



LUNDS
UNIVERSITET

Institutionen för psykologi

Forget your stereotypes: Memory inhibition as a tool to change attitudes

Alexander Holmbäck
Helena Lundström

Kandidatuppsats ht 2010

Handledare: Mikael Johansson

Abstract

In this study we investigated if stereotype conceptions against groups of people can be suppressed by retrieval practice and thereby affect implicit associations. Previous research of retrieval-induced forgetting has shown that interfering memories are inhibited and weakened by retrieval practice. Our hypothesis was that retrieval practice of features incongruent to a stereotype would cause memory inhibition of congruent interfering features and also a reduction of implicit associations. To induce memory suppression, we applied the retrieval-practice paradigm, the stimuli used for retrieval practice was groups of people as categories and stereotype features as items. To measure implicit associations, an Implicit Association Test was used. Our results showed that retrieval practice with social stimuli successfully induced a significant memory impairment but not a significant reduction of implicit associations. Based on the result, we discussed the possibility that retrieval of positive features could be a more effective method to reduce stereotypes than suppression of negative features. For future research within retrieval practice, we suggested an investigation of how pre-understanding of a stereotype could affect memory impairment. And for implicit associations, how susceptibility to reduction of implicit associations would differ between stereotypes or depend on a successful memory impairment.

Keywords: Stereotypes, Implicit Associations, Retrieval-practice, Retrieval-induce forgetting, memory inhibition.

Index

Introduction.....	4
Retrieval practice paradigm.....	6
Theories of Retrieval-induced forgetting.....	7
Stereotypes and implicit associations.....	11
Implicit Association Test.....	12
Retrieval practice and stereotypes.....	13
The present study.....	15
Method.....	16
Participants.....	16
Material.....	16
Design.....	18
Procedure.....	19
Results.....	20
RIF.....	20
IAT.....	21
Discussion.....	23
Social applications.....	27
Future research.....	27
References.....	29
Appendix A – Flowchart.....	31
Legend.....	31
Appendix B – Material.....	33
Legend.....	33

Introduction

Stereotyping, to assign qualities toward individuals based solely on their group belonging, is a contradictory phenomenon. The mechanism behind stereotyping, categorization, is essential for all kinds of learning. Still, side effects of stereotyping causes conflicts among groups of people and individuals as it influences our perception and aids to sustain false conceptions about the Other. Those side effects however are not inevitable. Even if sole individuals are most likely to generalize features among group identities based on her or his own experiences, the damage of stereotyping will never be substantial, for both individuals and society, unless the generalization is not shared with larger groups of people. When a stereotype is collective and shared among members of a culture, it can be maintained without the need of actual experience and is consequently in danger zone of confirming itself: Conception influences perception which in turn confirms the conception (Mason, Tatkow & Macrae, 2005). The consequences of collective stereotyping are often fatal for target groups and incompatible with the treaty of human rights stating every human's equal value. Hence, we hold that finding psychological means to identify and neutralize negative stereotyping is of high importance.

Within social psychology, several methods have been developed to understand the underlying mechanisms behind stereotype behavior as well as how to reduce them. What all methods have in common is that they focus on creating positive experiences of the group that is target of negative stereotypes. Examples of such methods are *intergroup contact* where different groups of people are working towards a common goal, or exposure to people from the out-group that contradicts the negative stereotype (Marx et al., 2005; McIntyre et al, 2003; Allport, 1954). The logic behind these strategies can be explained in terms of cognition: In situations like these individuals are forced to retrieve and attend to positive information and suppress negative misconceptions about each other, thereby rendering the stereotype schema obsolete and maladaptive. In the end, reducing stereotypes can consequently be seen as a cognitive task related to memory (Dunn & Spellman, 2003).

Recent research shows that forgetting is an active cognitive process that occurs when several similar memories compete for retrieval. When a stimuli activates several memories, some memories will be stronger associated to the stimuli than others, but the strongest memory does not necessarily have to be the most context appropriate. By recruiting inhibitory control mechanisms, the most context dependent memory can be retrieved while stronger but less context appropriate memories are inhibited (Anderson, 2003). For example, when learning a friend's new phone number

whom you call frequently, at first the old number will interfere when you try to dial the new number. But over time, the old number stops interfering and is ultimately no longer possible to retrieve. Not because time has passed, but because repeated inhibition has weakened the traces to the old number in order to aid retrieval of the new number. That memory traces are weakened by repeated inhibition is thought to apply on all kinds of memories, including those founding stereotypes. If an individual is forced to repeatedly retrieve positive and incongruent information to a group that is target to negative stereotypes, will information congruent to the stereotype be repeatedly inhibited and thereby weakened, just as the old phone number?

The research of memory inhibition caused by retrieval practice is an ongoing topic called *retrieval-induced forgetting*. The effect of retrieval-induced forgetting can be demonstrated by the retrieval practice paradigm developed by M.C. Anderson (1994).

As a complex and elusive concept, stereotyping comes in different flavors. A common way to classify stereotypes is to divide them into explicit and implicit, where explicit stereotypes are expressed and implicit hidden from the public view (Fazio, Jackson, Dunton & Williams, 1995). When explicit stereotypes can be assessed simply by questionnaires, more sophisticated methods apply to assess implicit stereotypes. The implicit measure is relevant when people are either unwilling or unable to report stereotype thinking. A widespread method to measure implicit stereotypes is called *Implicit Association Test* and measures a tendency to associate a target to an attribute. Stereotype thinking is usually detected by testing the tendency to associate a group of people to something negative rather than something positive (Greenwald et al., 2002). Given that retrieval practice of positive features causes an inhibition of negative features, could this inhibition reduce the negative implicit associations?

Previous research of retrieval-induced forgetting has applied simple and conventional semantic categories to test memory performance. Some studies have successfully shown a memory inhibition of stereotype related information associated to individuals (Brazel & Ringqvist, 2009; Dunn & Spellman, 2003). This study aims to investigate if memory inhibition of information based on pre-existing stereotypes that are associated to groups of people. When retrieving information that is incongruent to a stereotype, interference occurs with information congruent to the pre-existing stereotype. The purpose of this study is to investigate if this leads to 1) inhibition of existing explicit memories that associate stereotype features with groups, and 2) if this inhibition of explicit memories has potential to affect attitudes towards groups based on implicit memory.

To perform this investigation, we have designed an experiment that implements Implicit

Association Tests to measure changes in implicit associations and Retrieval Practice Paradigm to induce memory impairment. We will begin with the basics of these two paradigms and then explain how we intend to combine them in order to perform our experiment.

Retrieval practice paradigm

Retrieval-induced forgetting or short, RIF, refers to the memory impairment that occurs when some members of a category are subject to rehearsal and others are not. Retrieval-induced forgetting can be shown using the so called Retrieval-practice paradigm created by Anderson, Bjork & Bjork (1994). The experiment is divided into three phases; phase one is called *Study phase*, the second is called *Practice phase* and the third *Test phase*.

During the study phase, participants are asked to study and memorize a list of words that are organized in a category-exemplar manner (e.g. Fruit_Banana, Fruit_Kiwi, Animal_Cow). Several categories are used (fruits, animals etc) and every category contains several exemplars.

In the practice phase, some of the words are repeated using retrieval practice. The participants are asked to recall an exemplar based on the category and one or more letters of the word as a cue (e.g. Fruit_K for Kiwi). From one of the categories, only some exemplars occur (e.g. for the fruit category, Fruit_Kiwi is present but not Fruit_Banana), from another category none of the exemplars from the study phase are present. Thus, the words can be divided in three groups based on their role during practice phase. *Practiced words - Practiced category*, which refers to words that occurred during practice. *Unpracticed words - Practiced category*, which refers to words that did not occur during practice but belongs to a category with other exemplars that did occur. *Unpracticed words - Unpracticed category*, which refers to words that did not occur during the practice phase because it belongs to a category where no exemplars of the category were tested.

Between practice and test phase there is some kind of distraction task, usually a problem-solving task that does not require reading or understanding of words or letters. In the test phase, memory performance of all words from study phase is tested. The test approach can be either recall or recognition depending on research question. A recall approach implies that participants are asked to recall as many words as possible within a given a category, or that participants are tested on each word separately given the same cue as in the practice phase (Fruit_B for Fruit_Banana). A recognition approach implies that the participants are asked to select which of several words that occurred during study phase.

