When Suppressing One Stereotype Leads to Rebound of Another: On the Procedural Nature of Stereotype Rebound

Nicolas Geeraert

Abstract

A known consequence of stereotype suppression is post-suppressional rebound (PSR), an ironic activation of the suppressed stereotype. This is typically explained as an unintended by-product from a dual-process model of mental control. Relying on this model, stereotype rebound is believed to be conceptual. Alternative accounts predict PSR to be featural or procedural. According to the latter account, stereotype rebound would not be limited to the suppressed social category, but could occur for a target from any social category. The occurrence of procedural stereotype rebound was examined across five experiments. Suppression of one particular stereotype consistently led to rebound for social targets belonging to the same or a different stereotype in an essay-writing task (Experiments 1-3) and led to facilitation in recognition of stereotype-consistent words (Experiment 4). Finally, stereotype suppression was shown to impact on assessments of stereotype use but not on heuristic thinking (Experiment 5).

Keywords

stereotypes, suppression, rebound, procedural rebound

Received December 2, 2011; revision accepted April 11, 2013

An extensive body of research suggests that people experience great difficulty in mental control (Wegner, 1994, 2009). People attempting thought suppression are plagued with intrusions of the forbidden thoughts, particularly when under cognitive load (Wegner & Erber, 1992). When the intention to suppress is relaxed, the suppressed concept becomes more accessible than before, an effect known as post-suppressional rebound (PSR). For instance, participants suppressing white bear thoughts report subsequent more frequent intrusions of such thoughts (Wegner, Schneider, Carter, & White, 1987). Apart from white bears, PSR has proven useful in clinical settings (Wenzlaff & Wegner, 2000) such as dieting (Erskine, 2008; Ward, Bulik, & Johnston, 1996), smoking cessation (Erskine, Georgiou, & Kvavilashvili, 2010), emotional experience (Richards & Gross, 1999), and posttraumatic stress (Davies & Clark, 1998).

PSR has also been demonstrated with social stereotypes (Bodenhausen & Macrae, 1998; Bodenhausen, Todd, & Richeson, 2009; Monteith & Mark, 2009; Monteith, Sherman, & Devine, 1998). Participants instructed to suppress stereotypic thoughts next produce more stereotypic descriptions (Macrae, Bodenhausen, Milne, & Jetten, 1994; Wyer, Sherman, & Stroessner, 2000), modify their answers on a recognition test (Sherman, Stroessner, Loftus, & Deguzman, 1997), or alter their behavior toward members of the target social category (Macrae et al., 1994).

Modality of Rebound

Different models have been proposed to account for PSR, including explanations based on cognitive demand (Wegner, 1994), unintended priming of the forbidden thought (Macrae et al., 1994), motivational inferences (Forster & Liberman, 2001), and mental fatigue (Muraven, Tice, & Baumeister, 1998). Somewhat independently from the theoretical explanation is the modality of PSR. More specifically, at what level is rebound occurring? Although this issue has received

1University of Essex, Colchester, UK

Corresponding Author:
Nicolas Geeraert, Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK.
Email: geeraert@essex.ac.uk
little attention to date, some different modalities of PSR have been demonstrated: conceptual, featural, and procedural.

**Conceptual Rebound**

Wegner (1992, 1994) proposed a dual-process model of mental control, comprising a monitoring and operating process. The automatic monitoring process scans consciousness for any forbidden thoughts; if any are detected, the controlled operating process replaces the unwanted thought with a distractor. According to this model, mental control can and will often be successful. However, due to the cognitive demands of the operating process, a reduction in cognitive capacity (cognitive load, time pressure, etc.) would lead to ironic suppression effects (Wegner, 1994). Such cognitive interference of the operating process would then give free reign to hyper-accessibility caused by the monitoring process. Others have made this argument even more explicitly, stating that the monitoring process is actually priming the forbidden thought (Macrae et al., 1994). Although the dual-process model does not make explicit mention of rebound modality, it has nonetheless been designed to explain and account for the ironic effects of thought suppression. As such, according to the model, PSR occurs on a conceptual level, as suppression and rebound occur at the level of a mental representation of a particular object or class of objects. In terms of stereotypes, stereotypical suppression of a target (a hooligan) would be expected to lead to rebound on a similar social exemplar, belonging to the same social category (a hooligan).

**Featural Rebound**

A separate series of experiments suggests that rebound does not only occur on the level of concepts or mental representations. For instance, in one experiment, participants were told they possessed allegedly undesired personality traits and instructed to suppress these; subsequently, participants inferred these very same traits from the behavior of others (Newman, Duff, & Baumeister, 1997; Newman, Duff, Hedberg, & Blitstein, 1996). Recently, stereotype suppression was also shown to lead suppressors to incorporate the stereotypical attributes or traits into their self-concept or personal behavior (Follenfant & Ric, 2010; Wyer, Mazzoni, Perfect, Calvini, & Neilens, 2010). Importantly, in all of these studies, the rebound targets were not conceptually related to the suppressed attributes or stereotypes. Thus, what seems to be rebounding in these experiments is the attributes, not the actual stereotypes. Consequently, this suggests that the locus of suppression and rebound is at the level of exemplar features, as opposed to the actual exemplars. Although not originally hypothesized by Wegner’s dual-process model, the model could nevertheless account for these findings through the mechanism of spreading activation (Collins & Loftus, 1975). Interestingly, in terms of stereotypes, this account suggests that stereotypical suppression of a target (hooligan) may actually lead participants to attribute suppressed features (aggressive) to an unrelated target (a gay man) regardless whether this target is normally associated with these features.

