Binding Lies: Flexible retrieval of honest and dishonest behavior

Christina U. Pfeuffer[1], Roland Pfister[2], Anna Foerster[2], Franziska Stecher[3], & Andrea Kiesel[1]

[1] Albert-Ludwigs-Universität Freiburg, Department of Psychology, Freiburg, Germany

[2] Julius-Maximilians-Universität Würzburg, Department of Psychology III, Würzburg, Germany

[3] FernUniversität in Hagen, Department of Psychology, Hagen, Germany

Acknowledgments

This research was supported by the Deutsche Forschungsgemeinschaft [KI1388/5-1, Andrea Kiesel].

Correspondence:
Christina Pfeuffer
Albert-Ludwigs-Universitaet Freiburg
Cognition, Action, and Sustainability Unit, Department of Psychology
Engelbergerstrasse 41
79085 Freiburg, Germany
Tel. +49-761-203-9165
Email: christina.pfeuffer@psychologie.uni-freiburg.de

Abstract

Telling a consistent lie across multiple occasions poses severe demands on memory. Two cognitive mechanisms aid with overcoming this difficulty: Associations between a question and its corresponding response and associations between a question and its previous intentional context (in this case: honest vs. dishonest responding). Here, we assessed whether intentional contexts such as an honest versus dishonest mindset modulate the retrieval of stimulus-response (S-R) associations. In an item-specific priming paradigm, participants classified stimuli either honestly or dishonestly during a prime and a later probe. The results of three experiments yielded automatic retrieval of the previously primed motor responses (for both honest and dishonest responses) only when the intentional context repeated but not when it switched. These findings indicate interdependent associations between a stimulus, its intentional context, and the corresponding response, allowing for flexible, context-specific retrieval. Thus, humans benefit from prior learning history without incurring costs when the intentional context changes. This finding implies top-down control over the retrieval of S-R associations and provides new insights into the mechanisms of associative learning.

Keywords: Stimulus-response associations; associative learning; lying; dishonesty; context-specificity; top-down processes

Public Significance Statement

Lying poses severe memory demands, as liars have to face the difficulty of telling a consistent lie across various situations. There are two cognitive mechanisms that help liars to tell consistent lies: Associations between questions and the responses given to answer them and associations between questions and the previously chosen intentional context (lying vs. truth-telling). Here, we show that responses previously given in one intentional context only affect later responses to the same stimulus when this intentional context repeats (e.g., lie – lie), but not when we respond to the same stimulus in a different intentional context (e.g., truth – lie). Thus, benefits from responses previously learned in one intentional context (e.g., truth-telling), do not interfere with responding in a different intentional context (e.g., lying). These findings likely are not restricted to the context of lying, but might generalize to other intentional contexts.

Introduction

In everyday life, every moment we are faced with multiple stimuli. Deliberate processing of all of these simultaneous stimuli in depth before deciding whether and how to respond to them would render response selection very inefficient. Luckily, there are stimuli that we want to consistently respond to in the same way. By forming associations between stimuli and our responses to them, so-called stimulus-response (S-R) associations (e.g., Dennis & Perfect, 2013; Horner & Henson, 2009; for a review Henson, Eckstein, Waszak, Frings, & Horner, 2014), we are able to partly automatize and expedite response selection.

Crucially, however, automatically retrieving associated responses is only beneficial when the current situational context and our current intentions are in line with the associated response. A situation in which this is frequently not the case is when we decide to lie to a question we previously answered truthfully. Then, we deliberately do not want to retrieve the honest response we have previously given to a question. In line with recent ideas about the context-specific retrieval of S-R associations (Abrahamse, Braem, Notebaert, & Verguts, 2016), we examine whether the intentional contexts of lying and truth-telling can serve as top-down states modulating S-R retrieval. That is, we want to assess whether intentional contexts (e.g., lying vs. truth-telling) can be integrated into S-R associations, creating hierarchical (stimulus – intentional context – response) associations that allow for context-specific (i.e., intention-specific) S-R retrieval.

