Social Cognition, Vol. 35, No. 3, 2017, pp. 273–298

SOMETHING IN THE WAY YOU PRIMED ME: BELIEF MONITORING WHEN SOURCE IDENTIFICATION IS NOT POSSIBLE

Ana Sofia Santos and Leonel Garcia-Marques
University of Lisbon, Lisboa, Portugal

Diane M. Mackie
University of California at Santa Barbara

Tomás A. Palma, Rui Soares Costa, and Filipa de Almeida
University of Lisbon, Lisboa, Portugal

It has been shown that subtle contextual primes produce transient changes in stereotypes (Santos et al., 2012), an effect supposedly caused by both activation of the primed trait and failure of belief monitoring. The present research investigated people's ability to avoid the influence of primes. A first pilot experiment used a subliminal-priming paradigm and replicated the contamination found following subtle supraliminal priming (Santos et al., 2012). Experiment 1 made a previous episode of stereotypic assembling highly accessible, immediately before subliminal priming, and found that the primed information ceased to have an effect. Experiment 2 manipulated the diagnosticity of a previous stereotype-assembling episode for stereotype assessment. When the previous assembling episode was perceived as no longer diagnostic of one's beliefs, contamination occurred. The avoidance of mental contamination depends on the accessibility of stereotypic beliefs but also on its assumed diagnosticity. The working stereotype assembled seems to reflect a compromise between contextual contamination and belief monitoring, setting a functional limit on cognitive malleability of stereotypes.

Keywords: working-stereotype assembling, contextual contamination and belief monitoring, monitoring heuristic

The research was supported in part by Portuguese Science and Technology Foundation under the grant to the first author (PTDC/PSI-PSO/111992/2009).

Correspondence should be addressed to Ana Sofia Santos, CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Alameda da Universidade, 1649-013 Lisboa, Portugal; E-mail: sosantos@psicologia.ulisboa.pt

Stereotypes about human groups have the reputation of being resistant to change (see for instance, the famous Princeton trilogy, cf. Devine & Elliot, 1995). Although these knowledge structures can derive from totally illusory sources (Hamilton & Gifford, 1976), once created, they seem to be able to endure against even direct evidence of their inaccuracy (see Hamilton & Sherman, 1994; Hamilton & Trolier, 1986, for reviews).

Recent work has however revealed that stereotypes have other less obvious qualities: They can also be surprisingly flexible and context sensitive (e.g., Garcia-Marques, Santos, & Mackie, 2006; Wittenbrink, Judd, & Park, 2001). Santos and colleagues (Santos et al., 2012) showed that traits primed in a subtle way could be incorporated into stereotypes and thus, momentarily, change them. Such a mechanism compensates stereotypes' self-perpetuating tendencies, allowing stereotypes to change flexibly depending on the context.

But even context sensitivity can be too much of a good thing. Take the case of AD, a case reported by Conchiglia, Rocca, and Grossi (2007). AD suffered cardiac arrest, which damaged the frontal and temporal lobes of his brain. From that episode on, AD's identity became totally context dependent and thus very unstable. To assess the extent of AD's disturbance, Conchiglia and colleagues had actors create a number of different scenarios around AD and gauged his reaction. At a bar, for instance, an actor asked AD for a cocktail, prompting him to immediately fulfill the role of bartender, claiming that he was on a two-week trial but hoping to gain a permanent position. Taken to the hospital kitchen, AD quickly assumed the role of head cook, and said he had to prepare special meals for diabetic patients. He maintained these identities until the situation changed when he would assume his next role. Clearly, extreme context sensitivity is also maladaptive.

In this article, we focus on people's ability to avoid too much context sensitivity in their social beliefs. The processes underlying context dependency of stereotypes cannot be fully understood without examining the complementary process of belief monitoring, namely, the process whereby people "check" whether the activated information is "representative of their beliefs" and "appropriate." This article presents three studies demonstrating that participants are able to avoid the influence of contextually primed information, and the conditions under which they can do so.

CONTEXT-SENSITIVE STEREOTYPES

Recent empirical evidence suggests social stereotypes actually display a considerable degree of malleability (Garcia-Marques, Santos, & Mackie, 2006), paralleling the malleability first demonstrated in non-social category representations (Barsalou, Sewell, & Ballato, 1986; Barsalou, Spindler, Sewell, Ballato, & Gendel, 1987; Bellezza, 1984a, 1984b, 1984c). This malleability seems to depend on the incorporation of currently activated information into the stereotype (Garcia-Marques et al., 2006; Macrae, Bodenhausen, & Milne, 1995; Macrae, Mitchell, & Pendry, 2002; Santos et al., 2012; Wittenbrink, Judd, & Park, 2001). Such context sensitivity is

consistent with a number of findings showing malleability in explicit and implicit stereotypes when the context is manipulated (for a review, see Blair, 2002).

More recent studies (Santos et al., 2012) converged to show that stereotype-unrelated traits may become "one of the best descriptors of the group" when they have been subtly primed by an irrelevant linguistic task. When subjects were asked to list the traits that make up the stereotype of computer programmers, almost no one described them as friendly. But when "friendly" was primed, a large majority now reported that friendly was part of the computer programmer stereotype. The key point in these results is that, in contrast with most stereotype priming procedures (for reviews, see Hewstone, 1994; Richards & Hewstone, 2001), the unrelated information was never associated with group members or with the group as a whole. It was simply made accessible by an unrelated priming task. This feature of the paradigm is important because it ensures that the critical unrelated information was not deliberatively used to revise the stereotype. Instead, unrelated but accessible information was apparently incorporated when the stereotype was accessed. That is, these results represent a classic priming effect.

To explain this effect, in our earlier work, we adopted a constructionist perspective on stereotypes (see Garcia-Marques et al., 2006; Santos et al., 2012). From this perspective, because the information associated with or encoded in knowledge structures like stereotypes is vast and diverse, only part of that information will be available at a given moment according to the perceivers' current goals and tasks. Thus, instead of being activated in all-or-none fashion, information associated with or encoded in a stereotype must be re-assembled whenever needed, creating at that given moment what we have called the "working stereotype" (Santos et al., 2012). A working stereotype is a dynamic online structure that results from an assembling process, composed by the activated part of the information associated with or encoded in a stereotype together with contextually accessible information. Moreover, it is the working stereotype that gives support to and guides relevant processing.

Our proposal assumes that the working stereotype varies across time and context so that, for example, the stereotype of a "Professor" assembled in one context might be different from the stereotype "assembled" in another context because, although information closely associated with the stereotype is more likely to be both activated and incorporated, the re-assembling process is vulnerable to influence at a number of levels. First, given the large amount of information usually associated with a stereotype, stereotype assembly is not expected to be an exhaustive process. Second, other concepts activated in the immediate context become potentially available for incorporation when the stereotype is re-assembled. Such contamination typically happens because people often lack direct introspective access to the source of activation of these contextually activated concepts. In fact, one way of interpreting stereotype malleability is by conceiving stereotype malleability as a side effect of source confusion errors.

A convergent view can be found in the Source Monitoring Framework (SMF; Johnson, 2006), according to which thoughts/images/feelings that come to mind do not include abstract tags or labels that name their sources, but rather have

qualitative and quantitative characteristics that are more or less diagnostic of the source. Thus, people have to base their source attributions such as internal versus external ("did I see it or did it imagine it?"), different external sources ("did I hear it or did I see it?"), and different internal sources ("did I think about it or did I do it?"). According to the SMF, source judgments are based both on the typical features of each of the alternative type of source (for instance, if I did something I should be able to retrieve much more detailed information about the action than if I just thought about doing it) and in the convergence with our other relevant accessible information like other episodic memories, world knowledge, and so forth. Thus, a memory about a conversation between one of us and Mark Twain is easily attributed to a dream even if the mental image of the event is highly vivid and detailed, when we just remember that Mark Twain died before we were born. Note that a similar (dreamt) conversation between one of us and one of our neighbors may well lead to a source error if other retrieved information, or our world knowledge, does not readily suggest it's fantasy.

