further investigation. As already mentioned, the test was very demanding and this might be one of the reasons why no significance was shown when the scores of the OSPAN task was added as a between factor. The reason the test was made so demanding was that there were eight different outcomes in the DF-test (stereotype-congruent or –incongruent features for the out-group in either list 1 or 2 and stereotype-congruent or incongruent features for the in-group in list 1 or 2). To get a good measurement of each outcome, ten measurements were wanted for each outcome. For future research the test could be made less demanding by dividing the participants into two groups, where one group performs the test only considering the in-group and the other group considering the out-group. This design would require more participants, which was not possible in this study due to the time limit.

As already noted in the discussion, more research needs to be done considering WMC and DF. The different results attained so far are not yet satisfying when it comes to understand the underlying mechanisms of intentional forgetting.

References

Araya, T., Akrami, N., & Ekehammar, B. (2003). Forgetting congruent and incongruent stereotypical information. *The Journal of Social Psychology, 143,* 433-449

Aslan, A., Zellner, M., & Bäuml, K-H. T. (2010). Working memory capacity predicts listwise directed forgetting in adults and children. *Memory, 18,* 442-450

Brambilla, M., Ravenna, M., & Hewstone, M. (2011). Changing stereotype content through mental imagery: Imagining intergroup contact promotes stereotype change. *Group Processes and Intergroup Relations, 15,* 305-315

Coutant, D., Worchel, S., Bar-Tal, D., & van Raalten, J. (2011). A multidimensional examination of the "Stereotype" concept: A developmental approach. *International Journal of Intercultural Relations, 35,* 92-110

Conway, M. A., Harries, K., Noyes, J., Racsma'ny, M., & Frankish, C. R. (2000). The Disruption and Dissolution of Directed Forgetting: Inhibitory Control of Memory. *Journal of Memory and Language, 45,* 409-430

Conway, R. A. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and and user´s guide. *Psychonomic Bulletin & Review, 12,* 769-786

Delaney, P. F., & Sahakyan, L. (2007). Unexpected costs of high working memory capacity following directed forgetting and contextual change manipulations. *Memory & Cognition, 35,* 1074-1082.

Duncan, B.L. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of Personality and Social Psychology*, *34* , 590–598.

Gilovich, T., Keltner, D., & Nisbett, R. E. (2006). *Social Psychology.* New York: Ww Norton & co

Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464−1480

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197-216

Freud, S. (1938). *Basic writings.* New York: Modern Library

MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 271-279

MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review, 6,* 123–129.

Macrae, N. C., Bodenhausen, G. V., Milne, A . B., Ford, R . L. (1997) On the regulation of recollection: The intentional forgetting of stereotypes. *Journal of Personality and Social Psychology, 72,* 709-719

Redick, T. S., Heitz, R. P., & Engle, R. W. (2007). Working memory capacity and inhibition: Cognitive and social consequences and social consequences. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (pp. 125-142). Washington, DC: American Psychological Association

Rudman, L. A., Greenwald, A. G., Mellott, D. S., & Schwartz, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the implicit association test. *Social Cognition, 17,* 437-465.

Sahakyan, L. & Goodmon, L. (2007). The influence of directional associations on directed forgetting and interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 1035-1049

Sahakyan, L & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning Memory and Cognition, 28,* 1064-1072

Sherman, J .W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70,* 1126-1141

Sherman, J. W., Lee, A. L., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype efficiency reconsidered: Encoding flexibility under cognitive load. *Journal of Personality and Social Psychology, 75,* 589-606

Soriano, M. F., & Bajo, M. T. (2007). Working memory resources and interference in directed forgetting. *Psychológica, 28,* 63-85.

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Working memory capacity and retrieval limitations from long-term memory: An examination of differences in accessibility. *The Quarterly Journal of Experimental Psychology, 65,* 2397-2410

Wegner, D. M., & Schneider, D. J. (1989). Mental control: The war of the ghosts in the machine. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended Thought* (pp. 287-305). New York: Guilford Press.