Recent studies on the components of S-R associations have already assessed whether different classification tasks (e.g., size classification vs. mechanism classification) modulate the retrieval of S-R associations (e.g., Horner & Henson, 2009, Moutsopoulou, Yang, Desantis, & Waszak, 2015; Pfeuffer, Moutsopoulou, Pfister, Waszak, & Kiesel, 2017). For instance, Moutsopoulou et al. (2015) used an item-specific priming paradigm in which stimuli appeared only twice, once as a prime, and associations were formed, and once as a probe (lag 2-7 trials), and the associations formed in the prime were assessed. Between the prime and probe instance of a stimulus, the classification task participants had to perform (size vs. mechanism) could repeat or switch. By using task cues that indicated the current classification-action mapping, the authors also manipulated whether the response (left vs. right) participants had to perform to indicate the correct classification (larger/smaller for the size task; mechanic vs. non-mechanic for the mechanism task) repeated or switched between prime and corresponding probe. They found that participants showed better performance when stimulus classifications (i.e., when the classification task) repeated rather than switched and when responses repeated rather than switched. Importantly, however, there was no interaction between these effects (see also e.g., Horner & Henson, 2009; Pfeuffer et al., 2017). This led Moutsopoulou et al. (2015) to conclude that stimuli became independently associated with task-specific semantic classifications and motor outputs. Their findings, in turn, posit that bottom-up processes alone explain the retrieval of both S-R components (stimulus-classification and stimulus-action associations). Top-down processes like task-related intentions did not seem to play a role and there was no indication of a hierarchical organization of stimulus-classification and stimulus-action associations.

Conversely, a recent study by Waszak, Pfister, & Kiesel (2013) presented tentative evidence that top-down processes could affect the automatic retrieval of S-R associations. There, participants first trained two classification tasks (color versus shape classification of visual stimuli) on bivalent stimuli (i.e., coloured shapes). A task cue indicated which task was to be performed on a given trial. This training resulted in task-rule congruency effects, that is, faster responses when both classification tasks required the same keypress and slower responses when both classification tasks yielded different keypresses. Such task-rule congruency effects indicate that the irrelevant stimulus dimension activated the response associated with the currently irrelevant classification task. After initial training, four out of six shapes were presented only as distractors (i.e., whenever these shapes appeared, participants had to judge the colour of the stimulus, but never the shape of it), and two of the four shapes were re-instructed to the opposite keypress response. Re-instructed distractors did not any more yield any congruency effects in the following blocks, whereas the other distractors still yielded congruency effects independent of whether they still appeared as targets or not. These findings might indicate that intentionally formed task sets or intentional task set negations may have an influence on automatic retrieval. However, Waszak et al. (2013) could not unequivocally ascertain that top-down processes, like supraordinate intentions that modulated S-R retrieval, drove their results. Instead, the previously practiced S-R mapping and the currently instructed S-R mapping (see e.g., Liefooghe & De Houwer, 2018; Pfeuffer et al., 2017, for evidence on the instruction-based formation of S-R associations) could alternatively have led to the formation of competing S-R associations. If the latter were the case, their findings could be accounted for by bottom-up processes alone.

Here, we want to gain clear-cut evidence for the influence of top-down processes on S-R retrieval by assessing S-R retrieval in two distinct intentional contexts that we also experience in everyday life, truth-telling and lying. Lying can be a highly demanding task. Not only do liars have to cope with cognitive factors that render lying rather effortful per se, but they also have to maintain a particular dishonest account consistently across multiple situations.

These costs of lying received continued interest from psychological studies in the past and there is broad consensus that lie-telling is associated with additional effort as compared to responding honestly. When telling a lie for the first time, agents have to process a question posed to them and retrieve the honest answer in order to subsequently decide to conceal the truth and construct and tell a lie (e.g., Debey, de Houwer, & Verschuere, 2014; Foerster, Wirth, Herbort, Kunde, & Pfister, in press; Walczyk, Harris, Duck, & Mulay, 2014). Thus, telling a lie for the first time comes with a considerable amount of cognitive processing that becomes evident, for instance, in prolonged reaction times (RTs) when telling a lie as compared to telling the truth (e.g., Debey et al., 2014; Duran, Dale, & McNamara, 2010; Pfister, Foerster, & Kunde, 2014; Spence et al., 2001).