If we apply these notions to stereotype assembling, it is easy to see why contextually activated stereotype-unrelated information can be mistakenly perceived as representing one's own beliefs as long as discrepant relevant world knowledge or discrepant previous episodes are not readily accessible. In fact, we suggest that beliefs about social groups also have no special markings in memory that name them as valid individual beliefs and, as such, to avoid mental contamination, people need to distinguish beliefs (information that is taken as valid, self-descriptive, and adopted with a feeling of consent) from cultural stereotypes (when they are self-discrepant) and other types of unrepresentative beliefs. From our point of view, people capitalize on subtle indirect cues to infer that a belief is valid, in that it corresponds to a previously held belief, and as a way of belief monitoring not restricted to source monitoring. In that sense, the consistency of available information with our expectations about the features a typical belief has can serve as a cue to infer it corresponds to a previously held belief. And, in a similar vein, enough features sufficiently characteristic of "true" beliefs may lead to belief misattributions.

In our previous work, contextually activated stereotype-unrelated information was indeed incorporated into currently assembled stereotypes (Santos et al., 2012, Experiment 1). It is exactly because of this that belief monitoring is indispensable. We define belief monitoring as the set of processes that function to implement the goal of expressing one's beliefs about a given domain or in acting in terms of those beliefs, and that assesses the success of the implementation of this goal, in terms of consistency over time and/or with other beliefs, and in a way that fits the person's current goals and tasks. However, belief monitoring sometimes fails and the result is usually mental contamination (Wilson & Brekke, 1994).

In fact, and although we used identical priming procedures for each type of prime, stereotype-inconsistent primes were not incorporated in the working stereotypes in our previous work (Santos et al., 2012, Experiment 2). We argued that the incorporation of the primed stereotype-inconsistent information was prevented because, in this case, the contextually primed information ran directly against the participant's relevant world knowledge (i.e., stereotype-consistent informa-

tion) that is readily activated during stereotype assembling. Thus, the simultaneous accessibility of contextually primed stereotype-inconsistent and self-generated stereotype-consistent information seems to be a boundary condition to the malleability of stereotypes. Previous research has already suggested that when the source of activation can be ascertained, priming effects can be blocked or even reversed (Hess, Hinton, & Statham, 2004; Lombardi, Higgins, & Bargh, 1987; Shih, Ambady, Richeson, Fujita, & Gray, 2002; also Strack & Hanover, 1996). However, according to the source-monitoring framework, establishing the external source of a mental event is not the only way to avoid mental contamination. In fact, perceived discrepancies with other equally accessible but more diagnostic evidence may be taken as a cue to discard contextually primed information and avoid mental contamination.

To explore these possibilities, we conducted three experiments in which we increased the accessibility of trait constructs by priming them in a stereotype-unrelated context and measured the contamination by primed information on assembled stereotypes. We used subliminal priming to make the source of priming completely non-identifiable. A pilot experiment adapted a subtle contextual priming paradigm (Santos et al., 2012) to include subliminal priming procedures, to test if monitoring could operate even in conditions in which the source of mental contamination was completely absent. In Experiment 1, we made stereotypic beliefs accessible through a previous task, immediately before a subliminal priming procedure, so that a diagnostic accessible event could block the incorporation of subliminally primed concepts in assembled stereotypes. In that sense, Experiment 1 intends to conceptually reproduce previous findings (see Santos et al., Experiment 1). In Experiment 2, we tested whether accessible stereotype beliefs prevent mental contamination because a thorough scrutiny of one's own beliefs takes place or because a simple monitoring heuristic makes one's own beliefs distinctive in the experimental setting.

PILOT EXPERIMENT

As a necessary prerequisite for studying alternative means of discarding contextually primed information, and avoiding mental contamination, that do not depend on source identification, we needed to develop a paradigm in which the source of activation of information is completely unidentifiable. We thus primed concepts subliminally and checked whether such priming nevertheless produced the stereotype contamination results obtained in previous experiments with supraliminal priming (Santos et al., 2012).

The pilot experiment used a parafoveal subliminal priming procedure (Bargh & Pietromonaco, 1982) to activate concepts related to intelligence or friendliness. The priming procedure involved an initial vigilance task, in which participants pressed a button as quickly as they could in reaction to "flashes" appearing on a computer screen. Depending on the condition, trials contained either positive friendliness or positive intelligence traits (Garcia-Marques, Ferreira, Nunes, Garrido, & Gar-

cia-Marques, 2010; Rosenberg, Nelson, & Vivekananthan, 1968). The friendliness trait words were stereotypic of childcare professionals but stereotype unrelated for computer programmers, whereas the intelligence traits were stereotypic of computer programmers but stereotype unrelated to childcare professionals (Ferreira, Garcia-Marques, Toscano, Carvalho, & Hagá, 2011). Following this priming procedure, we assessed participants' stereotypes of either childcare professionals or computer programmers using a trait checklist task (Katz & Braly, 1933).

We expected subliminal exposure to an unrelated trait in an irrelevant context to increase the contamination by associated unrelated traits of the stereotype activated immediately after. These results would thus replicate the findings previously obtained with supraliminal priming (Santos et al., 2012, Experiments 1 and 2), even though the subliminal priming excluded any possible awareness of exposure, and any possibility of intentional contamination.

METHOD

Participants and Design

The participants were 80 Lisbon University students (43 females; M_{age} = 23.25, SD = 2.64), who volunteered for the study as partial fulfillment of course requirements. Participants were randomly assigned to the cells of a 2 (prime: intelligence or friendliness) × 2 (target group: computer programmers or childcare professionals) between-subjects factorial design.[1]

Procedure

Participants were tested in small group sessions of up to 10 people. On arrival, participants were seated in individual booths in front of computer monitors and told they would complete two separate studies: a vigilance study and a study on impressions of social groups.

Subliminal Priming Task. Participants were told that the perceptual vigilance task intended to explore how quickly and accurately people can identify the location of very briefly presented visual stimuli. Participants were instructed to maintain their gaze on an initially presented fixation point throughout the task, and to respond as quickly and as accurately as possible to flashes displayed on the screen. They should press one of two keys (Q, labeled LEFT and P, labeled RIGHT) to indicate whether the flash had occurred to the left or right of the fixation point.

The task was composed of four blocks of 25 trials (100 trials total), with each trial composed of the following sequence of events: (i) a cross (+) appeared on the cen-

1. The sample size for each study was determined beforehand based on our previous research in which we have used similar experimental designs (see Santos et al., 2012). Additionally, we inform that, in all experiments, the data was collected in one shot without prior statistical analyses and that we report all data, all manipulations, and all measures. Materials and data are available from the authors upon request.

ter of the screen for 1000 ms; (ii) a flash, composed of a prime (presented for 80 ms) followed by an 11-letter string mask (presented for 100 ms), appeared randomly in one of four parafoveal quadrants of the monitor; (iii) a blank response screen appeared for 3500 ms; (iv) finally the screen went black for between 2000 ms to 6500 ms between trials and for 7500 ms between blocks.

Manipulation of Primes. Participants were subliminally exposed to either 12 positive trait words perceived as closely associated with friendliness (generous, affectionate, humorous, friendly, sensitive, steadfast, kind, nice, reliable, sociable, happy, helpful) or 12 positive trait words closely associated with intelligence or intellectuality (skillful, motivated, determined, competent, clever, dynamic, methodical, efficient, quick, creative, insightful, rational).[2] Three neutral high frequency words (always, space, and during) unrelated to the stereotypes of either target groups were selected from a Portuguese word frequency database (Bacelar do Nascimento, Pereira, & Saramago, 2000) and added to each list.

Following 10 practice trials using other neutral words, participants responded to five blocks of 20 test trials. Within each block, participants responded to 20 words, 16 target primes (friendliness words in one condition and intelligence words in the other), and 4 neutral words. To make 16 target primes, each of the 12 target primes appeared once and four of the 12 target primes were randomly selected and presented twice. In a similar way, to make 4 neutral words, the 3 neutral words appeared once and one of the 3 neutral words was randomly selected and presented twice. Words were presented in a random order.