Retrieval-induced forgetting is demonstrated by comparing the memory performance between *Unpracticed words - Practiced category* and *Unpracticed words - Unpracticed category*.

The recall performance of words that were left out from a category where other words were practiced, are generally lower compared to words from the category where the participant did not practice any word. Different theories are used to explain this phenomenon.

Theories of Retrieval-induced forgetting

Even though the effect of retrieval-induced forgetting has been established for half a decade, the theories to explain the phenomena has shifted over time. According to the classical *response competition theory of independence* (also known as *blocking theory*) developed by McGeoch (1936), recalling one item will block access to related items. During retrieval, different memories associated with the same cue are competing to be recalled when this cue is present. When strengthening a target in the practice phase from a category, the recall of competitors from the same category will be blocked. According to this theory, the impaired recall of “Banana” is caused by strengthening “Apple”, the strength of “Banana” will remain unchanged but blocked. Further on, by retrieving the same item several times from a category based on a cue, a learning process will take place and improve the recall of the item.

Another theory, *associative unlearning*, explains the effect of retrieval-induced forgetting by positing that the practice phase weakens the connection between category and unpracticed item, at the same time as it strengthens the connection between category and practiced item. According to this theory, the representation of the item stays the same but we unlearn, or weaken, the connection between the unpracticed item and category when retrieving a practiced item (Melton & Irwin, 1940).

The two theories above are based on three assumptions: the competition assumption, the strength-dependence assumption and the retrieval-based learning assumption. The competition-based assumption posits that when presented with a cue, different memories that are associated with this cue will compete for recall and conscious awareness. The strength-dependence assumption implies that the ability to recall an item is dependent on competing items related to the same cue. Stronger association to competing items decreases the ability to recall the correct item. According to the retrieval-based learning assumption, retrieving is a learning event because it enhances subsequent recall of the retrieved item (Anderson, Bjork & Bjork, 1994).

The blocking and associative unlearning theories are now challenged by the theory of inhibition. The theory also involves all the assumptions mentioned above, but posits that forgetting is a result of interference caused by the competing memories, rather than being a passive side effect of our memory. When memories are competing for conscious awareness, inhibitory mechanisms are

recruited to retrieve the most appropriate memory. When provided with a cue, this given cue will probably be associated with several memories, some more strongly linked than others. The inhibitory control helps us to retrieve a weaker, yet more relevant target by inhibiting the stronger competitors. The competing memories will be weakened by the inhibition and this causes forgetting. In this case, the competing item “Banana” is inhibited in order to facilitate the retrieval of the target “Apple”. We will clarify some relevant properties of retrieval-induced forgetting that support the theory of inhibition.

Cue-independence. A property supporting that retrieval-induced forgetting affects the semantic encoding is that it appears to be cue-independent. This means that the impairment of competitors has a tendency to generalize to novel test cues that were not present during retrieval practice. This finding shows that impairment is caused by suppression of the competing memory itself, not by damaging the association between the cue and item. For example, the inhibition of the item “Banana” by practicing “Fruit_Apple”, can be measured when testing retrieval with another cue, “Food_B” instead of “Fruit_B”.

Retrieval specificity. According to the blocking theory strengthening one item will impair the recall of related items. This means that repeatedly studying items should impair the memory for competitors, but this is not the case. Several results from studies has shown that extra study exposure of target has failed to impair the recall of related competitors (Andersson & Bell, 2001). Instead, findings suggest that inhibition takes place to override interference from competing items while selectively retrieving the target item.

Interference dependence. Retrieval appears to be necessary but not sufficient to induce memory impairment. Inhibition occurs with memories that are interfering with the target item, so if there are no items interfering then no items are inhibited. Ergo, the amount of interference caused by a related item affects retrieval-induced forgetting. Anderson, Bjork & Bjork (1994) found that retrieval practice did not always cause impairment. They found that an important variable is whether the category exemplars are high or low in taxonomic frequency. Retrieval-induced forgetting occurs only if the unpracticed competing category items are high in taxonomic frequency, but not the other way around. For instance, it is more likely that practicing the untypical fruit kiwi causes an impaired recall of the typical fruit banana than vice versa. What moderates the strength of inhibition is the typicality of the competitors. A stronger association between a category and an unpracticed competitor will lead to more impairment.

Strength-independence. Early work on retrieval-induced forgetting was partly based on the

assumption that strengthening one item would impair the recall of other items associated to the same cue. Although, it appears that practiced items can be strengthened without causing impairment, even when interference dependence between the items is present. Hence, strength alone cannot predict forgetting. What seems to be the critical variable is the competitiveness of the unpracticed item, if the unpracticed items are potent competitors, a stronger inhibition will occur. (Anderson, 2003).

Moderating and masking factors. There are some factors that can either moderate or mask the effect of retrieval-induced forgetting. Moderating factors are those that influence the extent of memory impairment that the competitor items suffer from during retrieval practice. Masking factors affects the measurable, but not the actual, memory impairment. Moderating or masking factors might arise during all phases of the paradigm.

Different kinds of connections between items in the experiment seem to either reduce or magnify the inhibition and thereby moderate the effects. *Integration* refers to the inter-connections between different items that share the same category as retrieval cue. For example, when studying items such as coffee, tea, milk and juice from the category *Drinks* the participants can form interconnections between those items. For instance, they might form an interconnection between coffee and tea (hot drinks) or milk and juice (cold drinks). When integration between items occurs, it reduces inhibition during retrieval practice when the integration is formed between target and competitor. The opposite is true when the integration is formed between competitors. The integration can be caused by *associative relatedness* between items as one item has potential to evoke other items by association. The moderation of associative relatedness can be substantial if target is associated with competitor. The associative relatedness is a form of integration that occurs when items are linked to each other because they share a common associate, for example paper and pencil that regardless of their different semantic meaning both associates to stationary. Associative relatedness can also occur on a between-subject level, where integration is due to a more elaborate encoding.

A moderating factor similar to integration is when connections between items are established based on *semantic similarity* of the items. Just as with integration, different results will be obtained whether the similarity is between competitor-competitor similarity or a target-competitor. When there is a competitor-competitor similarity, inhibition disperses as the inhibition of one competitor will enhance the possibility that similar competitors are affected the same way. As with integration, the opposite occurs when there is a target-competitor similarity and the effect

of retrieval-induced forgetting weakens. This is because a target-competitor similarity causes an overlapping which reduces the memory impairment caused by inhibition of the target when the competitor is activated.

A possible masking factor is called *baseline deflation*. The baseline includes control items which are used to measure the impairment. If the baseline items themselves are affected by retrieval practice then the result shows less memory impairment for target than actually occurred and are thereby masked. Baseline deflation is a consequence of some kind of generalization between features in baseline and experiment.

Another masking factor is called *Cue-priming*. When performing retrieval practice the participant is presented with a category name and the first letters of the item they are supposed to retrieve. If successful, retrieval of practiced item is enhanced and the competitors inhibited. During this phase, the participant sees the same category name several times when practicing different items from the category. This can lead to an overall greater accessibility of the practiced category and facilitate retrieval of the associated items of the category (Andersson, 2003).

Duration of retrieval-induced forgetting. The findings considering the duration of retrieval-induced forgetting has been ambiguous. Some results suggest that the memory impairment caused by retrieval practice is temporary, and some that it has a more long lasting or even permanent effect. MacLeod and Macrae (2001) provided a study with evidence showing that the forgetting caused by retrieval practice recovers over time. Their participants were tested on recall either directly after the retrieval practice or 24 hours later. They detected retrieval-induced forgetting for the participants that had made the test immediately after practice, but not for those who made the test the following day. Although, there are some findings suggesting a more lasting effect. It appears that the effect of retrieval-induced forgetting can diminish over time under some circumstances but not always. For instance, remember the example about trying to recall a friend's new phone number. If practicing retrieval frequently, after a while the old number will stop interfering and be forgotten in advantage for the new number. It is unusual that the old number recovers from the inhibition and starts to interfere again, although it is possible. This can be explained in terms of interference. To the extent a memory continues to interfere, it also has to be inhibited: Repeated interference leads to repeated inhibition that results in impaired accessibility. It therefore appears that the duration of retrieval-induced forgetting depends on the circumstances. At this moment there is not enough known about whether inhibition recovers over time and more research should be done before we can make any predictions (Andersson, 2003).