However, previous studies also suggest that associative learning mechanisms help to ease the efforts associated with repeated lie-telling. In fact, telling a lie repeatedly can reduce or eliminate the cognitive effort that has to be invested initially (Dike, Baranoski, & Griffith, 2005; Polage, 2012; Van Bockstaele et al., 2012; Walczyk et al., 2012; Walczyk, Mahoney, Doverspike, & Griffith-Ross, 2009). For instance, performance costs associated with lying were significantly reduced for questions that participants have consistently given the same dishonest answer to as compared to questions that participants have consistently answered honestly or both honestly and dishonestly in equal proportion (Van Bockstaele et al. 2012; Foerster, Pfister, Schmidts, Dignath, Wirth, & Kunde, in press). These findings can be reconciled with associative learning accounts as these findings suggest that stimulus-response (S-R) associations between questions and participants’ responses are established, allowing for rapid, automatized retrieval of dishonest responses.

Furthermore, recent evidence suggests that not only a given dishonest response can be associated to the question. Rather, agents also seem to store the information of having responded dishonestly to a question and this contextual information is retrieved automatically when the question is re-encountered (Koranyi, Schreckenbach, & Rothermund, 2015). More precisely, participants in this study were instructed to provide honest or dishonest answers to several questions during an initial interview. Afterwards, interview questions and new questions served as primes in a computerized priming task. After each prime question, a target word appeared, and this target was either the word "honest" or the word "dishonest". Participants had to classify the target word by pressing a left or right button. When a question was presented as prime that had been answered dishonestly in the interview, participants tended to classify the "dishonest" target more easily than the "honest" target (though the effect was only significant when analyzing inverse efficiency scores rather than RTs). This finding suggests that, when participants re-encountered the questions, they automatically retrieved the contextual information of having provided an honest or dishonest answer to the question.

The previously reported two findings – automatic S-R associations for dishonest responses, and automatic retrieval of the intentional context – point towards an associative learning mechanism that aids with overcoming the difficulty of telling the same lie consistently across different occasions. Thus, all pre-requisites for a top-down modulation of S-R retrieval seem to be given in lying. This makes a comparison between the intentional contexts of lying and truth-telling ideal to assess the top-down influence of intentional contexts on S-R retrieval.

At present, it is unclear how the described mechanisms interact to retrieve a dishonest response. Two scenarios seem plausible here. Either, when encountering a particular question, the stimulus might automatically retrieve the associated responses and, simultaneously and independently, the corresponding contextual information (i.e., whether one had been honest or dishonest before; *independence hypothesis*). This would suggest that bottom-up processes alone can account for S-R retrieval in dishonest responding (c.f., Horner & Henson, 2009; Moutsopoulou et al., 2015; Pfeuffer et al., 2017). By contrast, associations might also be hierarchical and context-specific in that the contextual information is retrieved first, followed by the retrieval of the response component (stimulus – intentional context – response association; *interdependence hypothesis)*. If the latter were the case, responses could only be retrieved when learning and retrieval contexts match. This would suggest top-down control over the retrieval of S-R associations.

In the present study, we aim to go beyond preliminary findings in lying to directly test the interdependence hypothesis in an item-specific priming paradigm. In this paradigm, we use a well-documented feature of S-R associations: As previously described when encountering a stimulus, agents do not only retrieve its semantic classification, but also the motor response they had given to perform the classification (Dennis & Perfect, 2013; Horner & Henson, 2009; for a review Henson, Eckstein, Waszak, Frings, & Horner, 2014). Following previous studies on item-specific priming (Moutsopoulou et al., 2015), our participants were to classify everyday objects as small or large by pressing a left or a right key (see Figure 1). Each object was only presented twice throughout the experiment, once as a prime and once as a probe. Between the prime and probe instance of one specific object (appearing with a lag of 2-7 trials), the required motor response (left vs. right key press), could either repeat or switch (see Figure 2). In this setting, faster and more accurate responding to repetitions of the motor response than to switches would indicate that S-R associations between the stimulus and the motor response had been formed during primes and were automatically retrieved during probes (Dobbins, Schnyer, Verfaellie, & Schacter, 2004; Hsu & Waszak, 2012; Moutsopoulou et al., 2015).