**Procedural Rebound**

Other empirical evidence seems to fit neither of the above modalities. For instance, participants suppressing dispositional inferences subsequently relied more on dispositional language to describe others’ behavior in an unrelated task (Geeraert & Yzerbyt, 2007; Geeraert, Yzerbyt, Corneille, & Wigboldus, 2004). More recently, participants suppressing similarity use in a categorization task were shown to rely more on similarity in an unrelated memory task (Geeraert, Van Boven, & Yzerbyt, 2011). Rebound in each of these examples cannot be explained on the level of the concept or its features. Rather, these findings suggest that PSR also operates at the level of mental procedures. Thus far, procedural PSR has been demonstrated with dispositional attributes and similarity assessment, but actually not yet with stereotypes. Vis-à-vis stereotype rebound, this account would suggest what is rebounding may very well be stereotyping (a mental procedure) as opposed to a stereotype (a concept). In other words, stereotype rebound may not be limited to the suppressed social category but could occur for any target from any social category. Thus, people who suppress stereotypes of one social category (hooligan) may display higher levels of stereotypical thoughts about an unrelated category (a gay man).

According to the procedural model, the cognitive processes underlying stereotyping could be temporarily interrupted during stereotype suppression. However, these processes would be expected to resurge or rebound after the suppression phase. Obviously, stereotyping and prejudice are complex cognitive processes, potentially involving different subprocesses, such as categorization (Allport, 1954; Taylor, 1981). Importantly, the procedural account makes no claims with regard to the complexity and interaction of these different subprocesses. Rather it argues that stereotype suppression may impact on these processes, such that stereotyping is generally more likely following stereotype suppression.

Importantly, the predictions made by the procedural account are quite different from alternative models such as ego-depletion (Muraven & Baumeister, 2000). According to this influential approach, acts of self-regulation may lead to mental fatigue that would impede subsequent tasks. Likewise, stereotype suppression may be so cognitively demanding that it will lead to mental fatigue, and thus have knock-on effect for future acts of self-control (Muraven et al., 1998). Through the mechanism of self-control, the ego-depletion account offers a viable explanation for stereotype rebound, but more broadly it argues that suppression may lead to more heuristic processing. The procedural rebound account does not make such broad claims.
Overview of Current Experiments

The occurrence of stereotype rebound at the procedural level was examined across five experiments. Each experiment followed the same basic procedure: a suppression task followed by a rebound assessment task. In the suppression task, participants provided a written description of a typical day in the life of a social target (Macrae et al., 1994). One group of participants was induced to avoid using stereotypes (suppression condition), whereas a second group did not get such instruction (control condition). Experiments 1 to 3 all used a similar rebound assessment task; stereotype-related thoughts for different target categories were assessed by means of a written description. These targets belonged to either the same or a different social category than the target from the suppression task. After testing whether stereotypes rebound at the procedural level (Experiment 1), the extent to which stereotype rebound can be explained by a procedural account was examined (Experiments 2 and 3). A different rebound assessment task was used in the next study. Specifically, stereotype use was assessed by a primed lexical decision task (LDT; Experiment 4). In the final study, the boundary conditions of stereotype rebound were examined. Specifically, the consequences of stereotype suppression were examined on stereotypic thinking versus heuristic thinking (Experiment 5).

Experiment 1

Participants were exposed to a target belonging to the category of “chavs,” a British stereotype to denote teenagers who repeatedly engage in antisocial and delinquent behavior. Half of the participants were asked to suppress stereotypes. Participants were then exposed to a second, unrelated target, a “teenage mother.” Suppression participants were predicted to be more stereotypical toward the second target, thus displaying procedural PSR.

Method

Nineteen participants ($M_{\text{age}} = 20.2$ years) were randomly assigned to one of two conditions (suppression vs. control).

Pretest. Pretest participants ($N = 20$) were asked to think about both stereotypes and list five stereotypical attributes for each. Chavs were typically perceived as socio-economic challenged (e.g., unemployed), displaying deviant behavior (e.g., aggression, substance abuse) and verbal aggression (e.g., mouthy), and a typical physical appearance (e.g., hoodies). In contrast, teenage mothers were associated with promiscuous behavior (e.g., easy), immaturity (e.g., naive), situational descriptions (e.g., unmarried), and psycho-pathology (e.g., depression). All in all, the pretests suggest that these targets shared little or no semantic overlap.

<table>
<thead>
<tr>
<th>Table 1. Stereotypicality Ratings for Experiment 1 as a Function of Experimental Condition and Essay Order.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suppression ($n = 10$)</td>
</tr>
<tr>
<td>Essay 1</td>
</tr>
<tr>
<td>Essay 2</td>
</tr>
</tbody>
</table>

Note. Table shows mean ratings (standard deviation in parentheses). High scores indicate the essays were rated more stereotypical.

Materials and procedure. Participants were shown a photograph of a White male wearing a tracksuit and a hooded top, and instructed to compose a short essay describing a typical day in the life of the individual. Pilot testing had confirmed the individual on the picture was perceived as belonging to the category “chavs.” Suppression participants, but not controls, were also instructed to complete the task without thinking stereotypically about the target. Participants spent 5 min writing the essay. After a short filler task, participants were introduced to the second essay-writing task and shown a photograph of a teenage girl holding a toddler. Participants were provided with some additional background information about the target such as her name, age, and home town. After writing for another 5 min, participants answered a manipulation check. Specifically, they indicated to what extent they tried not to think stereotypically about the first target on a 7-point scale (1 = did not try at all, 7 = tried very hard). Finally, participants were debriefed.