Stereotype Trait Selection Task. After the vigilance task, participants were told that the next study was about how people form impressions of groups. Following Katz and Braly (1933), we had participants select the 5 traits that best described one of the two social groups from a 42-item trait checklist (see Appendix A) that included 10 key items. To assess incorporation of the primed material in the stereotypes, these key items included three synonyms of friendliness traits used as primes (thoughtful, tactful, attentive) and two other friendliness traits not used as primes (warm, understanding), and three synonyms of primed intelligence traits (resolved, capable, methodical) and two other intelligence traits not used as primes (cultured, organized).

Stereotype Trait Descriptiveness Ratings. Participants then rated how descriptive of the group the 12 different traits were, using 9-point scales anchored by the trait and its antonym. Eight items were critical. Four items assessed traits highly associated with friendliness (two synonyms of presented primes—thoughtless vs. *thoughtful*; hostile vs. *tactful*—and two words not used as primes—cold vs. *warm*; intolerant vs. *understanding*). Four items assessed traits highly associated with intelligence (two synonyms of presented primes—incapable vs. *capable*; undecided vs. *resolved*—and two words not used as primes: ignorant vs. *cultured*; unorga-

2. Previous pretested participants generated traits for either computer programmers or childcare professionals. Inspection of these lists indicated that the trait friendly is stereotype unrelated for computer programmers but is stereotypic of childcare professionals (mentioned by 45% of the participants). The trait intelligent is stereotypic of computer programmers (mentioned by 50% of the participants) but stereotype unrelated to childcare professionals (see also Garcia-Marques & Mackie, 1999; Santos, 2001, 2007). Further pretesting confirmed that the two traits were equally positive.

nized vs. *organized*). Other items comprised four stereotype-unrelated traits from the original list (ugly vs. attractive, unfaithful vs. faithful, poor vs. rich, fragile vs. strong).

After completion of the two stereotype response tasks, an individual funneled debriefing took place (see Appendix C for further details on this procedure). Participants' observations were recorded, particularly those pertaining to a connection between priming and the stereotype response tasks. No one reported that the screen blinking during the priming task influenced performance on the stereotype assessment task, presented as an independent experiment (Study 2) for participants (see Appendix C, question 3).

RESULTS AND DISCUSSION

Check on the Subliminal Priming Procedure. To ensure that prime presentation was subliminal, we ran an additional 27 participants and assessed their conscious awareness of prime content. In a *guessing* condition, 14 participants were told, before they performed the vigilance task, that the flashes they would see were actually words and that they should try to guess which exact words these were. Following Bargh and Pietromonaco (1982, Experiment 2), participants were encouraged and prompted to take a blind guess at the content of the word if necessary. Although these participants reported during and after the task that they did not know what the words actually were, they nevertheless followed instructions by making guesses on almost all trials. From the total of 336 trials, one answer matched the correct guess, that is, .997 of the guesses were incorrect. Two independent judges, blind to the experimental conditions, examined incorrect guesses for their relatedness to the primed concepts. Judges' ratings revealed no words related to friendliness or intelligence, suggesting participants were not able to identify the content of the priming words, and thus satisfying one criterion for unconscious processing.

As a more sensitive test of awareness, another 13 participants in a *recognition* condition completed the priming task. They were first alerted to the fact that the flashes they had seen on the screen were in fact words, and then repeated the priming task, this time selecting on each trial the word (from three options) they thought had just appeared. Each prime appeared twice in the 24 trials of the recognition test, giving participants two opportunities to detect the primes. Distracters were matched in length to the target words and were similar to these either in meaning or phonetically. The mean proportion of correct choices was .339, not significantly different from chance value (.333). No participant reported being aware of any word, or correctly guessed any relation between the priming task and the stereotype task.

Stereotype Assessment. Depending on the social group participants were asked to describe, the friendliness and intelligence primes were stereotype consistent or unrelated. Participants who described computer programmers after exposure to the intelligence primes and participants who described childcare professionals after exposure to the friendliness primes were in the stereotype-consistent prime condition. Responses were thus collapsed across these two groups. Participants

TABLE 1. Number of Participants Choosing Stereotype-Unrelated Traits, By Priming Condition—Pilot Experiment

	St-consistent Prime	St-unrelated Prime
St-unr trait chosen	4	22
St-unr trait not chosen	36	18

who described computer programmers after exposure to the friendliness primes and participants who described childcare professionals after exposure to the intelligence primes were in the stereotype-unrelated prime conditions. Responses were thus collapsed across these two groups. Before collapsing across social groups, we checked whether both groups exhibited the same frequency pattern. Indeed, we verified that, in both groups, more participants chose stereotype-unrelated traits under the stereotype-unrelated prime condition while the same did not happen in the stereotype-related prime condition.

We then counted the number of participants by condition that selected any of the five friendliness traits (the three synonyms of the friendliness traits primed or the two unprimed friendliness traits) and any of the five intelligence words (the three synonyms of the intelligence traits primed or the two unprimed intelligence traits). Table 1 shows participants' unrelated-trait choices as a function of stereotype-consistent and stereotype-unrelated prime condition.

In order to test our hypotheses, we ran a Fischer exact test. The test on this data showed that priming the stereotype-unrelated traits increased their choice as best descriptor of the group: $p = .0000$ (one tailed). Therefore, these results replicated the findings previously obtained (Santos et al., 2012, Experiments 1 and 2).

Trait Descriptiveness. First, we combined the responses to the two synonyms of the friendliness traits used as primes with the responses to the two non-primed friendliness traits to form a single friendliness index (Cronbach's alpha = .932). The responses to two synonyms of the intelligence traits primed and the responses to the two non-primed intelligence traits were also combined to form a single intelligence index (Cronbach's alpha = .787). Then, as the social group did not predict the expected interaction, $F(1, 76) = .0003$, $p = .986$, $MSe = 1.32$, $\eta_p^2 = .00$, we collapsed across social groups. We computed a 2 (stereotype-consistent trait vs. stereotype-unrelated trait primed) × 2 (ratings of stereotypic and stereotype-unrelated traits) ANOVA, the last factor being within-participants. We obtained two main effects. A main effect of prime indicating that the stereotype-unrelated prime lead to higher ratings for the two types of trait scales ($M = 6.75$) than the stereotype-consistent prime ($M = 5.86$), $F(1, 78) = 15.66$, $p < .001$, $MSe = 2.05$, $\eta_p^2 = .17$. A main effect of trait type showing that stereotypic traits were seen as more descriptive of the group ($M = 7.21$) than stereotype-unrelated traits ($M = 5.39$), $F(1, 78) = 102.55$, $p < .001$, $MSe = 1.29$, $\eta_p^2 = .57$. A significant interaction was also obtained, $F(1, 78) = 4.42$, $p = .039$, $MSe = 1.29$, $\eta_p^2 = .05$, showing that both types of traits were more affected by stereotype-unrelated primes than by stereotype-related primes, although in different degrees. Namely, results showed that stereotype-related traits were seen as more descriptive of the target group after being primed with stereotype-unrelated primes ($M = 7.47$) than after being primed stereotype-related primes ($M = 6.96$), $t(78) = 2.20$, $p = .031$, Cohen's $d = .50$, 95% CI [0.05, -0.94]. A similar but stronger

results' pattern was also obtained for stereotype-unrelated, that is, traits were seen as more descriptive of the target group following the stereotype-unrelated prime (M = 6.03) than after the stereotype-consistent prime (M = 4.76), $t(78) = 3.81$, $p < .001$, Cohen's $d = .86$, 95% CI [0.39, -1.31].

Thus, subliminal exposure to an unrelated trait in an irrelevant context increased the contamination by associated unrelated traits of the stereotype activated immediately afterward. That is, subliminal exposure to unrelated traits such as generous and happy increased the extent to which computer programmers were seen as warm and understanding. Importantly, this contamination occurred even though the subliminal priming excluded the possible awareness of exposure, and even though participants rated the target group on associated but not previously presented traits. Thus, subliminal priming resulted in contamination that replicated the contamination previously found following subtle supraliminal priming (Santos et al., 2012). Such findings provide even more compelling evidence that stereotypes are vulnerable to contextual contamination and are consistent with our reasoning that contamination is most likely to be found when source monitoring, or awareness of why or how particular material has become activated, is very difficult. These findings also indicated that we were able to create conditions in which source identification was rather improbable, which allowed us to explore means other than source identification by which such contamination might be resisted.