Crucially, we also varied between blocks whether participants were to classify objects honestly (i.e., appropriately) or dishonestly (i.e., intently provide an inappropriate classification) during prime and probe trials, respectively. Thus, intentional context (and with it the semantics of the motor response) either repeated or switched between the prime and probe instance of a stimulus. Based on the interdependence hypothesis, we predicted that the response sequence should only affect performance for context repetitions but not for context switches.

In line with the idea of a two-step process in lying that consists of activating the truth and subsequently inhibiting it during the generation of a lie, we focused on the activation versus inhibition of the honest response (Debey et al., 2014). That is, we chose the intentional contexts of responding truthfully (honest responding: activation of the truthful response) versus not truthfully (dishonest responding: inhibition of the truthful response). Note, however, that this focus on responding truthfully or not implies that our study design could not address additional aspects of lying (for such additional aspects, see the activation-decision-construction-action theory of lying; Walczyk et al., 2014).

Experiment 1

In Experiment 1, we assessed automatic retrieval of dishonest responses by manipulating the prime context (honest vs. dishonest responding) in an item-specific priming paradigm, and measuring its impact on honest responding in the probe.

More precisely, participants categorized objects as small or large, relative to a shoe box. A preceding cue, “K + G” or “G + K” (K for German “klein”, Eng. “small”; G for German “groß”, Eng. “large”), indicated whether a right or left key press was to be performed to classify the object as small or large, respectively. A colored context frame further indicated whether to classify objects honestly or dishonestly. For instance, if participants were instructed to lie during a trial, and an apple (an object smaller than a shoe box) was presented, they were to classify the apple as large, whereas they were to appropriately classify the apple as small when honest responding was required. In each block, participants either responded honestly or dishonestly for all prime trials, and they always responded honestly for the following probe trials (see Experiments 2 and 3 for an orthogonal design). Motoric (left or right) responses for each object either repeated or switched between the prime and probe instance of a stimulus, depending on the respective mapping for the small/large classification and the context.

For context sequence repetitions (honest ► honest), we expected faster responses for repetitions rather than switches of the motor response between prime and probe. For context switches (dishonest ► honest), an effect of response sequence should only emerge if stimuli were independently associated with response and context, but not if context and response were interdependently (i.e., hierarchically) associated with the stimulus.

Methods

Participants. Twenty-four participants (8 male, 5 left handed, mean age = 24.3 years) took part and received 12€ or course credit for their participation. G*Power (Erdfelder, Faul, & Buchner, 1996) determined a sample size of 24 participants based on the effect size of the difference between response repetitions and switches observed in studies using a similar item-specific priming paradigm (Moutsopoulou et al., 2015; Pfeuffer et al., 2017) to find an effect of $d_z = 0.60$ ($\alpha < .05$) with a power of 80%. The study was conducted in adherence to the standards set by the local ethics committee and participants provided written informed consent. Data of one participant were excluded, because less than 50% of his trials fulfilled the inclusion criteria for RT analyses. Two additional participants were excluded due to language difficulties and problems understanding the task. An additional exclusion criterion, error rates above 30%, was met by none of the remaining participants. The data of excluded participants were replaced with data from new participants.

Stimuli and apparatus. Participants sat in a sound attenuated room approximately 60 cm from a 19” LCD screen (resolution: 1024 x 768). Their left and right index fingers rested on two external keys placed in front of them to the left and right (inter-key distance: 13.5 cm).

A set of 512 distinct object images (size: 256 pixels x 256 pixels, about 8° visual angle) was adopted from previous studies (Brady, Konkle, Alvarez, & Oliva, 2008; Moutsopuolou, Yang, Desantis, & Waszak, 2015). The depicted objects had to be judged according to their real-life size in relation to a size referent (reference box: 37.5 cm x 30 cm x 13.5 cm). Half of the objects were smaller than the size referent and half of the objects were larger than the size referent. Each object appeared only twice throughout the experiment, once as a prime and once as a probe. Twenty-four additional objects were used in a preceding practice block.