Dependent measure. Two blind raters rated the essays for stereotypicality, of “chav” and “teenage mother” respectively, on a 7-point scale (1 = not at all stereotypical, 7 = totally stereotypical). Interrater reliability was good (both $r's > .75$), so mean scores were computed for both essays.

Results and Discussion

Manipulation check. First, the self-report measure was analyzed by means of a one-way ANOVA. Indicating the success of the experimental manipulation, suppression participants reported avoiding stereotypes more ($M = 5.80, SD = .79$) than controls ($M = 4.00, SD = 1.66$), $F(1, 17) = 9.45, p < .01, d = 1.49$.

Suppression and rebound. Stereotypicality of essay ratings was analyzed using a 2 (condition: suppression vs. control) x 2 (essay: 1 vs. 2) mixed model ANOVA, with repeated measures on the second factor. Analyses revealed the predicted interaction, $F(1, 17) = 21.25, p < .001, \eta^2_p = .56$ (see Table 1). As expected, suppression participants used less stereotypical language than did controls for Essay 1, $F(1, 17) = 6.74, p < .02, d = 1.26$. In the second description, however, stereotypicality was higher for suppression participants than for control participants, $F(1, 17) = 10.77, p < .01, d = 1.60$. Overall, this suggests that suppression participants successfully...
suppressed the chav stereotype, but subsequently used more stereotypes for teenage mother. This novel finding provides initial evidence for a procedural account of stereotype rebound.

**Experiment 2**

Having established the existence of procedural stereotype rebound, Experiment 2 compared the independent influences of conceptual versus procedural stereotype rebound. Suppression and control participants were again asked to describe a typical day of a male “chav.” Importantly, however, the target for the second essay varied across participants. Half of the participants were asked to describe a new target belonging to the same social category (a second male chav). The other participants were asked to describe a social target belonging to a different category (a teenage mother). This factorial design allows to test whether the rebound effect is similar for a stereotype of a different category compared with a stereotype of the same category.

**Method**

Sixty participants (M age = 23.0 years), writing two essays each, were randomly assigned to a 2 (condition: suppression vs. control) × 2 (stereotype category: same vs. different) design.

**Materials and procedure.** The materials and procedure were identical to the previous experiment. Suppression and control participants wrote an essay about a male “chav” target. However, the second target either belonged to the same category (male “chav”) or a different category (“teenage mother”). Participants had 5 min to complete each essay. All essays were rated for stereotypicality. Following good interrater agreement (r’s > .80), single mean rating scores were computed.

**Results and Discussion**

**Manipulation check.** Suppression participants reported suppressing harder (M = 5.90, SD = .76) than did controls (M = 3.93, SD = 1.76), F(1, 58) = 31.58, p < .001, d = 1.48, which suggested the manipulation was successful.

**Suppression and rebound.** To examine the occurrence of suppression and rebound, the essay ratings were analyzed by means of a 2 (condition) × 2 (essay) mixed model ANOVA. The two-way interaction was significant, F(1, 58) = 103.24, p < .001, ηp² = .64 (see Table 2). Follow-up analysis revealed that suppression participants used less stereotypical language than control participants in Essay 1, F(1, 58) = 23.44, p < .001, d = 1.26. In contrast, suppression participants described the second target in more stereotypical language than controls, F(1, 58) = 55.78, p < .001, d = 1.96. Taken together, these results indicate that suppression participants initially suppressed stereotypes, but subsequently showed PSR.

**Note.** Table shows mean ratings (standard deviation in parentheses). High scores indicate the essays were rated more stereotypical.

<table>
<thead>
<tr>
<th>Essay 1</th>
<th>Suppression (n = 30)</th>
<th>Control (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essay 2 (same category)</td>
<td>5.20 (1.85)</td>
<td>2.70 (1.58)</td>
</tr>
<tr>
<td>Essay 2 (across categories)</td>
<td>5.40 (1.64)</td>
<td>2.53 (1.32)</td>
</tr>
</tbody>
</table>

**Rebound modality.** To assess whether stereotype rebound was qualified by target category, the stereotypicality ratings for the second essay (see Table 2) were analyzed by means of a 2 (condition) × 2 (stereotype category) factorial ANOVA. This analysis revealed a main effect of condition, F(1, 56) = 54.73, p < .001, ηp² = .49, re-confirming that suppression participants relied more on stereotypes compared with control participants. Importantly, however, this effect was not qualified by the interaction (F < 1). The absence of the interaction, in the presence of the main effect, implies that PSR occurred to a similar degree for a target belonging to the same and a different category. This suggests that stereotype rebound occurs at a procedural level, not at a conceptual level.

**Experiment 3**

The results of Experiment 2 are novel and qualify earlier research on stereotype rebound. To further consolidate these findings, Experiment 2 was conceptually replicated but with two different stereotypical categories, namely “Essex girl” (first and second description) and a “man of working class” (second description). In essence, the stereotype of “Essex girl” is a British variation on the stereotype of a “dumb blonde.” The British stereotype for “working class” is more ambivalent, as it can either be positive or derogatory. It would include both people in lower tier jobs and those who are unemployed. The procedure for Experiment 3 was otherwise identical to that of Experiment 2.