EXPERIMENT 1

As suggested earlier, it may be possible to avoid contextual contamination even when it is impossible to identify the (actual) external source of activation, because other more diagnostic information about one's own beliefs is highly accessible. Thus, an activated unrelated trait may be ignored if it is discrepant with the output of a previous stereotype-assembling episode, particularly, if that episode is considered more diagnostic of one's own true beliefs. In this case, perceived true beliefs can override contextually primed information.

To ascertain whether accessibility of a previous stereotype-assembling episode could moderate contamination from contextual activation, we created conditions in which participants were completely unaware of contextual activation and then manipulated the accessibility of a previous episode of stereotype assembling of the target group. To do so, we used the parafoveal subliminal priming procedure from the pilot experiment to contextually activate either stereotype-consistent or stereotype-unrelated information about one of two groups (computer programmers or childcare professionals) outside participants' conscious awareness. In addition, we manipulated the accessibility of a previous stereotype-assembling episode. When a previous stereotype-assembling episode was relatively low in accessibility, we expected contamination by the subliminally primed material, indicating the vulnerability of stereotypes to contamination when source monitoring is difficult. When a previous stereotypic assembling episode was highly accessible, we expected contamination to be eliminated. In this condition, the accessibility of perceived stereotypic beliefs would make the contextually generated informa-

tion seem discrepant even though source monitoring was not possible. Thus, we wished to demonstrate one condition in which a form of belief monitoring other than source identification might protect stereotypes from contextual intrusions.

METHOD

Participants and Design

One hundred six Lisbon University students (62 females; M_{age} = 20.11, SD = 4.54) volunteered for the study in partial fulfillment of course requirements. Participants were randomly assigned to the cells of a 2 (accessibility of previously endorsed beliefs about the target group: high accessibility or non-accessibility) × 2 (prime: intelligence or friendliness) × 2 (target group: computer programmers or childcare professionals) between-subjects factorial design.

Procedure

The procedure closely followed the pilot experiment: participants first performed a vigilance study (which served to manipulate subliminally activated primes) and then offered their impressions of a social group (which served as the stereotype-assessment dependent measure). Before completing these two "separate" studies, participants were asked to perform one of two tasks intended to manipulate the accessibility of previously endorsed beliefs about the target group.

Stereotype Accessibility. Participants were asked to familiarize themselves with the list of traits later used in the stereotype-assembling task. In the high accessibility condition, participants were asked to imagine the traits they would use to describe the target group. In the non-accessibility condition, participants assessed the frequency with which each trait appeared in the children's literature. These tasks were followed by the vigilance task (i.e., the subliminal priming task) and by the stereotype-assessment task (in which participants are asked to choose the five traits that best describe the target group), using the same materials and following the exact same procedure of the pilot study.

RESULTS AND DISCUSSION

Check on the Subliminal Priming Procedure. To ensure that prime presentation was subliminal, we ran an additional 14 participants in conditions designed to assess their conscious awareness of prime content. Seven participants in the *guessing* condition were informed that the flashes they would see during the vigilance task were actually words and they were instructed to try to guess each word as it was presented. These participants reported during and after the task that they did not know what the words actually were, but they nevertheless followed instructions by making guesses on almost all of the trials. There were no correct guesses. Ratings from two independent judges blind to the experimental conditions who examined the incorrect guesses for their relatedness to the primed concepts indicated that participants were not able to identify the content of the priming words.

TABLE 2. Number of Participants Choosing Stereotype-Unrelated Traits, by Priming and Stereotype Accessibility Conditions—Experiment 1

	Non-accessibility		High Accessibility	
	St-consistent Prime	St-unrelated Prime	St-consistent Prime	St-unrelated Prime
St-unr trait chosen	6	18	9	9
St-unr trait not chosen	24	10	15	15

Seven participants in a *recognition* condition first completed the priming task, were alerted then to the fact that the flashes they had seen on the screen were in fact words, and then repeated the priming task, this time selecting on each trial the word (from three options) that they thought had just appeared. The mean percentage of correct choices (0.329) did not differ significantly from chance value (0.333).

No participant in the actual study reported being aware of any word or correctly guessed any relation between the priming and stereotype tasks.

Stereotype Assessment. Again, we first checked whether the priming manipulation had the same effect on both groups. We verified that, for both groups, in the non-accessibility condition, the frequency pattern revealed that more participants chose stereotype-unrelated traits when primes were stereotype unrelated than when primes were stereotype consistent. Regarding the high accessibility condition, we observed that, for both social groups, the number of participants who chose stereotype-unrelated traits was exactly the same in the two priming conditions. Because of these similarities between both social groups we collapsed their data as described in the pilot experiment (see Table 2 for frequencies). We then ran a Chi-square test on the collapsed data that indicated differences among conditions not expected by chance, χ^2 (3) = 12.34, p = .003, Cramer's V = .34. Follow-up standard residual analysis (Everitt, 1977) showed that the failure of the independence assumption was due to the high frequency of choice of the stereotype-unrelated traits when primes were stereotype unrelated and the stereotype was low in accessibility (n = 18 compared to the other three conditions with ns respectively of 6, 9, and 9), z = 3.84, p < .001, and to the high frequency with which the stereotype-unrelated traits were not chosen when the primes were stereotype consistent and the stereotype was low in accessibility (n = 24, compared to the other three conditions with ns respectively of 10, 15, and 15), z = -3.25, p =. 000.

As predicted, contamination was maximal when an unrelated trait was contextually primed and the internally generated stereotype was low in accessibility, converging with results from Santos et al. (2012). When the internally generated stereotypes were highly accessible, however, contamination was blocked. Thus, high accessibility of internally generated beliefs appears to set a boundary condition on contamination from contextual activation.

Trait Descriptiveness. First, we combined responses to the two non-primed friendliness items and to the two synonyms of the friendliness words used as primes to form a single friendliness index (Cronbach's alpha = .899), and responses to the two non-primed intelligence items and to the two synonyms of the intelligence words used as primes to form a single intelligence index (Cronbach's alpha = .745).

TABLE 3. Descriptiveness Ratings of Stereotypic and Stereotype-Unrelated Traits, By Priming and Stereotype Accessibility Conditions—Experiment 1

	Non-accessibility		High Accessibility	
	St-consistent Prime	St-unrelated Prime	St-consistent Prime	St-unrelated Prime
St trait scale	7.33 (.27)	7.39 (.28)	6.90 (.30)	7.25 (.30)
St-unrelated trait scale	5.58 (.29)	6.39 (.30)	5.12 (.33)	4.98 (.33)

Note. Numbers in parentheses are standard errors. $N = 106$.

We then checked whether there was a four-way interaction between social groups, accessibility conditions, priming conditions, and type of trait. Results showed a nonsignificant interaction, $F(1, 98) = .72$, $p = .397$, $MSe = 1.27$, $\eta_p^2 = .01$. Because the groups displayed the same pattern we combined their data.