Stereotypes and implicit associations

Maybe you believe that all Asians are intelligent or that every Irishman loves beer? Whether the stereotype is true or false, positive or negative, they help us categorize the people we meet in an easy and fast way. To find out about another person's beliefs, we can simply ask them. Although, the person we ask can, of course, choose how to answer the question. Sometimes we tend to overcompensate for our beliefs even to the point where we state the direct opposite, as noted in a study by Norton, Vandello & Darley 2004. The participants in the study hid their beliefs to appear less prejudiced. So, asking a person about their attitudes towards a group of people might not result in a trustworthy answer. There is an existing conflict between what we feel and what we are prepared to reveal publicly or even to ourselves.

As mentioned earlier, stereotypes can be divided into either explicit or implicit stereotypes for methodological reasons. During the past years, the research field concerning attitudes has shifted from examining explicit attitudes to a focus on implicit attitudes. The distinction between explicit and implicit processes is still not clarified and contemporary models are insufficient when it comes to integrating all of the available evidence. Research has shown that implicit stereotypes affect explicit behavior. Results reveal that implicit attitudes predicted a subtle and spontaneous behavior towards another person, such as seating distance or the amount of eye contact. Explicit attitudes on the other hand, did not predict behavior. (Rydell & McConnell, 2006).

As pointed out by Greenwald, McGhee & Schwartz (1998), implicit attitudes derive from past experiences that are not necessarily consciously remembered. Hence, implicit attitudes are unintentional and may not be available to our consciousness, therefore they cannot be measured by self-report. Implicit attitudes are thought to be automatic and unintentional since they are activated faster than conscious activity would be. The activation is triggered by a subliminal stimulus, not noted by the person.

Since implicit attitudes are automatic evaluations, triggered non-consciously and without the person's control, they are difficult to change (Rydell & McConnell, 2006). Although, research has shown that the implicit attitudes can be moderated by some factors. Barden, Maddux, Petty & Brewer (2004) showed that depending on the context, the implicit attitudes towards black people and Asians changed. The participants favored Asians when they were assigned a student context, but when the assigned context changed to a basketball court black people were favored. Still, changing implicit attitudes is an area of the literature that is largely unexplained. To capture implicit attitudes different methods have been developed. One common and widely used test is the Implicit

Association Test, or short, IAT.

Implicit Association Test

Even though Implicit Association Tests can be used to measure virtually all kinds of tendencies to automatic associations, its initial purpose was to explore and understand stereotyped behavior such as prejudices. The Implicit Association Test was developed by Greenwald, McGhee & Schwartz (1998).

IAT is a sorting task executed on a computer and the participants' job is to sort stimuli into categories according to a certain rule. The rule is to press a key on the left side whenever there is a stimulus that belongs to a category located on the left side of the screen, and a key on the right side when the stimulus belongs to a category located on the right side of the screen. The stimuli are divided into attributes and targets, the attribute stimulus has an evaluative representation (e.g. positive or negative words), the target stimulus consists of words or images that represents two groups of people (e.g. old or young).

The stimuli are presented by turn in the middle of the screen and the reaction times for every response is measured. The evaluative attribute stimuli are combined with the attitude target stimuli first in one way and then reversed. For example, the attribute category “Good” is first located in the same corner as the target category “Old” and then moved to the category “Young”. By combining attributes with target, two reaction tests are performed where the first test uses one combination of attribute and target (e.g. good words with old people and bad words with young people) and the second test the other combination (e.g. bad words with old people and good words with young people). The result is assessed by comparing the reaction times between the two combinations. If there is a faster mean reaction time when the subject use the same key to categorize negative words and pictures of old people compared to positive words, an automatic negative association for old people can be assumed. A slower reaction should be expected when the targets and attributes are incongruent, for instance when the respondents are meant to use the same key for positive words as for pictures of people they don't have a preference for.

IAT is presumed to measure automatically activated associations. By relying on automatic processes IAT is often credited for being resistant to faking. IAT is also thought to have a high predictive validity, which significantly exceeds self-report measures (Greenwald, Poehlman, Uhlmann & Banaji, 2009). In another study, the subjects' brain activity were registered while watching pictures of black and white faces. The results were correlated with their brain activity during an IAT and a *Modern Racism Scale* (an explicit measure), where a differential activation in

amygdala was detected. This activation was not present during the explicit test. This gives neurological support that IAT measures automatic associations related to social and emotional evaluation (Phelps et al., 2000).

IAT is widely used and accepted but there is some criticism against the test and an ongoing debate considering its validity. One question debated is whether the test truly captures the individual's evaluations or if it reflects the evaluations of the society. Some argue that the test results are a cultural expression rather than a personal. For instance, in an IAT study by Greenwald, McGhee & Schwartzman (1998), the participants were either Japanese Americans or Korean Americans, as expected the Japanese Americans showed a preference for Japanese names and the Koreans a preference for Korean names. But the results also showed that the effect was dependent on the level of acculturation in the Asian culture. The Japanese Americans showed a greater bias for Japanese names if they were more acculturated in Asian culture than other Japanese Americans, similar results were shown for the Korean Americans. It seems like the more a person is exposed to a culture, the more we adopt the culture's beliefs, including stereotypes. Considering this result, IAT may reflect cultural stereotypes rather than individual. Another critique against IAT is that the results are affected by the subject's familiarity with the attitude objects. The debate considers the fact that in-group objects are more familiar than out-group objects and might jeopardize the validity of the test. Hence, the test might not measure what we like/dislike but rather what is familiar to us. This was tested by Rudman, Greenwald, Mellott & Schwartz (1999) and the experiment showed that the in-groups were favored among Christians and Jews despite the fact that the groups were matched on population frequency. Despite the criticism, IAT has nonetheless proven to be a powerful method to assess automatic preferences.

Retrieval practice and stereotypes

Some exemplars are more strongly linked to a certain category; for example, it is easy to agree that a banana is an exemplar of the category fruit. The strongly linked exemplars are those that easily come to mind when presented with the category as a cue. A banana or an orange would for most people be considered as typical exemplars of the category fruit. Other fruits as perhaps kiwi or guava are not as strongly linked and therefore atypical for the category, although we know they are fruits they don't come to mind as easily. Due to interference dependence, strongly linked exemplars are more strongly inhibited than the atypical, weakly associated ones. In order to successfully recall an item from a category, the strongly associated items that belong to the same category must be strongly inhibited due to the fact that they are more likely to come to mind

(Bäuml, 1998). But when using the retrieval-practice paradigm with social stimuli the relationship between category and item becomes different. We all know that all bananas are fruits, what is not as certain is that all good basketball players are black or that all people with a bad memory are old. Even if these beliefs are common, we cannot be sure that everyone would agree with this belief or that such a relationship between category and feature would be sufficient to cause interference and thereby inhibition. This is due to the fact that every feature more or less fits into all groups of people. Even if one feature is commonly associated with a group of people, it could almost equally be applied to another group. For instance, even if we commonly associate the feature fanatic with Muslims it does not prevent us from associating it with Swedes as well. Still, interference dependence could be acquired by strong submission to a stereotype. If the subject believes in a stereotype, a sufficient association between the group of people (e. g. category) and the stereotype feature (exemplar) should arise (Dunn & Spellman, 2003). If this is true then these strongly associated stereotype features would suffer more forgetting than others.

To investigate if social stimuli would suffer from a retrieval-induced forgetting effect, Macrae and MacLeod (1999) used features as exemplars and individuals as categories, and found evidence for an effect. The participants were presented with ten different features for each of two individuals, Bill and John. After the study phase, participants performed retrieval practice and rehearsed half of the features for either Bill or John. Macrae and MacLeod were able to demonstrate a result consistent with previous research, participants' memory for the unpracticed traits of the person whose other traits had been practiced. This result demonstrates that retrieval-induced forgetting does occur even with social stimuli.

An important moderating factor concerning retrieval practice and social stimuli is discussed by Dunn and Spellman (2003) and is referred to as *cross-item integration*. A cross-item integration occurs when one feature evokes thoughts of other associated features. Features that normally do not seem to have a connection, like cheap and wise, might be associated with each other when providing a stereotype label such as old. This integration can potentially reduce the inhibition during retrieval practice when occurring between features within categories. Dunn and Spellman offers an explanation of this effect similar to associative relatedness, called mediated retrieval. If believing in a stereotype for a category of people, a strengthened link between the features within the category will occur. When trying to inhibit some of these features through retrieval practice, stereotype believers will suffer less impairment as inhibited features has co-occurred with practiced features and thus becomes activated through mediated retrieval. Note however that using

stereotypes per se is not sufficient to cause cross-item integration, but rather the fact that there is an inter-feature connection present that may prevent forgetting.