**Table 2. Stereotypicality Ratings for Experiment 2 as a Function of Experimental Condition and Essay Order (Top Panel), or Experimental Condition and Essay Content (Bottom Panel).**

<table>
<thead>
<tr>
<th>Essay 1</th>
<th>Suppression (n = 30)</th>
<th>Control (n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essay 1</td>
<td>1.57 (.94)</td>
<td>2.95 (1.26)</td>
</tr>
<tr>
<td>Essay 2 (across categories)</td>
<td>5.40 (1.64)</td>
<td>2.53 (1.32)</td>
</tr>
<tr>
<td>Essay 2 (same category)</td>
<td>5.20 (1.85)</td>
<td>2.70 (1.58)</td>
</tr>
<tr>
<td>Essay 2 (different category)</td>
<td>5.60 (1.43)</td>
<td>2.37 (1.03)</td>
</tr>
</tbody>
</table>

**Method**

Forty-eight participants (M age = 21.4 years), writing two essays each, were randomly assigned to a 2 (condition) × 2 (stereotype category) design.

**Pretest.** Pretest participants (N = 20) were asked to list attributes of the two social categories “Essex girl” and “working class man.” There was no substantial overlap between attributes, suggesting the two categories to be semantically different.
Table 3. Stereotypicality Ratings for Experiment 3 as a Function of Experimental Condition and Essay Order (Top Panel), or Experimental Condition and Essay Content (Bottom Panel).

<table>
<thead>
<tr>
<th></th>
<th>Suppression (n = 24)</th>
<th>Control (n = 24)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Essay 1</td>
<td>2.63 (1.37)</td>
<td>3.67 (1.50)</td>
</tr>
<tr>
<td>Essay 2 (across categories)</td>
<td>5.54 (1.38)</td>
<td>4.04 (1.73)</td>
</tr>
<tr>
<td>Essay 2 (same category)</td>
<td>5.46 (1.14)</td>
<td>3.71 (1.59)</td>
</tr>
<tr>
<td>Essay 2 (different category)</td>
<td>5.63 (1.63)</td>
<td>4.38 (1.86)</td>
</tr>
</tbody>
</table>

Note. Table shows mean ratings (standard deviation in parentheses). High scores indicate the essays were rated more stereotypical.

Materials and procedure. The suppression target category was “Essex girl.” The target category for the second essay was either “Essex girl” (same category) or “working class man” (different category). For each of the three targets, photographs were selected and pretested to ensure they were a good representation of the target category. Participants were also presented with some additional background information about each target. Again, all essays were rated for stereotypicality. Following good interrater agreement ($r > .80$), mean rating scores were computed.

Results and Discussion

Manipulation check. Suppression participants reported suppressing more ($M = 5.13, SD = 1.62$) than did controls ($M = 3.17, SD = 1.58$), $F(1, 46) = 17.95, p < .001, d = 1.25$, indicating our manipulation was again successful.

Suppression and rebound. A 2 (condition) × 2 (essay) mixed model ANOVA on the stereotype ratings revealed a significant interaction, $F(1, 46) = 24.60, p < .001, \eta^2_p = .35$ (see Table 3). For Essay 1, suppression participants were less stereotypical compared with control participants, $F(1, 46) = 6.31, p < .02, d = .74$. Once more, the pattern was reversed for the second target, showing that suppression participants were more stereotypical than control participants, $F(1, 46) = 11.10, p < .01, d = .98$, thus displaying rebound.

Rebound modality. A 2 (condition) × 2 (stereotype category) ANOVA on Essay 2 revealed a main effect of condition, $F(1, 44) = 10.89, p < .01, \eta^2_p = .20$, re-affirming that suppression participants were more stereotypical than control participants. As before, the interaction was not significant ($F < 1$). Overall, these findings indicate that stereotype rebound was equally strong independent of the target category.

Experiment 4

In the previous experiments, target categories were selected and pretested to ensure they did not share any stereotypic descriptions or attributes. Despite the fact that the social categories did not semantically overlap, it cannot be completely ruled out that these categories are related at some level. For instance, although substantially different, the social categories in Experiments 1 and 2 are likely to share some demographical aspects such as age and socio-economic status. For this reason, a novel set of different social categories was selected deliberately varying age, socio-economic status, and valence. As before, the suppression target was chav, and the rebound assessment categories were chavs, the elderly, and businessmen.

In addition to this, Experiment 4 also adopted a different paradigm to assess stereotype use during the rebound phase. As before, suppression and control participants were asked to describe a typical day of a male “chav.” This was followed by a primed LDT. Reaction times in a LDT are a common measure of stereotype activation (Dovidio, Evans, & Tyler, 1986; Wittenbrink, Judd, & Park, 1997), and have previously been used as a measure of stereotype rebound (Macrae et al., 1994, Study 3). In the current experiment, each trial was preceded by a prime, denoting one of three social categories (chav, the elderly, and businessmen), followed by a probe (word or nonword). Prime and probe words were either stereotypically consistent (e.g., chav-noisy, chav-rude, elderly-forgetful) or inconsistent (i.e., chav-caring, businessmen-weak). Suppression participants were expected to respond quicker to probe words when these were preceded by a stereotype-consistent prime.

Method

Thirty-four participants ($M_{age} = 21.1$ years) were randomly assigned to one of two conditions (suppression vs. control). Due to a poor command of English, one participant failed to understand the instructions of the experiment. This participant was removed from the data prior to analyses.