We computed a 2 (high vs. non-accessibility of previously endorsed beliefs about the target group) × 2 (stereotype-consistent trait vs. stereotype-unrelated trait primed) × 2 (ratings of stereotypic and stereotype-unrelated traits) ANOVA, the last factor being within-participants. We present the relevant means for this analysis in Table 3. The stereotypic traits were "intelligent" for computer programmers and "friendly" for childcare professionals, whereas the stereotype-unrelated traits were "intelligent" for childcare professionals and "friendly" for computer programmers. Two main effects, a two-way interaction, and a marginal three-way interaction emerged. One main effect indicated that overall descriptiveness ratings were higher under low than under high accessibility conditions ($M = 6.68$ vs. $M = 6.06$), $F(1, 102) = 5.87$, $p = .017$, $MSe = 3.36$, $\eta_p^2 = .054$. A second main effect showed that, as expected, stereotypic traits were seen as more descriptive of the group than stereotype-unrelated traits ($M = 7.22$ vs. $M = 5.52$), $F(1, 102) = 109.55$, $p < .001$, $MSe = 1.38$, $\eta_p^2 = .518$. The two-way interaction, $F(1, 102) = 3.96$, $p = .049$, $MSe = 1.38$, $\eta_p^2 = .037$, revealed that, as predicted, the manipulation of high versus non-accessibility had no effect on descriptiveness ratings of stereotypic traits ($M = 7.07$ vs. $M = 7.36$), $t(102) = 1.00$, $p = .318$, Cohen's $d = .20$, 95% CI [-0.19, -0.58], but had an effect on descriptiveness ratings of the stereotype-unrelated traits ($M = 5.05$ vs. $M = 5.99$), $t(102) = 3.01$, $p = .003$, Cohen's $d = .59$, 95% CI [0.20, -0.98]. Finally, the marginal three-way interaction, $F(1, 102) = 3.71$, $p = .057$, $MSe = 1.38$, $\eta_p^2 = .035$, suggested that, under high accessibility, it didn't matter whether stereotype-consistent versus stereotype-unrelated traits were primed because stereotypic traits ($M = 6.89$ vs. $M = 7.25$) did not differ from each other, $t < 1$, as well as the stereotype-unrelated traits that were always considered equally descriptive following stereotype-consistent versus stereotype-unrelated traits primed ($M = 5.12$ vs. $M = 4.98$), $t < 1$. However, under non-accessibility conditions whereas we found a similar absence of priming effects for stereotypic traits ($M = 7.33$ vs. $M = 7.39$), $t < 1$, stereotype-unrelated traits were considered much worse descriptors when stereotype-consistent ($M = 5.58$) than when stereotype-unrelated primes were used ($M = 6.39$), $t(102) = 1.93$, $p = .056$, Cohen's $d = .52$, 95% CI [-0.02, -1.03]. More importantly, our main prediction was that priming would be much more effective in the case of stereotype-unrelated primes under non-accessibility conditions in the rat-

ings of stereotype-unrelated traits. To test this specific hypothesis, we contrasted the descriptiveness ratings of stereotype-unrelated traits in the non-accessibility/stereotype-unrelated prime condition with the other three cases (i.e., non-accessibility/stereotype-consistent prime, high accessibility/stereotype-unrelated prime and high accessibility/stereotype-consistent prime). As predicted, stereotype-unrelated traits were rated as more group descriptive in the non-accessibility/stereotype-unrelated prime condition ($M = 6.39$) than in the other three conditions ($M = 5.55$), $t(102) = 3.31$, $p = .001$, Cohen's $d = .65$, 95% CI [0.29, -1.17]. The same contrast performed on the ratings of the stereotypic traits was not significant, $t < 1$.

In sum, Experiment 1 replicated the findings of the pilot experiment when (as was probably true in the pilot experiment) previously endorsed stereotypic beliefs had low accessibility. When this was true, the stereotype was contaminated by the contextually activated unrelated information. In contrast, contamination from the contextually activated stereotype-unrelated information was much reduced when stereotypic beliefs were highly accessible. This was the case even when participants had no awareness of the source of activation of the stereotype-unrelated information, a circumstance that typically leads to the misattribution to one's own internal thought process. That is, even when primed information is easily confused with one's own response to a situation, its influence seems to depend on other cues about the validity of the accessible information. Source identification is therefore one means by which belief monitoring can exclude contamination, but belief monitoring is not restricted to source monitoring. In this experiment, even though participants could not identify the (subliminal) source of priming, they were nevertheless able to resist its influence when their own beliefs were recently activated. Note again that these ideas are akin to the ideas from the source-monitoring framework (Johnson, 2006; Johnson, Hashtroudi, & Lindsay, 1993).

EXPERIMENT 2

Experiment 1 suggested that even when it is impossible to identify an external source of activation, it might be possible to avoid contextual contamination as long as diagnostic belief-related information is already accessible (when group target-relevant beliefs were assembled exhaustively in the absence of perceived sources of contamination). We assumed that when supposedly diagnostic beliefs associated with the target group have already been made highly accessible, contamination would be eliminated because, in this condition, the accessibility of valid information would make the contextually generated information seem discrepant and would probably trigger the need for monitoring and editing of information.

However, the fact that a contextually activated unrelated trait was more likely to be ignored when stereotype-consistent traits were highly accessible, it does not directly tell us what are the mental processes involved or, if a monitoring mechanism is at work, how it operates. One possibility is that there is no conflict detection or monitoring at all, but that it is the subliminal priming that is relatively weaker in

terms of making the primed trait accessible for deliberate use than a previous assessment of the stereotype made recently. When the stereotype is made accessible by the previous stereotype assessment task, that information is more accessible for the later judgment and somehow overshadows subliminally primed traits; when the stereotype is not previously assessed, the subliminal primes are sufficiently accessible (with no overshadowing) to contaminate the working stereotype. To discard this interpretation, in the next experiment we will create conditions in which a previous stereotype assessment occurs in two different conditions but that differs in terms of diagnostic value for monitoring.

But even if we accept the obtained results are due to monitoring, the nature of this monitoring is far from established. At least two alternative explanations seem plausible. One would be that the detection of a discrepancy between a previous assembling episode and the currently accessible information might trigger a relatively thorough search through one's network of beliefs, trying to find out whether the currently accessible information in one's mind fits with the constellation of related beliefs that are retrievable in that point in time (i.e., "when I think of members of this group, intelligent comes to mind but I know that members of this group tend to avoid mathematics and science subjects in general"). If the currently accessible information would poorly fit with other related beliefs, this currently accessible information would not be considered part of one's beliefs and mental contamination would be avoided. On the other hand, the discrepancy itself might be a sufficient basis for rejecting the currently accessible traits because the traits chosen in the earlier task did not include those. Thus, the discrepancy between the traits chosen in the previous task and the presently accessible traits may be used as a simple heuristic such that only the traits chosen in both assessments are incorporated in the working stereotype and the discrepant ones are rejected. Again, mental contamination would in this case be avoided.

Israel and Schacter (1997) suggested that a similar monitoring heuristic, one that relied on perceived distinctiveness, helped reduce false memories. People expect to remember critical details of an experience as distinctive and thus use distinctiveness to make recognition decisions. If an item presented in the study list seems in some way distinctive from non-presented items, people accept it as having occurred. If an item or event lacks the expected distinctive information, people use this absence to reject the item as having occurred. However, if the presence or absence of distinctiveness is made non-diagnostic of occurrence, then false memories re-emerge (Schacter, Israel, & Racine, 1999).

In a similar way, participants in Experiment 1 might have compared a currently accessible (because it had been primed) trait with the traits they had reported earlier as their beliefs in the initial stereotype assessment task. If the currently accessible trait had not been chosen earlier, it could safely be rejected as a non-endorsed belief and excluded because nothing in the previous assessment task seemed to suggest that outcome of the task was biased. Such a comparison process thus operates as a simple heuristic for excluding other than internally generated information without the need of a more systematic and exhaustive memory search.

To ascertain whether belief monitoring took the form of an exhaustive search through one's own belief network or a simple heuristically driven decision that excludes traits that did not appear in both assessments, participants were first asked to produce various trait descriptions, then they went through our subliminal priming paradigm, and then we again assessed their stereotypes. When participants had just accessed their (uncontaminated) stereotype, we predicted that they would not be affected by primed material, consistent with the results of Experiment 1. In the critical condition, however, participants had to choose, in the first task, traits that either described the target group or an unrelated group. Thus, when participants had their stereotype assessed, the choice of a given trait in the previous task was no longer diagnostic of one's beliefs (because the trait could have been chosen to describe the unrelated group). If the inconsistency between previously generated beliefs and currently activated information generates thorough scrutiny, then the diagnosticity of previous trait choice would not matter and mental contamination should still be avoided. However, if the successful monitoring found in Experiment 1 was due to a heuristic rejection of traits that do emerge in two comparable diagnostic assessments of the stereotype, making the previous task non-diagnostic of one's own beliefs should allow mental contamination to re-emerge. If mental contamination occurred in the low diagnostic monitoring experimental condition, it would constitute evidence for the role of heuristic monitoring (Schacter et al., 1999).