As mentioned above, an individual's stereotype thinking isn't always as straight forward as it may seem. An individual can believe in a stereotype without being willing or able to report it, or in some ways accepting or acting out parts of a stereotype but not others. We do not intend to assess if or how much people actually believe in stereotypes, but only rely on the fact that they have knowledge about them. Assuming that the association between categories and features are established by having knowledge about the stereotype, not by submitting to them.

The present study

Our first intention is to contribute to the existing literature about retrieval-induced forgetting by investigating an unexplored relation between category and item, namely stereotypes, an arbitrary relation based on social and cultural conceptions. This will carry the concept of retrieval-induced forgetting closer to the domain of social psychology, which gives opportunities to use new means to investigate the consequences of memory inhibition.

This brings us to our second intention, which is to examine whether memory impairment of specific features that is caused by retrieval practice has potential to affect implicit associations. This is based on the assumption that implicit associations are preceded by repeated exposure of explicit information. Information that in one way or another connects a category with something negative, i.e. information on the news about terrorist attacks that associates the group Muslims with violence and fear. Or the show “Biggest loser” that strongly suggests overweight as something undesirable, reinforcing a common view that fat people represent something unwanted. If the negative associations comes from explicit information, retrieval practice could be used inhibit the negative information on behalf of positive information and consequently reduce the result on an IAT.

Even if retrieval practice is a method to simulate an explicit learning process, the memory inhibition of the features is an automatic and thereby implicit measure. In this study it is therefore relevant to distinguish between specific and general implicit associations. The specific refers to the association between a group of people and specific features. The general implicit associations refer to the tendency to associate a group of people to something negative or positive. From a stereotype perspective, the retrieval practice is a way to activate the stereotype that causes the negative implicit association, and inhibit the same with retrieval practice. This is crucial to our hypothesis because it is the connection between specific features and the negative implicit association that makes us believe that the result on an IAT can be reduced by retrieval practice. What we aim to investigate is:

1) Does retrieval practice cause memory inhibition when using groups of people as categories and features as items when the only connection between groups and features are culturally based stereotypes? 2) Given that there is a measurable implicit preference for one group compare to its opposite, can the implicit preference be heaved by inhibition of non-preferable features and excitation of preferable features related to the non-preferred group? These questions leads to the following two hypothesis: 1) Retrieval practice will cause memory inhibition for the unpracticed items in the practiced category, when using categories and items based on social and cultural conceptions. 2) Implicit associations are preceded with explicit information and can thereby be changed by explicit means. Retrieval practice of incongruent features will thereby reduce the result on an IAT.

Method

Participants

A total of 32 (15 women) individuals participated in this study, with a mean age of 26.31 (min=21, max=52). The participants were recruited from friends and family and took part of the test voluntarily. They all gave informed consent and participated without any reward.

Material

Retrieval practice. A total of 144 features were selected and distributed over 4 domains (36 features in each domain) (Appendix B). The four domains were *ethnicity, age, skin color and weight*. Every domain consisted of opposite pair-wise stereotype categories (with 18 features each), *Muslim/Swede, old/young, dark skinned/light skinned and fat/thin*. The choice of categories aimed at selecting culturally-specific negative associations and was based on a recent meta-study made by Greenwald, Uhlmann, Poehlman & Banaji (2009). *Swede, young, light skinned and thin* are the categories labeled *preferred*, the opposite categories will be labeled *non-preferred* because its relative status as less preferred compared to its opposite. Within each category, half of the features were negative and half of the features were positive (9 positive and 9 negative features for every category).

The positive features for the preferred categories and the negative features for the non-preferred categories makes up an all-embracing group of features that we choose to call *congruent features* because their congruence to the implicit widespread relatively negative or positive association. Likewise, the negative features for the preferred categories and the positive features for the non-preferred categories are called *incongruent features*.

We used our own stereotyped thinking to generate 200 positive and negative features that we thought people in general would easily associate to the a specific category. To determine the strength of the relation between feature and category, a pretest was conducted. 13 participants took part of this a priori survey where they were told to categorize the 200 features into the eight different categories. The instruction for the participants was not to categorize the features based on their own believes, but how they thought most people would agree upon.

Based on the result from the pretest, the features in every category were organized in the following way: Positive features for the preferred categories were strongly related whereas negative features were weakly related. The opposite was true for the non-preferred categories where the positive features were weakly related and the negative strongly related. Consequently, all the features that are congruent to their categories are also strongly related to its category and all incongruent features are weakly related to their categories. The reasons for organizing the words in this manner was: 1) To activate the stereotype that causes the negative implicit association in order to inhibit the stereotype with retrieval practice. 2) To gain use of interference dependence, stating that inhibition of target is strongest when connection between target and category are weak and the connection between competitor and category are strong. Because the congruent features are those that will be excited and the incongruent features are those that will be inhibited, the congruent features should have a strong relation to its categories and the incongruent features should have weak relation to its categories. As congruent and incongruent features do not co-occur and thus do not cause mediated retrieval, this organization should not make the material sensitive to cross-item integration (Appendix B).

IAT. Four different IATs were used, one for each domain. Materials for the evaluative attributes were the same for all IATs and consisted of seven positive and seven negative Swedish words. For the ethnicity domain, seven typical Swedish names and seven common Muslim names were used as target stimuli. For the age domain, seven pictures of young faces and seven pictures of old faces were used as target stimuli, the pictures were black and white and showed the eyes of the faces only. For the skin color domain, 14 animated faces were used as target stimuli, seven of those with the skin tone darkened. For the weight domain, seven pictures of fat people and seven pictures of normal people were shown. The pictures were black and white and the entire face was shown. The material, instructions and design of our IAT was taken from the tests available at the Project Implicit website (<http://implicit.harvard.edu>).

Distraction Task. For the distraction phase, a list of 60 different types of easy problem

solving tasks was used. The problem was either a mathematics expression, or series of characters with one character left out.

Design

The initial layout of the experiment was to present an IAT, followed by a retrieval practice, followed by the same IAT again. This layout would let us observe the difference between the score on the first and the last IAT to see if the retrieval practice has potential to reduce the result. To allow monitoring of practice effects, another IAT was added to serve as a baseline. The only difference between the IAT experiment and IAT baseline was that the two categories in IAT baseline did not undergo retrieval practice.

In the retrieval-practice paradigm, a test was presented for all words in order to measure the effect of retrieval induced forgetting. As the test also was a kind of retrieval practice, there was no way to make sure that the memory impairment actually took place for the domain used in the IAT without jeopardizing a potential effect on the IAT. Therefore, the two remaining domains were used as RIF experiment and RIF baseline. To sum up, the four domains were used as IAT experiment, IAT baseline, RIF experiment or RIF baseline. This design made it possible to investigate the impact that retrieval practice had on the explicit forgetting of features related to certain categories, and also the impact retrieval practice had on the result on an IAT. The domains were rotated between participants in order to counterbalance confounds due to the specificity of the different stereotypes. However, the domains *ethnicity* and *skin color* were always assigned to either RIF experiment and IAT experiment or IAT baseline and RIF baseline at the same time. This to reduce overlapping effects between the two domains because of their similar nature. A consequence was that the domains *age* and *weight* became paired in the same way during rotation (Figure 1).

Rotation of domains				
#	RIF Experiment	RIF Baseline	IAT Experiment	IAT Baseline
1	<i>Ethnicity</i>	<i>Age</i>	<i>Skincolor</i>	<i>Weight</i>
2	<i>Age</i>	<i>Ethnicity</i>	<i>Weight</i>	<i>Skincolor</i>
3	<i>Skincolor</i>	<i>Weight</i>	<i>Ethnicity</i>	<i>Age</i>
4	<i>Weight</i>	<i>Skincolor</i>	<i>Age</i>	<i>Ethnicity</i>

Figure 1. The domains were rotated between participants in order to compensate for domain specific confounders. Ethnicity and skincolor were paired so that they always occurred in the experiment condition at the same time.

The full layout can now be summarized. Two IATs were committed at the beginning and the

end of the test. Between these two IAT sessions, a retrieval practice was used in order to excite incongruent features and inhibit congruent features for the domain that serves as IAT experiment. The reduction of the IAT was assessed by comparing the reduction with a second domain, IAT baseline. A third domain was used as RIF experiment, the purpose of this domain was to measure the forgetting caused by retrieval practice. The forgetting of the features within the domain was assessed by comparing the result with the fourth domain that serves as RIF baseline (Appendix A).