Responses were instructed via the cues “K + G” and “G + K”, corresponding to the first letters of the German words for small (“klein”) and large (“groß”). Items were to be classified by pressing the key that spatially corresponded to the applicable item classification (truth trials: appropriate classification, lie trials: purposely inappropriate classification).

To support participants´ interpretation of the task as “lying” versus “truth-telling”, throughout the experiment, the silhouette of a person was displayed in the background of the screen. In the instructions, this person was introduced to the participants and participants were instructed to imagine that they gave the honest versus dishonest response to this person when performing the size classification task.

Design and procedure. Participants´ task was to classify the presented objects according to their size as fast and accurately as possible. Accurate responses were defined as giving the appropriate response on prime truth trials and as giving the inappropriate response on prime lie trials. Before each block, participants were instructed to respond honestly throughout the block (truth-truth block) or to lie during the prime trials and then respond honestly during probe trials (lie-truth block). In this setting, honest responding meant that participants were instructed to judge the item´s real-life size and provide the appropriate classification, whereas participants were instructed to purposely provide the inappropriate classification during lie trials. The color of a frame (orange vs. blue) around the centrally presented stimulus provided information regarding the current context (truth-telling vs. lying) in each trial.

Each trial started with an inter-trial interval of 1000 ms, during which the color of a frame around the center of the screen indicated the current context (truth vs. lie; see Figure 1). The colored frame remained visible throughout the entire trial. Subsequently, a cue (700 ms) indicated the current key-classification mapping and participants classified the following item accordingly via a left or right key press (maximum duration: 2000 ms). False responses, that is, inappropriate classifications during truth trials or accidental appropriate classifications during lie trials or response omissions triggered a specific feedback for 500 ms (“Fehler!”, Eng.: “error!”; “zu langsam!”, Eng.: “too slow!”).

Blocks consisted of eight trials, four prime trials followed by four corresponding probe trials. After eight practice blocks (4 truth-truth, 4 lie-truth) participants continued with 128 blocks of the experiment proper (1024 trials, 64 blocks truth-truth and 64 blocks lie-truth, 128 trials per condition). For each block four new items were randomly selected. An individual item only appeared in one block, once as a prime and 2 to 7 trials later once as a probe (see Figure 1).

Crucially, not only the context, but also the required response (left vs. right key press) could switch or repeat between the prime trial and the probe trial of a specific item (see Figure 2). This was realized by varying the cue and resulted in four possible combinations of the factors prime context (truth vs. lie) and response sequence (repetition vs. switch).

Results

Prime responses. For the analysis of error percentages (PEs) response omissions (0.6%) were excluded. For the analysis of prime trial RTs response omissions and error trials (11.9%) as well as outliers (1.2%) were excluded, with outliers being defined as RTs that deviated from their individual cell mean by more than 3 standard deviations. RTs and error rates were subjected to paired *t*-tests comparing prime lie and prime truth trials.

Prime lie trials during which participants were instructed to purposely provide inappropriate classifications were associated with both longer RTs and larger error rates (see Figure 3), and this difference was significant for both measures; RTs, $t(23) = 8.80$, $p < .001$, $d_z = 1.80$; error rates, $t(23) = 9.43$, $p < .001$, $d_z = 1.93$.

Probe responses. Participants committed errors on 7.2% of the probe trials and omitted responses on 0.2% of the probe trials. For the analysis of probe trial error rates, trials with response omissions were excluded. Furthermore, probe trials with preceding errors or response omissions in the corresponding prime trial were also excluded. On average, this led to the exclusion of 12.6% of trials for the error rate analysis. For the analysis of probe trial RTs, trials with response omissions or errors as well as outliers, defined as RTs deviating by more than 3 standard deviations from their individual cell means, were excluded (1.3%). Moreover, probe trials with response omissions or errors in the corresponding preceding prime trial were excluded. For RT analysis, these criteria lead to the exclusion of 18.2% of the trials, on average.