We used the same parafoveal subliminal priming procedure used in the previous experiments and the same stereotype accessibility manipulation used in Experiment 1. In addition, we manipulated the diagnosticity of trait choice for self-belief assessment. When stereotypic beliefs were relatively low in accessibility, we expected to replicate the results from Experiment 1, that is, contamination by the subliminally primed material. When stereotypic beliefs were accessible and diagnostic, we expected contamination to be eliminated. This condition replicated the high accessibility condition used in Experiment 1. However, we expected that if a monitoring heuristic were used, even the presence of highly accessible stereotypic beliefs would not be able to prevent contamination, if trait choice in the previous task was made non-diagnostic. As so, no difference results would appear between the non-accessibility and the non-diagnostic accessibility conditions (two conditions we would expect to be susceptible to contamination).

METHOD

Participants and Design

One hundred twenty Lisbon University students (81 females; M_{age} = 21.13, SD = 3.59) volunteered for the study in partial fulfillment of course requirements. Participants were randomly assigned to the cells of a 3 (diagnostic accessibility of previously endorsed beliefs about the target group: non-accessibility, diagnostic-accessibility, or non-diagnostic accessibility) × 2 (prime: intelligence or friendliness)

× 2 (target group: computer programmers or childcare professionals) between-subjects factorial design.

Procedure

Stereotype Accessibility. In the task intended to manipulate the degree and diagnosticity of previously activated traits, participants were familiarized with the list of traits they would later use to describe the stereotype of a group. In the non-accessibility condition, participants estimated the frequency with which each trait appeared in children's literature. In the diagnostic-accessibility condition, participants were asked to imagine which traits describe the target group of the stereotype-assessment task. As so, it was unlikely that participants would choose the critical primed trait (unrelated to the stereotype of the target group) to describe the target group in this initial task. In this case there was a diagnostic discrepancy between accessibility due to priming during the stereotype-assessment task and not having chosen the same critical trait in the previous stereotype-accessibility task. In the non-diagnostic accessibility condition, participants were asked to imagine which traits describe two groups: the target group (e.g., computer programmers) and another group (usually described with different traits, e.g., childcare professionals). As so, it was most likely that participants would choose the critical primed trait (unrelated to the stereotype of the target group) to describe the other group in this initial task. In this case there was no discrepancy between accessibility of the critical trait due to priming during the stereotype-assessment task and having chosen the same critical trait in the previous stereotype-accessibility task. The list included the most common traits associated with each of the groups.

After the stereotype-accessibility task, participants always performed the vigilance study, which served to subliminally activate different primes, and then completed the stereotype-assessment dependent measure from Experiment 1. Note that for the two groups we used, the target group and critical primed traits were counterbalanced.

RESULTS AND DISCUSSION

Stereotype Assessment. As in the previous experiments, we first inspected whether both social groups exhibited a similar pattern of frequencies. Indeed, in non-accessibility and the non-diagnostic accessibility conditions, we observed that more participants chose stereotype-unrelated traits in the stereotype-unrelated prime condition than in the consistent-prime condition. In the diagnostic accessibility conditions this pattern was not observed. Given the similarity between both social groups we collapsed their data as described in previous experiments (see Table 4). Because we predicted no differences between the non-accessibility and the non-diagnostic accessibility conditions (two conditions we expected to be susceptible to contamination) and the data was nearly identical for these two conditions, $\chi^2(1) = .01$, $p = .91$, Cramer's $V = .02$, we collapsed across them in the analysis.

A 2 (stereotype-consistent vs. stereotype-unrelated prime) × 2 (stereotype-unrelated trait chosen vs. not chosen) × 2 (diagnostic accessibility vs. contamination susceptible conditions) Chi-square test of independence was computed, $\chi^2(3) = 13.52$, $p = .009$, Cramer's $V = .41$, indicating differences across conditions not ex-

TABLE 4. Number of Participants Choosing Stereotype-Unrelated Traits, By Priming and Stereotype Accessibility Conditions—Experiment 2

	Non-Accessibility		Diagnostic Accessibility		Non-diagnostic Accessibility	
	St-consistent Prime	St-unrelated Prime	St-consistent Prime	St-unrelated Prime	St-consistent Prime	St-unrelated Prime
St-unr trait chosen	5	12	4	5	5	11
St-unr trait not chosen	15	8	16	15	15	9

pected by chance. Follow-up standard residual analysis (Everitt, 1977) showed that the failure of the independence assumption was due to the high frequency of inclusion of the stereotype-unrelated trait when the prime was stereotype unrelated (n = 23 compared to the other three conditions with ns respectively of 10, 4, and 5), z = 5.01, $p < .001$, and to the high frequency with which the stereotype-unrelated trait was not included when the prime was stereotype consistent, but only in contamination susceptible conditions (n = 30 compared to the other three conditions with ns respectively of 16, 17, and 15), $z = -4.51$, $p < .001$.

To further explore our hypothesis, we performed a direct comparison between the diagnostic and the non-diagnostic accessibility conditions. As can be seen in Table 4, only 5 out of 20 participants (i.e., 25%) in the diagnostic accessibility condition who were primed by stereotype-unrelated primes chose unrelated traits, whereas in the non-diagnostic accessibility condition 11 out of 20 participants (i.e., 55%) chose such traits, $\chi^2(1) = 3.75$, $p = .053$, Cramer's $V = .31$.

As predicted, contamination was maximal when an unrelated trait was contextually primed and the stereotypic beliefs were low in accessibility or lacked diagnosticity, converging with the results from Santos et al. (2012). When an unrelated trait was contextually primed but the stereotypic beliefs were both accessible and diagnostic, however, contamination didn't occur.

Trait Descriptiveness. Once more, we combined responses to the two synonyms of the friendliness words used as primes and to the two non-primed friendliness items to form a single friendliness index (Cronbach's alpha = .913), and combined responses to the two synonyms of the intelligence words used as primes and to the two non-primed intelligence items to form a single intelligence index (Cronbach's alpha = .650). We then checked whether there was a four-way interaction between social groups, accessibility conditions, priming conditions, and type of trait. Results showed a nonsignificant interaction, $F(2, 108) = .35$, $p = .702$, $MSe = .98$, $\eta_p^2 = .01$. Because the groups displayed the same pattern we combined their data.

We computed a 3 (non-accessibility vs. diagnostic-accessibility vs. non-diagnostic accessibility) × 2 (stereotype-consistent trait vs. stereotype-unrelated trait primed) × 2 (ratings of stereotypic and stereotype-unrelated traits) ANOVA, the last factor being within-participants. The relevant means appear in Table 5. The stereotypic traits were "intelligent" for computer programmers and "friendly" for childcare professionals whereas the stereotype-unrelated traits were "intelligent" for childcare professionals and "friendly" for computer programmers. One main

TABLE 5. Descriptiveness Ratings of Stereotypic and Stereotype-Unrelated Traits, By Priming and Stereotype Accessibility Conditions—Experiment 2

	Non-Accessibility		Diagnostic Accessibility		Non-diagnostic Accessibility	
	St-consistent Prime	St-unrelated Prime	St-consistent Prime	St-unrelated Prime	St-consistent Prime	St-unrelated Prime
St trait scale	6.88(.27)	7.63(.27)	7.61(.27)	7.41(.27)	7.50(.27)	7.30(.27)
St-unrelated trait scale	5.59(.32)	6.38(.32)	5.89(.32)	5.28(.32)	4.84(.32)	5.80(.32)

Note. Numbers in parentheses are standard errors. $N = 120$.

effect, a two-way interaction, and a three-way interaction emerged. The main effect indicated that, as expected, stereotypic traits were seen as more descriptive of the group than stereotype-unrelated traits ($M = 7.39$ vs. $M = 5.63$), $F(1, 114) = 175.56$, $p < 0.001$, $MSe = 1.06$, $\eta_p^2 = .606$. The two-way interaction, $F(2, 114) = 3.53$, $p = .033$, $MSe = 1.06$, $\eta_p^2 =. 058$, revealed that, as expected, the manipulation of accessibility had no effect on ratings of stereotypic traits ($M = 7.40$ vs. $M = 7.51$ vs. $M = 7.25$), all ts < 1. But the descriptiveness ratings of the stereotype-unrelated traits were affected by the manipulation of accessibility, that is, the stereotype-unrelated traits received higher ratings in the non-accessibility condition ($M = 5.98$) when compared to the non-diagnostic accessibility ($M = 5.32$), $t(114) = 2.10$, $p = .038$, Cohen's $d = .48$, 95% CI [0.02, -0.91], but not when compared to the diagnostic accessibility ($M = 5.58$), $t(114) = 1.27$, $p = .207$, Cohen's $d = .29$, 95% CI [-0.16, -0.72].