Procedure

As mentioned, the retrieval practice was prepended and appended by two identical IAT sessions. The session started with a message informing the participant that this session will test prejudices. It contained two tests for two domains, the first test was for the domain that served as IAT experiment, the second test was for the domain that served as IAT baseline. Each test was identical in its structure and consisted of seven phases. The first phase was a target learning procedure where the participant during 20 trials learned to place the target stimuli (e.g. Pictures of old and young faces) under the correct category (e.g. Young-Left, Old-Right). The second phase was an attribute learning procedure which was similar to the first phase but with attribute stimuli (e.g. words such as “Flower” or “War”) and attribute categories (e.g. Good-Left, Bad-Right) used instead. The third phase was a paired test learning procedure where both the attribute and target stimuli and categories were used for 20 trials. The fourth phase was a paired test equal to phase three but with 40 trials. The fifth phase was a reversed attribute learning procedure identical to phase two but with swapped sides for the attribute categories. The sixth phase was a reversed paired test learning procedure where the swapped attribute categories were combined with the target stimuli. The seventh phase was identical to the sixth but with 40 trials instead of 20. All stimuli in all phases were randomly selected except one rule in the paired test where the stimuli alternated between attribute and target. After the first IAT session, the participant was forwarded to the next block, retrieval practice. After the second IAT session, the participant was thanked for the contribution and informed that the experiment was over.

The retrieval practice started with a message informing the participant that this session was testing the ability to remember words. The study phase comprised all 144 features from all domains for all participants. During the practice phase, the incongruent features for the domains that served as IAT experiment and RIF experiment were used (36 of the 144 words). In the test phase, all features from the domains that served as RIF experiment and RIF baseline were used which made up 72 of the 144 features. To remember 144 features presented without a brake is an unendurable

task that we didn't want anyone to experience, also we would endanger that the memory performance would be too low. To overcome this issue we split the features in three chunks with one third of the features in each chunk and made three separate complete retrieval practice exercises. The number of features in every category was multiple by three so every chunk was identical except the features used. For every chunk there was a study phase, a practice phase, a distraction task and a test phase. In the study phase, every feature was presented by turn for 4000ms prefixed by its category and a hyphen. Prior every word a fixation cross was presented for 500ms. All stimuli were randomly selected.

In the practice phase, the participants were asked to type the next letter in a feature based on a cue that was made up of as many letters that was required to make the cue unique relative to the other cues in the same session. However a minimum of two letters were always used. Every cue was prefixed by its category and a hyphen and was presented by turn for 4000ms or until the participant responded. Prior every word a fixation cross was presented for 500ms. No feedback was presented. All stimuli were randomly selected.

The distraction task was made up of 20 randomly selected problems where a one-key user response was expected. Every problem was presented for 5000ms or until the participant responded. After response or expiration, a feedback message was presented for 2000ms.

The test phase was performed in the same way as the practice phase except that RIF experiment and RIF baseline categories and features were used. At the end of the first two chunks a message was displayed commending the participants well done and instructed them to forget all words and make place for new words in the next chunk. At the end of the third chunk a similar message was displayed omitting the instruction to make place for new words, the participant was then redirected to the second IAT session.

Results

RIF

When calculating the results of RIF, we compared the recall rate depending on two factors. The category that was subject to retrieval practice was abbreviated with a P. The category P contained two groups of features, practiced (P+) and unpracticed (P-). The control category (C) did not appear in retrieval practice at all but was none the less divided into two groups, C+ and C-, for comparison reasons. We used a mixed ANOVA to compare the recall rate depending on the two factors *practice/unpractice* and *baseline/experiment*. The result demonstrated a significant main effect of practiced (+) and unpracticed (-) features, $F(1,31) = 34.082, p < .001$ and a significant

interaction effect between both factors $F(1,31) = 34.455, p < .001$.

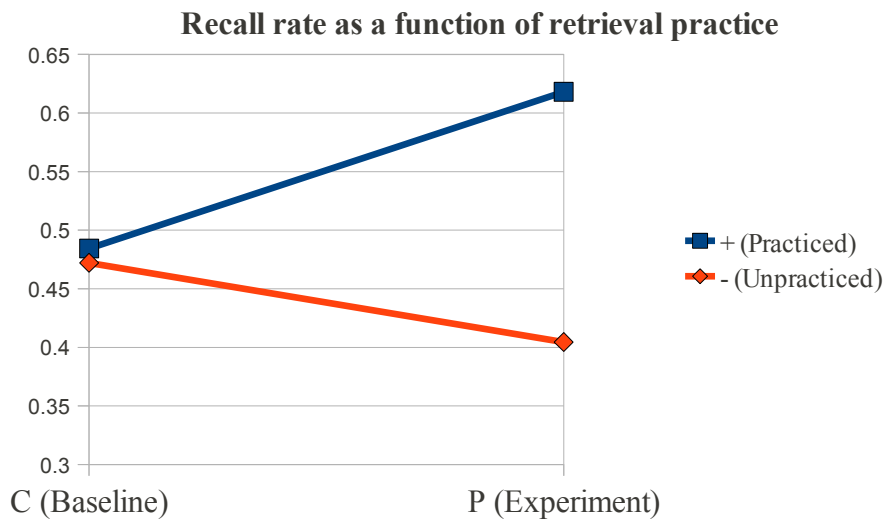


Figure 2. The recall rate as a function of retrieval practice. P- and P+ is unpracticed and practiced features respectively and belongs to a practiced category. C- and C+ are both unpracticed features from an unpracticed category.

In order to examine the effect of retrieval practice, the unpracticed features from the unpracticed category (C-) were compared with the unpracticed features from the practiced category (P-). The mean recall rate for C- was .472 ($SD = .129$) and .404 ($SD = .133$) for P-. A two-tailed paired-samples t-test gave the result $t(31) = 2.310, p = .028$, which showed a statistically significant RIF effect. A corresponding test between C+ and P+ gave a mean recall rate for C+ of .472 ($SD = .129$) and .404 ($SD = .133$) for P+ and $t(31) = -4.308, p < .001$ which confirmed that retrieval practice improved the recall rate (Figure 2).

IAT

To assess the participants' results on the IATs, the reaction times were converted according to the GNB scoring algorithm (Greenwald, Nosek & Banaji, 2003). The participants performed two IATs: the first one served as experiment (IAT-E) and the second served as baseline (IAT-B). Two identical sessions with both IAT-E and IAT-B were committed in the beginning and at the end of the test. The first two tests were abbreviated IAT-E1 and IAT-B1 and the second two tests were abbreviated IAT-E2 and IAT-B2. The result of an ANOVA revealed a significant main effect between baseline (B) and experiment condition (E) $F(1,31) = 3.186, p < .001$, but failed to reveal a significant interaction effect as hypothesized $F(1,31) = 1.136, ns$.

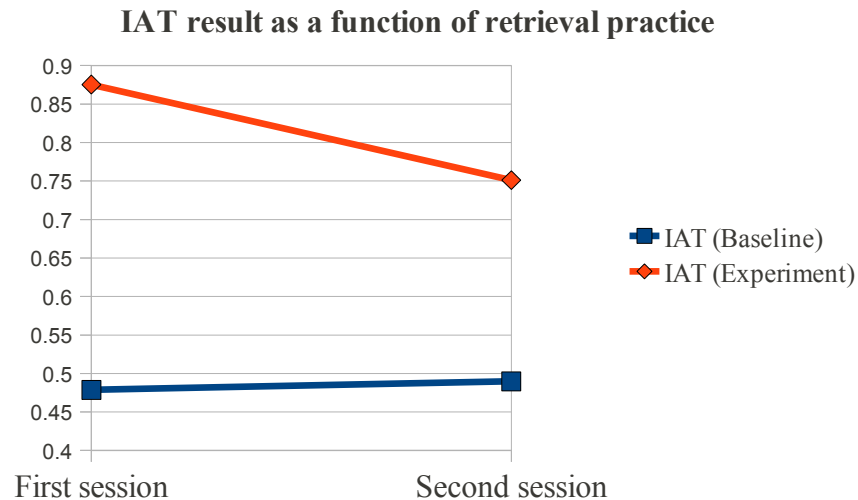


Figure 3. IAT result as a function of retrieval practice. First session is IAT before retrieval practice and second session is after retrieval practice. Baseline (blue) is the IAT for the domain that did not occur in the retrieval practice whereas Experiment (red) did occur.