Pretesting. A series of different social categories were selected. For each category, a list of approximately 40 stereotypical words or attributes was pretested for valence and the extent to which the attribute was a good description of the social category (stereotypicality). Following the pretesting, a set of three distinct social categories were selected (chav, elderly, and businessmen), each with an associated list of eight stereotypically consistent words and eight stereotypically inconsistent words. All six lists were matched for stereotypicality and approximately matched for word length. In terms of valence, the list of stereotypical chavs words was negative, the list of elderly attributes was relatively neutral, and the list consistent for businessmen was moderately positive. Obviously, the target categories also varied in age and socio-economic status.

Pilot test. Prior to the experiment, the materials and procedure were pilot tested. The pilot study revealed suppression participants to be significantly quicker, relative to controls,
to recognize words preceded by stereotypically consistent primes (p < .03). Although lacking the appropriate control trials (i.e., stereotypically inconsistent primes), the pilot study provided some initial evidence for stereotype rebound through a primed LDT.

Materials and procedure. The suppression task was identical to that of Experiments 1 and 2; suppression and control participants were asked to describe a typical day in the life of a male “chav” target. Again, the essay of the suppression task was rated by two blind judges. Following good interrater agreement (r = .90), mean rating scores were computed.

After the essay and a short filler task, participants were introduced to an ostensibly unrelated task, a primed LDT. The LDT was programmed in Inquisit, and responses had to be given on a Cedrus RB-834 response pad. Participants were informed that each trial would consist of a fixation cross, followed by a blue word (i.e., the prime), and finally followed by a letter string printed in black (i.e., the probe). They were told to keep their attention on the fixation cross and blue word, and to only respond to the probe. Each trial in the primed LDT started with a fixation point for 200 ms (“+”), immediately followed by the prime word for 500 ms (“chav,” “businessmen,” or “elderly”). After a 200 ms interval, the probe (word or nonword) appeared and stayed on the screen until a correct response was given. Participants first completed a practice block consisting of 24 trials, followed by two experimental blocks, each consisting of 96 randomized trials. In each block and for each social category (chav, businessmen, and elderly), participants responded to 8 probe words that were stereotypically consistent, 8 probe words that were stereotypically inconsistent, and 16 nonwords. On completion of the primed LDT, participants were asked to respond to the manipulation check, and fully debriefed.

Results and Discussion

Prior to analyses, reaction times were examined for the presence of outliers (Ratcliff, 1993). Inspection of frequencies revealed the distribution of rt’s to be positively skewed. For each participant, a mean rt was computed and Z-scored. All participants’ scores fell within a range of 2 SD from the mean. To deal with outliers on the level of individual trials, rt’s below 300 ms and above 1,500 ms were removed from the data. A mean rt score was computed for each pair of prime (chav vs. businessmen vs. elderly) and probe (consistent vs. inconsistent). On the basis of the practice trials, an individual baseline score was computed for each participant. Finally, the baseline score was subtracted from each of the mean scores of the six prime–probe pairs, resulting in six facilitation scores per participant.

Suppression task. First, the manipulation check was analyzed. As before, suppression participants reported avoiding stereotypes more (M = 5.65, SD = 1.00) than did controls

(\(M = 2.94, SD = 1.73\), \(F(1, 31) = 10.74, p < .001, d = 1.99\). The stereotypicality ratings of the essay were analyzed subsequently. As expected, suppression participants’ descriptions were less stereotypical (M = 2.24, SD = 1.55) than the descriptions of control participants (M = 4.06, SD = 2.34), \(F(1, 31) = 7.05, p < .02, d = .95\). Again, these findings suggest that suppression participants successfully suppressed stereotypes.

Primed LDT. Rt facilitation scores were analyzed by means of a 2 (condition: suppression vs. control) × 3 (prime: chav vs. businessmen vs. elderly) × 2 (probe: consistent vs. inconsistent) mixed model ANOVA, with repeated measures on the two latter factors. None of the main effects were significant (all \(F\)'s < 1). More interestingly, the predicted interaction of condition by probe was significant, \(F(1, 31) = 7.15, p < .02, \eta_p^2 = .19\). To examine this interaction, the facilitation scores were plotted (see Figure 1). Simple main effect analyses showed that suppression participants had greater facilitation for stereotype-consistent probe words (\(M = 44 ms, SD = 22 ms\) than for probe words that were stereotypically inconsistent (\(M = 27 ms, SD = 22 ms\), \(F(1, 31) = 4.51, p < .05, d = .76\). Although not significant, the pattern of results was somewhat reversed for control participants, \(F(1, 33) = 2.77, p = .11\). Thus, suppression participants, but not controls, more quickly recognized words that were preceded by a stereotypically related prime. Importantly, the three-way interaction did not reach significance (\(F < 1\)), suggesting the interaction of condition by probe was not qualified by type of prime. In other words, whether the primed stereotype in the LDT was similar or different from the suppression category did not affect the facilitation effect.

As a set, these findings clearly indicate that suppression participants were faster to recognize words associated with different social categories. In as far as LDT reaction times

![Figure 1. The facilitation scores for Experiment 4, as a function of experimental condition and prime-probe consistency or inconsistency.](image-url)
show stereotypic activation (Dovidio et al., 1986; Wittenbrink et al., 1997), this would imply that suppressors showed higher levels of stereotype activation for related and unrelated stereotypes. The findings from the primed LDT clearly match those of a procedural account of stereotype rebound.