RTs and error rates were subjected to separate 2 x 2 repeated measures analyses of variance (ANOVAs) with the factors prime context (truth vs. lie) and response sequence (response repetition, RR, vs. response switch, RS). Paired *t*-tests were used to further investigate interactions of prime context and response sequence. Please note that results of Experiment 1 are plotted in accordance with the analyses in Experiment 2 and 3 for the sake of comparison (Figures 3, 4, and 5).[1]

RTs. Participants were slower to respond to items that they had lied to during prime trials in comparison to items they had classified honestly, $F(1,23) = 32.72$, $p < .001$, $\eta_p^2 = .59$ (see Figure 3A). The main effect of response sequence (response repetition vs. response switch) was not significant, $F(1,23) = 2.46$, $p = .130$, $\eta_p^2 = .10$, whereas, the interaction of prime context and response sequence reached significance, $F(1,23) = 8.06$, $p = .009$, $\eta_p^2 = .26$. There was a significant difference between response repetitions and response switches when having responded honestly in the prime trial, $t(23) = 4.08$, $p < .001$, $d_z = 0.83$, but not after lying in the prime trial, $t(23) = 0.69$, $p = .500$, $d_z = 0.14$. Thus, when participants responded honestly in both prime and probe of an item, probe RTs were significantly larger for response switches in comparison to response repetitions, whereas probe RTs did not differ between response repetitions and switches when participants had lied during prime trials.

Error rates. Error rate analysis showed a significant main effect of prime context, $F(1,23) = 8.34$, $p = .008$, $\eta_p^2 = .27$, with responses being more error-prone for items for which participants had provided lies during the prime trial (see Figure 3B). Again, the main effect of response sequence was not significant, $F(1,23) = 2.27$, $p = .146$, $\eta_p^2 = .09$. Furthermore, a significant interaction between prime context and response sequence, $F(1,23) = 7.74$, $p = .011$, $\eta_p^2 = .25$, was qualified by significantly increased error rates for response switches in comparison to response repetitions for probe trials with honest prime responses, $t(23) = 2.95$, $p = .007$, $d_z = 0.60$, and a marginally

[1] In Experiment 1, results in the lie-truth blocks might have been affected by the task switch from the last prime trial to the first probe trial (Debey, Liefooghe, de Houwer, & Verschuere, 2015; Foerster, Wirth, Kunde, & Pfister, 2017). However, additional RT and error rate analyses that excluded the first probe trials yielded the same pattern of results.

significant inverse pattern of decreased error rates for response switches for probe trials with dishonest prime responses, $t(23) = 1.97$, $p = .061$, $d_z = 0.40$.

Discussion

Automatic S-R translation, as evident in significantly increased RTs and error rates for response switches compared to repetitions, only occurred when participants responded honestly during both prime and probe trials, that is, when the context repeated. When participants lied about the classification of objects during primes and were required to provide an honest answer during subsequent probes, RTs and error rates for response repetitions and switches did not differ significantly (though they descriptively showed the reverse tendency).

This pattern of results is in line with the interdependence hypothesis. Participants formed S-R associations between stimuli and motor responses during both contexts, truth-telling and lying, but these S-R associations were only retrieved when context repeated rather than switched. An alternative account for the results of Experiment 1, however, is that participants might have formed automatic S-R associations between stimuli and motor responses only during honest responding but not during lying. Experiment 2 rules out precisely this alternative explanation.

Experiment 2

Experiment 2 employed the same design as Experiment 1, except that we varied orthogonally whether participants responded honestly or dishonestly during primes as well as during probes, so that all four possible combinations of prime context (truth vs. lie) and probe context (truth vs. lie) were realized. Based on the findings of Experiment 1, we predicted an impact of response sequence only if prime and probe context matched (context repetitions) but not if they differed (context switches). This would suggest that both during truth-telling and lying, context-specific S-R associations between the stimulus and the motor response are formed, yet are thus only retrieved when priming and retrieval context match.

Methods

Participants. Twenty-four new participants (10 male, 2 left handed, mean age = 23.9 years) took part, provided written consent, and received 12€ or course credit for their participation. One participant was excluded, because his data provided less than 50% of usable trials and another participant was excluded due to disturbing noise outside the laboratory. An additional exclusion criterion, error rates above 30%, was met by none of the remaining participants. Additional participants were recruited in place of the excluded participants.