Despite the lack of an interaction effect, we investigated the difference of impact between experiment and baseline condition, a two-tailed paired samples t-test was performed on the difference between IAT-E1 and IAT-E2 and the difference between IAT-B1 and IAT-B2. The result for IAT-E1 and IAT-E2 was $M = .124$ ($SD = .550$) $t(31) = 1.276$, $p = .212$ and for IAT-B1 and IAT-B2 was $M = -.011$ ($SD = .466$), $t(31) = -.134$, $p = .895$. In line with the hypothesis, there was a higher mean for the difference between IAT-E1 and IAT-E2 than for the difference between IAT-B1 and IAT-B2 (Figure 3). Although, none of the mean differences was significant.

The same test was performed on the half of the participants that showed the highest susceptibility to retrieval-induced forgetting, using a median split. The result for IAT-E1 and IAT-E2 was $M = .180$ ($SD = .477$) $t(15) = 1.506$, $p = .153$ and for IAT-B1 and IAT-B2 was $M = -.005$ ($SD = .515$), $t(15) = -.39$, $p = .969$. Showing a greater mean difference for the high susceptibility group compared to all participants.

The test was also performed on the half of the participants that attained the highest result on the first IAT-E. The result for IAT-E1 and IAT-E2 was $M = .296$ ($SD = .506$) $t(15) = 2.340$, $p = .034$ and for IAT-B1 and IAT-B2 was $M = .086$ ($SD = .484$), $t(15) = .713$, $p = .487$. Showing a significant change between E1 and E2, but not for the change between B1 and B2, for the group that attained the higher result on IAT-E1.

By comparing the difference between IAT-E1 and IAT-E2 with the difference between IAT-B1 and IAT-B2 an index describing the actual reduction of implicit associations was given. This index is referred to as IAT reduction ($IAT\ Reduction = (IAT-E1 - IAT-B1) - (IAT-E2 - IAT-B2)$). A high IAT Reduction indicated a greater reduction from IAT-E1 to IAT-E2 than IAT-B1 to IAT-B2. A between-subjects ANOVA was used to examine how the means of IAT Reduction were distributed among the four domains. The number of participants is 8 for all domains. For ethnicity $M = -.030$ and $SD = .471$. The corresponding numbers for age were $M = .176$ and $SD = .982$. For skin color $M = -.080$ and $SD = .562$. And for weight $M = .475$ and $SD = .745$. $F(3,28) = .991$ $p = .411$. Showing a mean difference between the domains that is not significant. None of the domains had an IAT Reduction that was significantly distinguished from zero (Figure 4).

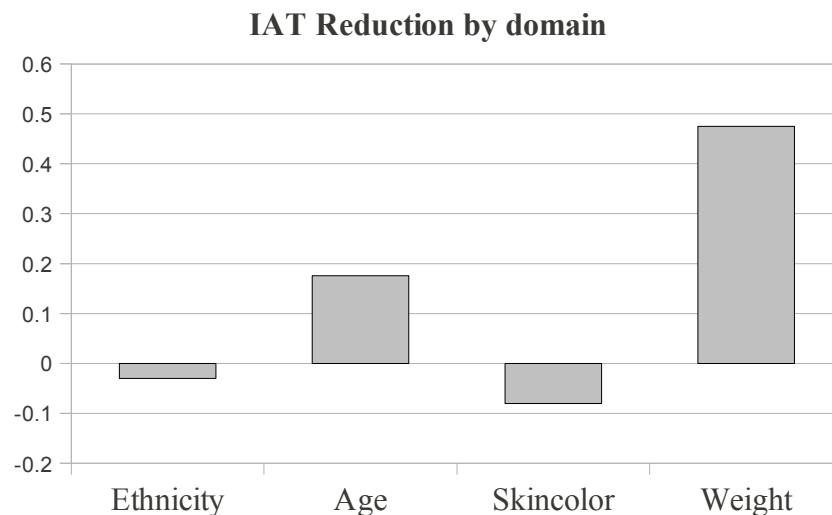


Figure 4. Mean values for the compensated IAT Reduction, mean values are presented for each domain.

Discussion

One of the main purposes of this study was to show that memory impairment occurs when performing retrieval practice with groups of people as categories and stereotype features as items. In other words, that the categories and the relation between categories and items are based on social and cultural conceptions. Our hypothesis stated that the effect of retrieval-induced forgetting would be present because the categories and features were sufficiently culturally rooted. Hence, that interference from pre-existing congruent conceptions forces the participants to inhibit features in order to retrieve context appropriate incongruent features. The results attained show a significantly lower recall rate for the unpracticed features from the practiced categories compared to the unpracticed features from the unpracticed categories. This relatively lower recall rate is consistent

with previous demonstrations of retrieval-induced forgetting and verifies our hypothesis. This appliance of the retrieval-practice paradigm is novel to the field of retrieval-induced forgetting and a significant result with this material is remarkable and brings new opportunities to the field.

Past research of retrieval-induced forgetting has employed well-demarcated categories with conventional and apparent definitions such as fruits or animals. In contrast, the categories used in this study may differ in meaning between individuals. Despite using more arbitrary categories than conventional, they proved to be solid enough to cause interference. A possible explanation is that there is an existing univocal cultural belief about the groups of people used in this study. And even if the participants submit to these cultural believes or not, they still have knowledge about them. Consequently, this knowledge appears to be sufficient to create strong conceptions for these groups which means that categorization of people based on stereotypes does not have to be more arbitrary than categorization of fruits. This make sense when for example thinking of banana, for most people a strong exemplar of the category fruit, which according to experts in the field is more related to categories such as herbs or berries. Still, fruit would work perfectly as a cue even for those that disagree because they know that the majority categorize banana as a fruit.

The features were organized so that the features meant to be inhibited were strongly related to its category and the features meant to be excited was weakly related to its category. A reason for this was that we wanted to gain use from the interference dependence principle. This enhances the inhibition during retrieval practice and applies when the relation between competing features and category are strong and connection between target and category are weak. Although, the study about retrieval-induced forgetting and stereotypes by Dunn and Spellman showed that a high-belief in a stereotype feature causes cross-item integration that makes it non-equivalent to a strongly linked item (Dunn & Spellman, 2003). The reason is that cross-item integration works as a moderating factor and prevents forgetting because of mediated retrieval. However, our experiment differs from Dunn & Spellman's in one important aspect, we used congruent features for inhibition and incongruent for excitation. Therefore, we assumed that our material was not sensitive to mediated retrieval because the links that mediates retrieval are weak or non-existent between positive-negative or negative-positive features. Our significant RIF effect confirmed that the assumption was successful. However, as we did not control for cross-item integration, its effect remains uncertain. This is true also for another moderating factor mentioned in the introduction, namely semantic similarity. Even though many of our features had overlapping semantic meanings, the barrier between congruence and incongruence was likely to prevent the semantic similarity to affect our

result.

The impaired recall rate of the inhibited words can also be described in terms of implicit associations. Remember that the words that were subjected to inhibition were strongly related and congruent to their categories. If the connection between the category and feature is inhibited, an automatic association between the category and the features might have been weakened. A problem with this conclusion is that it conflicts with the findings supporting cue-independence which shows that it is the actual item that has been inhibited and not the connection between the category and the item. But cue-independence has been shown in conditions where cross-item integration has been absent. The fact that the features more or less fit into every category could possibly make the features relevant only with their categories and thereby making them cue-dependent. The design of this experiment offers no way to find out whether the features are cue-dependent or not and we refer to the *Future Research section* for further discussion.

The next main purpose of this study was to investigate if the forgetting of negative features had potential to change the result on an IAT. A premise for our hypothesis was the notion that explicit learning precedes implicit associations. By making the participant inhibit negative information about stereotype groups and excite positive information, through retrieval practice, we presumed that this would affect the implicit associations and thereby the result on an IAT. The results attained shows that the retrieval practice did not cause a significant change on implicit associations. Although, data indicates that a change might have occurred. The mean difference between the experiment IATs before and after retrieval practice was higher than baseline. Furthermore, the change between the experimental IATs was significant when the participants that had the highest result on the initial experiment IAT were used, whereas the change on baseline IAT's was not significant under the same circumstances. This suggests that a significant result might be able to attain with either more participants or a design more focused on the implicit changes.