**Experiment 5**

All of the previous experiments demonstrated that post-suppressional stereotype rebound occurs to an equal extent for different social groups, regardless of whether the social category of the rebound task was initially suppressed or not. In other words, these findings seem to indicate that the consequences of stereotype suppression are in fact broader than the conceptual account allows for. Importantly, however, this naturally raises questions on the boundary condition of the stereotype rebound phenomenon. Does the act of suppression, like self-control, leave people in a state of temporary mental fatigue (Muraven et al., 1998)? Or, through some other mechanism, do suppressed individuals make qualitatively different judgments, be they social or not? For a long time, psychologists have made a distinction between two cognitive systems: System 1 and System 2 (Kahneman, 2011). Perhaps thought suppression and/or stereotype suppression renders people more likely to temporary rely on System 1. If this were the case, then suppressors would be expected to rely more readily on heuristic judgments (Tversky & Kahneman, 1974), be less cognitively reflexive (Frederick, 2005), or be persuaded by peripheral cues (Petty & Cacioppo, 1986). Crucially, stereotype thinking is itself heuristic in nature; thus, this explanation provides a very viable alternative to the procedural rebound account.

The final experiment sought to address this alternative account by directly pitting these two accounts against each other. Similar to the previous experiments, suppression and control participants were first asked to describe a typical day in the life of a male “chav.” This time, the suppression task was followed by one of two types of assessment to which participants were randomly assigned. Participants either completed an assessment of stereotype rebound or an assessment of System 1 thinking. The former was operationalized by means of a word fragment completion task and the latter assessment consisted of two different tasks. First, borrowing from the persuasion literature, a measure was constructed to assess the extent to which participants were more or less persuaded by a message containing fewer strong arguments or more weak arguments (Petty & Cacioppo, 1984). Second, the cognitive reflection task was administered to participants (Frederick, 2005).

This factorial design allows us to compare the heuristic assessments with the stereotype assessment. If the heuristic account is accurate, then heuristic thinking is to occur equally strongly as stereotype rebound. In other words, a main effect of condition would be expected, but no interaction. If, however, the procedural account is accurate, then stereotype rebound is to occur more strongly than heuristic thinking. This would be evident from an interaction between suppression condition and mode of assessment.

**Method**

Forty-eight participants (M_age = 20.0 years) were randomly assigned to the 2 (condition: suppression vs. control) × 2 (assessment: heuristic vs. stereotypic) design.

**Pretesting.** For the persuasion task, 2 message topics had to be selected, each with three strong and nine weak arguments. Initially, 25 topics were compiled and pretested for general agreement. Of these, 6 neutral topics were selected. For each selected topic, a series of arguments were generated and pretested for argument strength. Finally, 2 topics were selected (“a tax should be imposed on junk food and fatty snacks” and “people should pay for the quantity of waste they produce”) with three strong and nine weak arguments for each.

The stereotype task consisted of a word fragment completion task. A series of words were pretested for their level of stereotypicality of “chav.” Thirty words were selected for the word fragment completion task: 12 words with high stereotypicality and 18 words that were not associated with the stereotype.

**Materials and procedure.** The suppression task was identical to that of Experiments 1, 2, and 4; suppression and control participants were asked to describe a typical day in the life of a male “chav” target. Two blind judges rated the essays for stereotypicality. Following good interrater agreement (r > .70), mean rating scores were computed. After completing the essay and a short filler task, participants were introduced to either the persuasion task or the stereotype assessment task.

The persuasion task was modeled after Petty and Cacioppo (1984). Participants were presented with an alleged impression formation task. Each participant was shown two messages (tax on junk food and waste payment), varying in number of arguments (three strong and nine weak arguments). They were told they would be shown some samples of what other people had written, and for each person asked to try and form an impression of that person. Participants were told that a student “was asked to think about the following policy” (tax on junk food or waste payment) and that “this individual was asked to list the main reasons why this would be a good policy.” A box then displayed the message containing the arguments (three strong or nine weak). Participants were told that because their personal opinion about the policy could influence their rating of the person, they first had to respond to some questions regarding their personal opinion. Participants then rated the policy on four 7-point semantic differential scales (bad/good, harmful/beneficial, foolish/wise, unfavorable/favorable), which were averaged into a single score (α = .89). In line with the cover
story, participants next responded to a few filler questions regarding the target student.

Participants responded to two different messages, giving their opinion on each topic. For each participant, a difference score was computed between their attitude toward each of the topics, such that higher scores indicated more attitude change for the message containing nine weak arguments relative to the message containing three strong arguments. To the extent that persuasion by the peripheral route is an indication of System 1 or heuristic thinking (Petty & Cacioppo, 1986), the difference score reflects the level of heuristic thinking. The range of difference scores had a possible range from −6 to +6, and an actual range from −1.75 to +4.00, with $M = 1.01$ and $SD = 1.63$. To allow comparison with the scores from the stereotype assessment task, the difference score was Z-scored.

Following the persuasion task, participants were asked to respond to the cognitive reflection task (Frederick, 2005). The three items were presented as a forced choice with the correct answer and intuitive (but incorrect) answer provided. The number of incorrect responses was counted into a single score, with higher scores reflecting more intuitive (heuristic) thinking. Participants’ actual scores ranged from 0 to 3, with $M = 1.67$ and $SD = .96$. These were Z-scored.