Finally, to better understand the three-way interaction, $F(2, 114) = 3.11$, $p = .049$, $MSe = 1.06$, $\eta_p^2 =. 052$, we computed a contrast on the responses to the stereotype-unrelated scales corresponding to an interaction between type of prime (stereotype-consistent vs. stereotype-unrelated) and contamination susceptibility after collapsing non-accessibility and non-diagnostic accessibility (i.e., non-accessibility/non-diagnostic accessibility collapsed vs. diagnostic-accessibility). The contrast was indeed significant, $t(114) = 2.72$, $p = .007$, Cohen's $d = .71$, 95% CI [0.19, -1.30], showing that the effect of exposure to the stereotype-unrelated prime relative to the stereotype-consistent prime was weaker in the diagnostic accessibility than in the contamination-susceptible conditions (see Table 5). The same contrast was nonsignificant for the stereotypic scales, $t(114) = 1.02$, $p = .312$, Cohen's $d = .27$, 95% CI [-0.26, -0.82]. In addition, we computed the remaining orthogonal contrast, that is, the contrast corresponding to an interaction between type of prime (stereotype-consistent vs. stereotype-unrelated) and contamination susceptibility after excluding non-diagnostic accessibility (i.e., non-accessibility vs. non-diagnostic accessibility) for stereotype-unrelated rating scales. The contrast was significant, $t(114) = 7.69$, $p = .006$, Cohen's $d = 1.74$, 95% CI [1.28, -2.23], showing that the effect of stereotype-unrelated relative to the stereotype-consistent prime was weaker in the diagnostic accessibility than in the contamination-susceptible conditions (non-accessibility and non-diagnostic accessibility) conditions. Again the same contrast was nonsignificant in the case of the stereotypic scales, $t(114) = 1.04$, $p = .310$, Cohen's $d = 0.24$, 95% CI [-0.21, -0.67].

GENERAL DISCUSSION

In this article we argued that there are functional limits on the cognitive malleability of stereotypes. Just as excessive inflexibility in stereotypes cannot be maximally adaptive, excessive malleability would render stereotypes useless. The results from the experiments described here suggest that even when priming is subliminal, and thus the priming source cannot be ascertained, contextual contamination can still be avoided as long as a previous stereotype-assembling episode is accessible at the time of priming. Experiment 2 extended these results by suggesting that this avoidance of mental contamination occurs not because the previous assessment of the stereotype somehow overshadows subliminal priming, given that we required previous assessment of stereotype in two different conditions (diagnostic and non-diagnostic) and subliminal failed to be "overshadowed" in the latter condition. As so, it seems that this avoidance of mental contamination depends not only on the degree of accessiblity of a previous episode of stereotype assembling but also on the perceived diagnosticity of this episode.

Hence, although stereotype assembly is susceptible to contextual intrusion, the working stereotype that is assembled at a given moment seems to be the result of a compromise between contextual contamination and belief monitoring. The balancing of these two mechanisms allows us to benefit from both the predictive advantages conferred by stereotype stability and the adaptive flexibility of stereotype permeability.

Before we proceed with the discussion of future developments of the present work, three caveats are in order. The first caveat is that, although we interpreted the difference between diagnostic and non-diagnostic conditions in Experiment 2 as due to the fact that monitoring would produce more informative results in the former than in the latter condition, it is possible that our manipulation of non-diagnosticity instead of making this condition less informative, simply made the cognitive conflict between previous stereotype assessment and subliminal priming less conspicuous. In that sense, monitoring was not necessarily triggered in non-diagnostic conditions. Only further research can differentiate these alternative accounts.

The second caveat is that even if we agree that the emergence of priming effects in the accessible/non-diagnostic condition in Experiment 2 is compatible with the use of a diagnosticity-based monitoring heuristic, we recognize that this evidence is only indirect. Nevertheless, we would argue that it presents compelling evidence against the idea that a discrepancy between a previous assembling episode and the currently accessible information leads to an exhaustive and systematic search. It is true that our manipulation rendered the discrepancy non-diagnostic but participants could always search their own memories for more diagnostic instances of stereotype assembling and apparently they did not. In any case, it is our goal to continue to explore these effects with more process-based measures like reading time or event-related potentials or with measures that allow for the estimation of the role of the different processes involved (e.g., Krieglmeyer & Sherman, 2012).

The third caveat is that although the ideas and research inspired by the notion of a working stereotype seemed promising to us, we easily admit that further research is needed to elaborate this notion; in particular, its relationship with working memory. Thus, the next paragraph is meant to provide some possible follow-up ideas regarding the relationships between our notion of working stereotype and the more general notion of working memory.

WORKING STEREOTYPES AND WORKING MEMORY

There seems to be a consensus regarding some features of working memory, namely that it corresponds to "those mechanisms or processes that are involved in the control, regulation and active maintenance of task relevant information in the service of complex cognition including novel, as well as familiar, skilled tasks" (Miyake & Shah, 1999, p. 450). Thus, selective attention and resistance to interference in the sake of goal fulfilling or task completion are crucial aspects of working memory (Robison & Unsworth, 2015). Although belief assessment is normally not included when most authors talk about working memory guiding complex cognition, belief assessment about the typical features is certainly a complex task that involves, for instance, separating cultural stereotypes from individual beliefs (e.g., Devine & Elliot, 1995). In this sense, for instance, it would be interesting to explore whether working memory capacity predicts vulnerability to context priming as we used in our studies.

Moreover, belief assessment could be seen as part of the working self, a concept that is described by its proponents as such: "According to this view the goals of the working self form a subset of working-memory control processes organized into interconnected goal hierarchies that function to constrain cognition, and ultimately behavior, into effective ways of operating on the world" (Conway & Pleydell-Pearce, 2000, p. 265). It is obvious that belief assessment must be a part of self-expression, a component arguably very important to the working self. In fact, the same authors in a footnote recognize that: "We do this while at the same time recognizing that beliefs and attitudes are also significant aspects of the self" (footnote 2, same page). We agree and we see our own work as a step in that direction.

CONCLUSION

Decades of research has demonstrated that established constructs are conservative and self-perpetuating, but more recent research has suggested mechanisms that allow considerably more flexibility and revision than previously emphasized in the literature. Although this development has opened up new lines of research and a new understanding of the flexibility of our mental life, it also raises the question of why, if such mechanisms exist, our mental lives are not thrown into chaos. In this article, we present research that we hope contributes to beginning to delineate the careful balance between processes of conservatism and processes of change in our

mental representations, processes that are not limited to the operation of stereotypes. We hope that pursuing this line of research will help document the interplay of mechanisms that make our mental processes maximally adaptive.

APPENDIX A

List of Words Used as Primes in the Subliminal Priming Task, Experiments Pilot, 1, and 2

12 Positive Trait Words Associated with Friendliness
Generous
Affectionate
Humorous
Friendly
Sensitive
Steadfast
Kind
Nice
Reliable
Sociable
Happy
Helpful
12 Positive Traits Words Associated with Intelligence or Intellectuality
Skillful
Motivated
Determined
Competent
Clever
Dynamic
Methodical
Efficient
Quick
Creative
Insightful
Rational
3 Neutral Words
Always
Space
During

APPENDIX B

LIST of 42 Personality Traits For Stereotype Assessment Task, Experiments Pilot, 1, and 2

Traits	
Strong	Undecided
Warm	Vulgar
Rich	Unfaithful
Disorderly	Tactful
Faithful	Unpretentious
Thoughtful	Discreet
Cultured	Fragile
Trouble-maker	Intolerant
Attractive	Incapable
Resolved	Ignorant
Capable	Offensive
Peaceful	Attentive
Cold	Sophisticated
Methodical	Understanding
Respectful	Distrustful
Naive	Polite
Ugly	Ill-mannered
Organized	Unorganized
Disrespectful	Thoughtless
Show-off	Poor
Arrogant	Hostile

APPENDIX C

FUNNELED DEBRIEFING PROCEDURE

Now that you have finished your participation in this study, we would like to ask for your opinion about a couple of factors that can be important to, in the future, improve the relation between the participants and the computer in the research context. Next you will be asked a few questions.