The fact that the participants that had the highest results on the initial experiment IAT also appeared to be most susceptible to reduction is possible to explain with the mechanisms behind retrieval-induced forgetting. If the associations between groups of people and negative features are strong, which is the case when getting a high result on IAT, then a stronger inhibition of these negative features will be needed in order to retrieve positive features. Even though the features used in retrieval practice are more specific than simply negative or positive, they are most likely to represent a general negative or positive association. Thus, a negative association in its entirety will be suppressed with a couple of negative features and not just the features in themselves. That the

initial negative association is strong will therefore lead to a stronger inhibition of that association and consequently a greater reduction caused by retrieval practice. Note that the retrieval practice is designed with the purpose to affect the IAT result in a direction that is incongruent to the common conception. This means that even if the subjects initial result on the IAT is not in line with common stereotypes, the retrieval practice would still have potential to affect the IAT result.

Presumably, the participants that showed a higher susceptibility to retrieval-induced forgetting would also show a greater reduction on the IAT. The results showed only a slightly larger reduction on the IAT compared to baseline. But the assumption regards an inter-domain comparison and the analysis presented in the Results section was performed between domains as the IAT and RIF was tested with different domains across participants. A more appropriate analysis would be to compare the reduction of IAT based on RIF susceptibility, not within participants, but within domains. But this is not possible either because different participants had different susceptibility. This catch 22 renders a comparison between reduction of IAT and susceptibility to RIF unreliable with our design unless the number of participants is substantially increased. If we had >30 participants for every condition, it would be possible to establish the differences between domains and create a measure of RIF susceptibility and IAT reduction that is independent of domain. We did not have sufficient resources nor time to collect the extra 88 participants needed for the comparison.

Our intention was to investigate the effect retrieval practice has on implicit associations in its entirety, not to investigate the differences between domains, we did not intend to compare the domains. Some differences between the domains in the results are still interesting to discuss. We found a difference in reduction of IAT results between the domains. The power of reduction seems to be greatest for weight followed by age and non-existent for the domains ethnicity and skin color. The number of participants tested in each group was only eight, which can explain the non-significant result between domains. Given that stereotypes toward Muslims and black people are more intense, has a longer tradition and are more widespread than for overweight or old people, this result is not surprising. Furthermore, the stereotypes connected to Muslims and black people are often related to fear and violence, which is considered more threatening than stereotypes commonly connected to overweight or old people. Although this is only hindsight speculations, it would be interesting to further investigate the differences between domains.

The discussion above touches another implication of our assumption about affection of implicit associations. The attitudes that are reflected in the result of an implicit association test can be seen as a result of decades of life experiences. When assuming that implicit associations based

on a well founded attitude can be changed by 20 minutes of feature retrieval, we also ignore the three-dimensional aspect of the implicit associations, that they are not only a current set of links between objects, but also a sum of life experiences. If implicit associations are well founded in our past experiences, a much more vigorous intervention than retrieval practice is needed in order to cause a modification.

There was a significant main effect between IAT experiment and IAT baseline, which is something to be concerned about as the domains were rotated between participants. The only likely explanation to this outcome is that the domain that served as experiment always occurred prior the domain that served as baseline in both IAT sessions. The significant difference is possibly due to an order effect as a consequence of the design and is also likely to reveal a confounding factor. To eliminate this confound the order of the IAT's should have been either rotated or randomized between participants.

Social applications

The results of this study demonstrate that it is possible to inhibit negative features associated with stereotypes on behalf of positive features. In other words, by frequently thinking about a group of people incongruent to the stereotype, the negative associations will be more difficult to retrieve. Since the research considering the duration of retrieval-induced forgetting is not sufficiently developed, it is problematic to discuss the lasting of our effect. It would be logical to assume that people with a solid stereotype framework (high believers) would recover faster from forgetting than those without, as they possess conscious knowledge to sustain the stereotypes. It is also likely that a solid stereotype framework results in greater interference dependence because of a stronger connection for congruent features. This results in the paradox that those who suffer strongest memory impairment also are those who recover most rapidly. Note however that it is not our intention, nor do we think it is possible, to alter solid well founded attitudes of individuals without their consent.

Our target group for a potential social application of the effect showed in our experiment are those willing to change their negative automatic associations (both explicit and implicit) caused by stereotypes. It is required that the individual is willing to change since retrieval practice is an active process. The lesson learned from this experiment is that retrieving positive thoughts about a target group is an effective way to prevent negative thoughts. A social application would consequently be to actively in our everyday life challenge our stereotypes by retrieving positive relevant information rather than actively trying to prevent retrieval of negative information based on a stereotype. To

consciously prevent negative information can still be an effective method in itself, but by retrieving positive information this prevention seem to occur automatically and unconsciously.

Future research

Although many of our questions have been answered in this study, there are still threads unanswered that are worth investigate. Some suggestions has already been mentioned in this discussion in order to refine the test. However, given the result of this study, new questions arises.

Features without congruence. When we designed this experiment, we used all available means to achieve inhibition. One of those means was to organize the features so that they had as obvious relation as possible to their categories. However, it is still an open issue to what extent our results were because the participants had a pre-understanding of the categories, or if the participants assigned the features to the categories during the experiment. This makes it difficult to determine whether we suppressed existing or assigned stereotypes. We suggest an experiment where the features are randomly distributed among the categories in order to find out if inhibition is stronger when the material is organized according to a common stereotype framework.

Testing cue-dependency. To determine whether the inhibition was due to a suppression of the actual features or a weakened connection to the category, we suggest a study that controls for cue-dependency. This can be done either by using a recognition test without the categories as cues, by using synonymic category names to describe the same group of people, or by assigning the same features to different categories. The result of such a study is important in terms of stereotype associations: Our intention is to weaken the connection between features and groups of people, not to erase features from peoples' vocabulary.

Susceptibility to reduction might depend on inhibition. We discussed the possibility that Susceptibility to reduction of implicit association might differ depending on the severity of memory impairment. The current experimental design supports that type of analysis but the number of participants has to be increased substantially.

Susceptibility to reduction might depend on domain. As the negative implicit associations for the domains age and weight might be easier to reduce, it would be interesting to carry these to an experiment focused only on the reduction of implicit associations by retrieval practice.

References

- Allport, G.W. (1954). *The nature prejudice*. Oxford and Reading, MA: Addison-Wesley.
- Anderson, M.C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*, 415–445.
- Anderson, M.C., & Bell, T. (2001). Forgetting our facts: The role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology, General*, *130*, 544–570.
- Anderson, M.C., Bjork, R.A., & Bjork, E.L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 1063–1087.
- Barden, J., Maddux, W.W., Petty, R.E., & Brewer, M.B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology*, *87*, 5-22.
- Brazel, C., & Ringqvist, K. (2009). *Retrieval-induced forgetting of emotional traits and its effect on implicit attitudes*. Masteruppsats. Lunds Universitet, Institutionen för psykologi.
- Bäuml, K. (1998). Strong items get suppressed, weak items do not: The role of item strength in output interference. *Psychonomic Bulletin & Review*, *5*, 459–463.
- Dunn, E.W., & Spellman, B.A. (2003). Forgetting by remembering: Stereotype inhibition through rehearsal of alternative aspects of identity. *Journal of Experimental Social Psychology*, *39*, 420–433.
- 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.
- Greenwald, A. G., Uhlmann, E. L., Poehlman, T. A., & Banaji, M. R. (2009). Understanding and Using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Psychology*, *97*, 17-41.
- 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.
- Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review*, *109*, 3–25.
- 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

- MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science, 12*, 148–152.
- Macrae, C.N., & MacLeod, M.D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology, 77*, 463–473.
- Mason, M. F., Tatkow, E. P., & Macrae, C. N. (2005). The look of love: Gaze shifts and person perception. *Psychological Science, 16*, 236-239
- Marx, D. M., Stapel, D.A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparison under threat. *Journal of Personality and Social Psychology, 88*, 432-446
- McGeoch, J. A. (1936). Studies in retroactive inhibition: VII: Retroactive inhibition as a function of the length and frequency of presentation of the interpolated lists. *Journal of Experimental Psychology, 19*, 674–693.
- McIntyre, R. B., Paulson, R. M., & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology, 39*, 83-90.
- Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology, 3*, 173–203.
- Norton, M.I., Vandello J.A., & Darley, J.M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology, 85*, 47-62.
- Phelps, E. A., O'Connor, K. J., Cunningham, W. A., Funayama, E. S., Gatenby, J. C., Gore, J. C., et al. (2000) Performance on indirect measure of race evaluation predicts amygdala activation. *Journal of Cognitive Neuroscience, 12*, 729-738.
- 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.
- Rydell, R.J., & McConnell, A.R. (2006). Understanding implicit and explicit attitude change: A system of reasoning analysis. *Journal of Personality and Social Psychology, 91*, 995–1008.