For the stereotype assessment task, participants were presented with 30 different word fragments, 12 of which were stereotypically associated with “chavs.” The critical word fragments were constructed in such a way that they could be solved in more than one manner, of which only one solution was stereotypical. For instance, the word fragment RU_E could be solved as “rude,” but also as “rube,” “rule,” “rune,” “ruse,” and “rute.” All of these solutions are valid English words, but only the former is stereotypical of the target category. The 12 critical word fragments were presented in random order within a list of 30 word fragments. Overall, 98.6% of the word fragments were solved with a correct word. For each participant, the number of words solved in a stereotypical manner was counted, thus resulting in a single overall score. This stereotypicality score had an actual range from 0 to 12, with $M = 7.71$ and $SD = 4.71$. To allow comparison with the scores from the persuasion task, the stereotypicality score was Z-scored.

On completion of the persuasion task or stereotype assessment task, participants were asked to respond to the manipulation check, and fully debriefed.

Results and Discussion

Suppression task. The manipulation check indicated that suppression participants avoided using stereotypes to a greater extent ($M = 5.54$, $SD = 1.41$) than did controls ($M = 3.75$, $SD = 1.82$), $F(1,46) = 30.84$, $p < .001$, $d = 1.64$. Next, examination of the essay ratings showed that suppression participants were indeed less stereotypical in their description ($M = 2.40$, $SD = 1.67$) than controls ($M = 4.17$, $SD = 2.32$), $F(1,46) = 18.67$, $p < .001$, $d = 1.27$.

Assessment task. Half of the participants completed the persuasion task, the other half the word fragment completion task. Given the different metrics of both tasks, the analyses were conducted on the Z-scores. The data were analyzed by means of 2 (condition: suppression vs. control) × 2 (assessment: stereotypes vs. persuasion) full factorial ANOVA. Looking at the main effects first, there was no effect of assessment type ($F < 1$), nor condition, $F(1,44) = 1.69$, $p = .20$. However, the analysis did reveal an interaction of condition by assessment, $F(1,44) = 4.35$, $p < .05$, $n_p^2 = .09$. The mean Z-scores were plotted (see Figure 2). To further decompose the interaction, simple main effects were analyzed. In the word fragment completion task, suppression participants completed more target words in a stereotypical manner ($M = 9.92$, $SD = 3.26$) than did controls ($M = 5.50$, $SD = 5.02$), $F(1,44) = 5.73$, $p < .03$, $d = .72$. No such difference emerged in the persuasion task ($F < 1$). In fact, suppression participants had slightly lower persuasion scores ($M = .83$, $SD = 1.94$) than did controls ($M = 1.19$, $SD = 1.30$).

The ANOVA was repeated comparing the stereotype assessment with the cognitive reflection task. The patterns of results were identical; no main effects were found, but the two-way interaction was significant, $F(1,44) = 5.43$, $p < .03$, $n_p^2 = .11$. The simple main effect analysis showed no difference between suppression participants ($M = 1.50$, $SD = .79$) and controls ($M = 1.83$, $SD = 1.12$) on the cognitive reflection task ($F < 1$). Taken together, these findings suggest that stereotype suppression leads to a stereotype rebound, but not to a more heuristic thinking style overall.

General Discussion

Across five experiments, participants were made to suppress stereotypes or not. The manipulation check and analyses of participants’ written essays showed that suppression...
participants actively suppressed stereotypes resulting in less stereotypical descriptions of a typical day in the life of a target (Macrae et al., 1994). Subsequently, however, suppression participants used more stereotypical language, compared with controls, in an essay about an unrelated target (Experiment 1), suggesting that stereotype rebound (Bodenhausen et al., 2009; Monteith & Mark, 2009) does indeed take place at a procedural level. The occurrence of procedural rebound has been shown for dispositional inferences and similarity use (Geeraert et al., 2011, 2004), but a procedural rebound of stereotypes had not been demonstrated before.

Importantly, stereotype suppression of one particular category led to stereotype rebound regardless whether the target belonged to the same or a different social category (Experiments 2 and 3). Speculatively, these findings could suggest that stereotype PSR may not be operating at the conceptual level at all (Macrae et al., 1994; Wyer et al., 2000), but only at the procedural level. Interestingly, procedural rebound did not only manifest itself in explicit measures but also in implicit measures (Experiment 4). Experiment 5 examined whether stereotype suppression had a more general influence on heuristic thinking. Although suppression participants did display rebound, their ability for cognitive reflection was unaffected (Frederick, 2005), nor did it affect their susceptibility to peripheral cues in a persuasion message (Petty & Cacioppo, 1984). All in all, these findings indicate that stereotype rebound may be more general than initially thought (Bodenhausen et al., 2009; Monteith & Mark, 2009). The conscientious social perceivers trying not to be racist in one context may be inadvertently sexist or ageist in another context.

**Modality and Mechanism**

Interestingly, a procedural modality of stereotype rebound can be informative, but not completely conclusive, with regard to its underlying cognitive mechanism. The absence of evidence for rebound at a conceptual level may potentially put in question the explanations building on Wegner’s (1994) dual-process model of mental control. More specifically, it is uncertain whether a procedural stereotype rebound can be explained due to the cognitive demand of the operating process (Wegner, 1994) or the unintentional priming caused by the monitoring process (Macrae et al., 1994).