1. During this study did you think the SIZE OF THE TEXT that appeared on the screen was:
   a) Appropriate
   b) Large
   c) Small
2. During this study did you think the CONTRAST between the letters and the screen was:
   a) Appropriate
   b) Large
   c) Small

3. During Study 1, during the visual localizing stimuli task at the right and left sides, you have probably noticed that, during the stimuli presentation the screen blinked. Do you think that might have somehow influenced your performance on the study you participated in next (Study 2)?
   If you think YES, press the key "s."
   If you think NO, press the key "n."
4. In the case you answered NO to the previous question press the "space bar."
5. In the case you answered YES, we would like to ask you to tell us what might have happened. Press the option (a, b, c, or d) that best corresponds to your answer.
   "The screen blinking during the stimuli presentation has influenced my performance on Study 2 because…"
   a) …I paid less attention to the tasks.
   b) …I paid more attention to the tasks.
   c) …Words appeared that induced my answers.
   d) …I ended up much more tired during the study.

REFERENCES

Bacelar do Nascimento, M. F., Pereira, L. A. S., & Saramago, J. (2000). Portuguese corpora at CLUL. In *Second international conference on language resources and evaluation—Proceedings, volume II* (pp. 1603-1607). Athens: European Language Resource Association.

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology, 43*, 437-449.

Barsalou, L. W., Sewell, D. R., & Ballato, S. M. (1986). Assessing the stability of category representations with graded structure. Unpublished manuscript.

Barsalou, L. W., Spindler, J. L., Sewell, D. R., Ballato, S. M., & Gendel, E. M. (1987). Assessing the instability of category representations with property generation. Unpublished manuscript.

Bellezza, F. S. (1984a). Reliability of retrieval from semantic memory: Common categories. *Bulletin of the Psychonomic Society, 22*(4), 324-326.

Bellezza, F. S. (1984b). Reliability of retrieval from semantic memory: Information about people. *Bulletin of the Psychonomic Society, 22*, 511-513.

Bellezza, F. S. (1984c). Reliability of retrieval from semantic memory: Noun meanings. *Bulletin of the Psychonomic Society, 22*(5), 377-380.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242-261.

Conchiglia, G., Rocca, G. D., & Grossi, D. (2007). On a peculiar environmental dependency syndrome in a case with frontal-temporal damage: Zelig-like syndrome. *Neurocase: The Neural Basis of Cognition, 13*, 1-5.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*, 261-288.

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes really fading? The Princeton trilogy revisited. *Personality and Social Psychological Bulletin, 21*, 1139–1150.

Everitt, B. (1977). *The analysis of contingency tables.* London: Chapman and Hall.

Ferreira, M. B., Garcia-Marques, L., Toscano, H., Carvalho, J., & Hagá, S. (2011). Para uma revisão da abordagem multidimensional das impressões de personalidade: O culto, o irresponsável, o compreensivo e o arrogante. *Análise Psicológica, 29*, 315-333.

Garcia-Marques, L., Ferreira, M. B., Nunes, L. D., Garrido, M. V., & Garcia-Marques, T. (2010). False memories and impressions of personality. *Social Cognition, 28*, 556-568.

Garcia-Marques, L., & Mackie, D. M. (1999). The impact of stereotype-incongruent information on perceived group variability and stereotype change. *Journal of Personality and Social Psychology, 77*, 979-990.

Garcia-Marques, L., Santos, A. S., & Mackie, D. M. (2006). Setereotypes: Static definitions or dynamic knowledge structures? *Journal of Personality and Social Psychology, 91*(5), 814-831.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in the interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology, 12*, 392-407.

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., vol. 2, pp. 1-68). Hillsdale, NJ: Erlbaum.

Hamilton, D. L., & Trolier, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J. Dovidio & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 127-163). New York: Academic.

Hewstone, M. (1994). Revision and change of stereotypic belief: In search of the elusive subtyping model. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 5, pp. 69-109). Chichester, England: Wiley.

Hess, T. M., Hinton, J. T., & Statham, J. A. (2004). Explicit and implicit stereotype activation effects on memory: Do age and awareness moderate the impact of priming? *Psychology and Aging, 19*, 495-505.

Higgins, E. T., & Brendl, C. M. (1995). Accessibility and applicability: Some "activation rules" influencing judgment. *Journal of Experimental Social Psychology, 31*, 218-243.

Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology, 13*, 141-154.

Israel, L., & Schacter, D. L. (1997). Pictorial encoding reduces false recognition of semantic associates. *Psychonomic Bulletin and Review, 4*, 577-581.

Johnson, M. K. (2006). Memory and reality. *American Psychologist, 61*, 760-771.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.

Katz, D., & Braly, K. W. (1933). Racial stereotypes of 100 college students. *Journal of Abnormal Social Psychology, 28*, 280-290.

Krieglmeyer, R., & Sherman, J. W. (2012). Disentangling stereotype activation and stereotype application in the stereotype misperception task. *Journal of Personality and Social Psychology, 103*, 205-224.

Lombardi, W. J., Higgins, E. T., & Bargh, J. A. (1987). The role of consciousness in priming effects on categorization: Assimilation versus contrast as a function of awareness of the priming task. *Personality and Social Psychology Bulletin, 13*, 411-429.

Loersch, C., & Payne, B. K. (2011). The situated inference model: An integrative account of the effects of primes on perception, behavior, and motivation. *Perspectives on Psychological Science, 6*(3), 234-252.

Macrae, C. N., Bodenhausen, G. V., & Milne, A. B. (1995). The dissection of selection in person perception: Inhibitory processes in social stereotyping. *Journal of Personality and Social Psychology, 69*, 397-407.

Macrae, C. N., Mitchell, J. P., & Pendry, L. F. (2002). What's in a forename? Cue familiarity and stereotypical thinking. *Journal of Personality and Social Psychology, 38*, 186-193.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Richards, Z., & Hewstone, M. (2001). Subtyping and subgrouping: Processes for the prevention and promotion of stereotype change. *Personality and Social Psychology Review, 5*, 52-73.

Robison, M. K., & Unsworth, N. (2015). Working memory capacity offers resistance to mind-wandering and external distraction in a context-specific manner. *Applied Cognitive Psychology, 29*, 680-690.

Rosenberg, S., Nelson, C., & Vivekananthan, P. S. (1968). A multidimensional approach to the structure of personality impressions. *Journal of Personality and Social Psychology, 9*, 283-294.

Santos, A. S. C. (2001). A instabilidade dos estereótipos: Experiências em cognição social com uma metodologia de teste-reteste longitudinal. Unpublished manuscript. Faculdade de Psicologia e de Ciências da Educação, Universidade de Lisboa.

Santos, A. S., Garcia-Marques, L., Mackie, D. M., Ferreira, M. B., Payne, B. K., & Moreira, S. (2012). Implicit open-mindedness: Evidence for and limits on stereotype malleability. *Journal of Experimental Social Psychology, 48*, 1257-1266.

Santos, A. S. C. (2007). A relatividade dos estereótipos: Sensibilidade ao contexto e mecanismos explicativos. Unpublished manuscript. Faculdade de Psicologia e de Ciências da Educação, Universidade de Lisboa.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1– 24.

Shih, M., Ambady, N., Richeson, J. A., Fujita, K., & Gray, H. M. (2002). Stereotype performance boosts: The impact of self-relevance and the manner of stereotype activation. *Journal of Personality and Social Psychology, 83*, 638-647.

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

Srull, T. K., & Wyer, R. S. (1980). Category accessibility and social perception: Some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology, 38*, 841-856.

Strack, F., & Hanover, B. (1996). Awareness of influence as a precondition for implementing correctional goals. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action* (pp. 579-596). New York: Guilford.

Wilson, T. D., & Brekke, N. C. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116*, 117-142.

Wittenbrink, W., Judd, C. M., & Park, B. (2001). Spontaneous prejudice in context: Variability in automatically activated attitudes. *Journal of Personality and Social Psychology, 81*, 815-827.