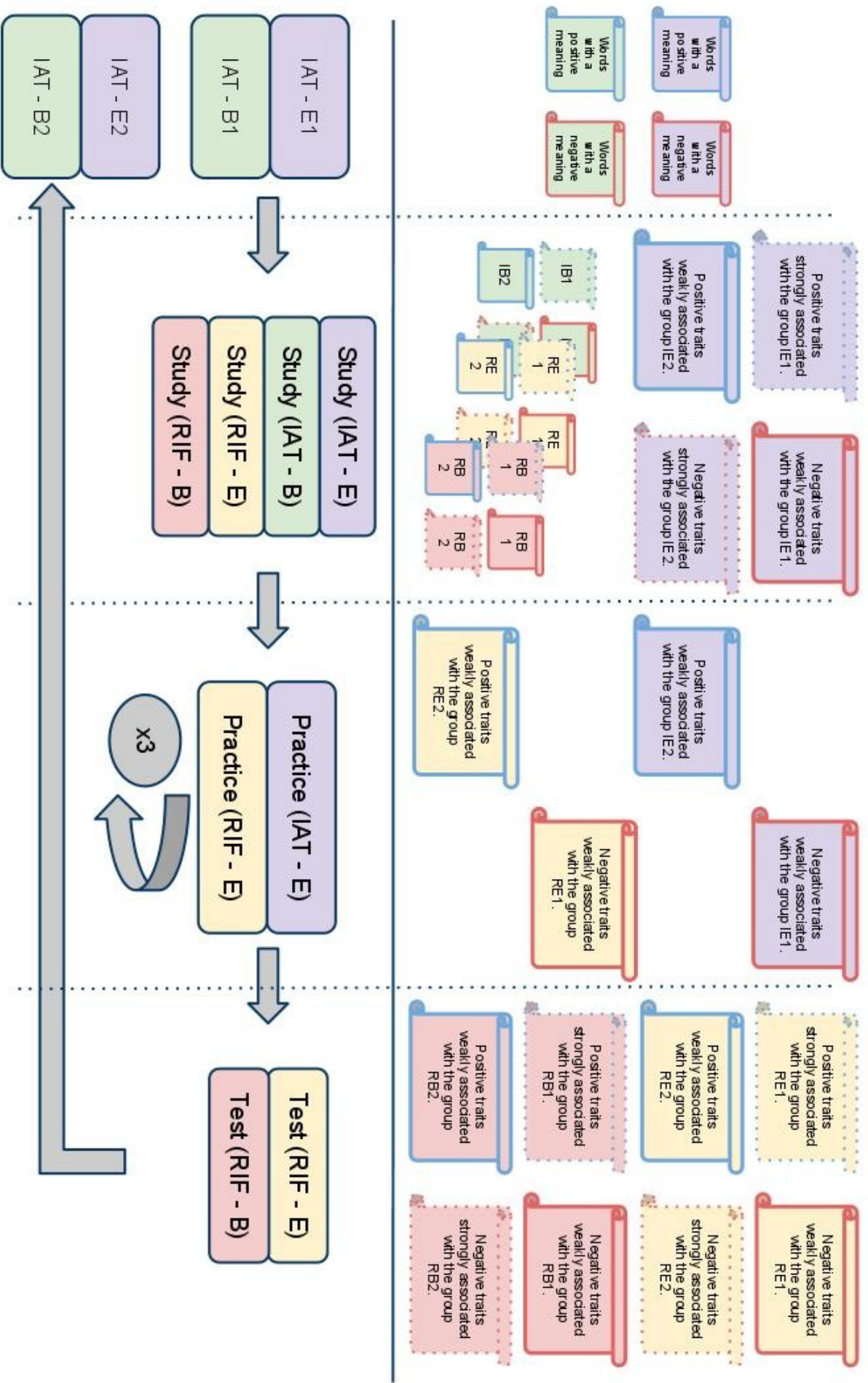
Appendix A – Flowchart

Legend

Categories that in this study are referred to as *preferred* are represented with number one (1). Categories that are referred to as *non-preferred* are represented with the number two (2). The color blue and the letters *IE* represents the IAT experiment condition, green and *IB* represents the IAT baseline condition, yellow and *RE* represents the RIF experiment condition and finally, red and *RB* represents the RIF baseline condition. Lists with a dotted frame contain features that are meant to inhibit and are congruent to their categories. Lists with a solid frame contain features that are meant to excite and are incongruent to their categories. Lists with a blue frame contains positive features, lists with a red frame contains negative features. The domains are rotated among the conditions between subjects in the following way:

Rotation of domains

#	RIF Experiment	RIF Baseline	IAT Experiment	IAT Baseline
1	<i>Ethnicity</i>	<i>Age</i>	<i>Skincolor</i>	<i>Weight</i>
2	<i>Age</i>	<i>Ethnicity</i>	<i>Weight</i>	<i>Skincolor</i>
3	<i>Skincolor</i>	<i>Weight</i>	<i>Ethnicity</i>	<i>Age</i>
4	<i>Weight</i>	<i>Skincolor</i>	<i>Age</i>	<i>Ethnicity</i>



Appendix B – Material

Legend

The titles for every block of words are in this study referred to as *domains* and the subtitles as *categories*. Categories on the left side are referred to as *preferred* and categories on the right side as *non-preferred*. Bold features are in this essay depending on the context referred to as: *Congruent*, *practiced*, *meant to excite* or *targets*. Gray features are referred to as: *Incongruent*, *unpracticed*, *meant to inhibit* or *competitors*.

Ethnicity

<i>Svensk</i>		<i>Muslim</i>	
Blyg	Ansvarsfull	Artig	Aggressiv
Försiktig	Behärskad	Begåvad	Fanatisk
Introvert	Duktig	Driftig	Fientlig
Missunsam	Förnuftig	Hjälpsam	Förtryckande
Reserverad	Hederlig	Intelligent	Hotfull
Sluten	Ordentlig	Medgörlig	Kontrollerande
Stel	Pålitlig	Omsorgsfull	Oresonlig
Timid	Rationell	Osjälvisk	Samvetslös
Tråkig	Skötsam	Vänlig	Våldsam

Age

<i>Ung</i>		<i>Gammal</i>	
Ansvarslös	Äventyrlig	Djupsinnig	Bitter
Dumdristig	Bekymmerslös	Erfaren	Envis
Förvirrad	Energisk	Hövlig	Glömsk
Osäker	Impulsiv	Klok	Inskränkt
Otålig	Initiativtagande	Omtänksam	Negativ
Rastlös	Kreativ	Rar	Pessimistisk
Tanklös	Nytänkande	Snäll	Snål
Uppkäftig	Påhittig	Väluppfostrad	Svårlärd
Vårdslös	Uppfinningsrik	Vis	Vresig

Skincolor

<i>Ljushyad</i>		<i>Mörkhyad</i>	
Arrogant	Disciplinerad	Bekräftande	Finesslös
Cynisk	Flitig	Engagerad	Främmande
Dryg	Förståndig	Exotisk	Kriminell
Elitistisk	Idérik	Företagsam	Lynnig
Intrigerande	Mottaglig	Intressant	Obegåvad
Känslökall	Smart	Öppen	Ohederlig
Lömsk	Talangfull	Social	Opålitlig
Nonchalant	Trevlig	Spontan	Påstridig
Tillgjord	Vettig	Tjänstvillig	Påträngande

Weight

<i>Smal</i>		<i>Tjock</i>	
Åsidosättande	Ambitiös	Bussig	Äcklig
Fåfång	Effektiv	Djup	Asocial
Fördomsfull	Extrovert	Fantasifull	Destruktiv
Manipulativ	Hälsosam	Glad	Karaktärslös
Nedlåtande	Ordningssam	Godhjärtad	Klumpig
Okänslig	Redig	Härlig	Labil
Respektlös	Sund	Humoristisk	Lat
Skoninglös	Sympatisk	Lättsam	Oduelig
Ytlig	Vårdad	Rolig	Passiv