Alternatively, procedural rebound may also be the result of ego-depletion or mental fatigue (Muraven & Baumeister, 2000; Muraven et al., 1998). According to this rationale, mental control and suppression are demanding processes that tax a limited cognitive resource. Ego-depletion has been shown to affect future tasks, such that depleted participants may choose a more heuristic approach in subsequent social judgments. Although heuristic processing has been shown to be less cognitively demanding than more elaborate processing (Chaiken & Trope, 1999), it is also more prone to biased judgment and higher stereotype use (Bodenhausen, Macrae, & Sherman, 1999; Kahneman, 2011). Thus, the rationale is that suppression leads to depletion, and that in turn this makes people more biased (prone to use stereotypes) in future judgments. Undoubtedly, this approach may provide a viable alternative explanation for the findings of Experiments 1, 2, and 3. It is less clear, however, to what extent mental fatigue can account for the reaction time speed-up in the primed LDT (Experiment 4). In addition, there was no evidence for the idea that stereotype suppression may lead to more heuristic thinking in the final study (Experiment 5). Elsewhere, it has also been argued that ego-depletion does not adequately explain dispositional rebound (Geeraert & Yzerbyt, 2007).

A procedural explanation of stereotype rebound may perhaps be more easily explained through the motivational inference model (Förster & Liberman, 2001; Liberman & Förster, 2000). These authors suggest that participants may experience difficulty to suppress, which they subsequently attribute to a motivation to use the suppressed construct. In other words, PSR may be caused by participants inferring that they are motivated to use the suppressed concept. Importantly, however, participants are not always explicitly instructed to suppress. For instance, participants have been reported to spontaneously suppress stereotypes in front of a mirror (Macrae, Bodenhausen, & Milne, 1998). Similarly, in some experiments of procedural rebound, participants were never instructed to suppress (Geeraert et al., 2011, Exp 2; Geeraert et al., 2004, Exp 2). If participants are unaware of the suppression, then the motivational inference approach is perhaps less likely.

Obviously, alternative mechanisms are also possible. One interesting explanation that would fit the procedural rebound is a hydraulic model (Lorenz, 1966). The idea of a hydraulic mechanism was conceptualized in the context of aggression. Building on the Freudian notion of catharsis (Dollard, Doob, Miller, Mowrer, & Sears, 1939), humans (and animals) are believed to have a build-up of action-specific energy or a drive to behave in a particular way. This energy or drive will build up or accumulate until a level is reached by which the action occurs spontaneously, provided the necessary external stimuli are present (Lorenz, 1966). According to a hydraulic explanation of rebound, participants would have a drive to respond in a certain (stereotypic) way, when interrupted the need or pressure will build up until the next occasion at which it will released. A critical assumption of this explanation is that the cognitive processes underlying stereotyping are fully automatic, in the sense that people may engage in these processes continuously. After all, if a mental process is not active at the outset, then it cannot be interrupted or suppressed. Thus, it has to be assumed that stereotyping is a largely automatic tendency that people readily engage in. Obviously, this notion resonates with the idea of the cognitive miser (Taylor, 1981) and the literature on automaticity (Barth, Chen, & Burrows, 1996; Devine, 1989). Although these ideas are perhaps relatively underinvestigated, the notion of catharsis still has a role to play in modern theories of social cognition (Denzler, Förster, & Liberman, 2009).
Future Research

Future research should try to further disentangle the three different modalities of PSR. Although the current research unequivocally supports a procedural account of stereotype rebound, the interplay between conceptual and procedural rebound should be further examined by establishing boundary conditions and potential moderators. For instance, the current research does not investigate the role of individual differences (Monteith, Spicer, & Tooman, 1998). In addition, further studies should try to unravel what mechanism underlies procedural rebound of stereotypes, and contrast a featural account of stereotype rebound (Newman et al., 1996) with conceptual or procedural rebound.

This research clearly suggests that stereotype suppression and subsequent rebound are more pervasive than previously thought as suppression of one stereotype is likely to lead to activation of another.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Alternatively, the data could have been analyzed by means of a 2 (essay: 1 vs. 2) × 2 (stereotype category: same vs. different) × 2 (condition: suppression vs. control) ANOVA. Importantly, however, there is no distinction between stereotype category for Essay 1. Thus, the study is not a full 2 × 2 × 2 design. Analyzing the data as a fractional factorial ANOVA may potentially lead to spurious results. For this reason, the analysis strategy was to analyze the stereotypicality data by means of two separate 2 × 2 ANOVAs.

2. Four different booklets were constructed, varying the order of presentation and the three strong and nine weak arguments across topics. There was no effect of order or topic (Fs < 1).

3. The difference score was computed by subtracting participants’ attitude on the topic with three strong arguments from their attitude on the topic with nine weak arguments, that is, \[ \text{Attitude} = \text{attitude} \_9 - \text{attitude} \_3 \].

4. The simple main effect analyses were conducted on the Z-scores. However, for ease of understanding, the actual scores of the word fragment completion and persuasion task are given.

5. To examine whether stereotype suppression had any effect on the persuasion task, a follow-up analysis was conducted. A 2 (condition) × 2 (argument number: 3 vs. 9) mixed model ANOVA revealed a main effect of argument number (p < .01), indicating that overall participants were persuaded more strongly by a higher number of arguments. Crucially, the interaction of condition by argument number did not approach significance (F < 1).

References


Geeraert, N., & Yzerbyt, V. Y. (2007). How fatiguing is dispositional suppression? Disentangling the effects of procedural...