Stimuli, apparatus, design, and procedure. Stimuli and apparatus were equivalent to Experiment 1. The design of Experiment 2 equaled Experiment 1 with the exception that in Experiment 2, all four possible combinations of prime and probe context were realized (truth-truth vs. truth-lie vs. lie-lie vs. lie-truth). This resulted in 32 blocks per block type and 64 trials per condition.

Results

Prime responses. Response omissions (0.5%) were excluded from prime trial analyses. Furthermore, for RT analysis trials with erroneous responses (12.0%) and outliers were excluded (1.1%). RTs and error rates were subjected to paired *t*-tests comparing prime lie and prime truth trials.

In accordance with the results of Experiment 1, we found a significant difference between lying and truth-telling during prime trials in both RTs, $t(23) = 9.76$, $p < .001$, $d_z = 1.99$, and error rates, $t(23) = 5.79$, $p < .001$, $d_z = 1.18$. Participants responded slower and committed more errors when they were lying (see Figures 4).

Probe responses. Participants on average committed errors on 9.1% of the probe trials and omitted responses on 0.4% of the probe trials. Equivalent to the analyses of

Experiment 1, response omissions were excluded. Moreover, probe trials with preceding errors or response omissions in the corresponding prime trial were also excluded. For the analysis of probe trial RTs, outliers were additionally excluded (0.9%). These exclusion criteria lead to an average exclusion of 19.3% of the trials for error rate analysis and 19.7% of the trials for RT analysis.

RTs and error rates were subjected to a 2 x 2 x 2 repeated measures ANOVAs with the factors prime context (truth vs. lie), response sequence (response repetition vs. response switch), and context sequence (context repetition vs. context switch). Specifically, context sequence indicates whether the same context repeated from prime to probe (i.e., truth-truth, lie-lie) or switched from prime to probe (i.e., truth-lie, lie-truth). Paired *t*-tests were used to further investigate two-way interactions.

RTs. Our analysis again showed a significant main effect of prime context, $F(1,23) = 36.48$, $p < .001$, $\eta_p^2 = .61$ (see Figure 4A). Participants were significantly slower to classify items that they had lied to during prime trials in comparison to items they had honestly classified during prime trials. Moreover, the main effect of response sequence, $F(1,23) = 6.00$, $p = .022$, $\eta_p^2 = .21$, reached significance. Overall, response switches were associated with significantly longer RTs in comparison to response repetitions. The main effect of context sequence failed to reach significance, $F(1,23) = 1.21$, $p = .282$, $\eta_p^2 = .05$, however, prime context and context sequence significantly interacted, $F(1,23) = 64.27$, $p < .001$, $\eta_p^2 = .74$. Subsequent paired *t*-test showed that, when context repeated, probe RTs were faster after participants had told the truth in the prime (and responded honestly in the probe) as compared to when participants had lied during the prime (and responded dishonestly in the probe), $t(23) = 8.37$, $p < .001$, $d_z = 1.71$. Conversely, when context switched, probe RTs were faster after participants had lied in the prime (and were telling the truth in the probe) as compared to when participants had told the truth in the prime (and lied in the probe), $t(23) = 6.11$, $p < .001$, $d_z = 1.25$.

Importantly, context sequence and response sequence interacted significantly, $F(1,23) = 4.59$, $p = .043$, $\eta_p^2 = .17$. Subsequent *t*-tests revealed that response switches were only associated with longer RTs than response repetitions when the context repeated, $t(23) = 3.04$, $p = .006$, $d_z = 0.62$, but not when the context switched, $t(23) = 0.53$, $p = .600$, $d_z = 0.11$. The two-way interaction of prime context and response sequence and the three-way interaction did not reach significance, *F*s < 1.

PEs. For error rates, we found a significant main effect of prime context, $F(1,23) = 6.15$, $p = .021$, $\eta_p^2 = .21$, with participants committing more errors when having lied during primes instead of having honestly classified items (see Figure 4B). The main effects of response sequence, $F < 1$, and context sequence, $F(1,23) = 2.19$, $p = .153$, $\eta_p^2 = .09$, failed to reach significance. Prime context and context sequence interacted significantly, $F(1,23) = 12.15$, $p = .002$, $\eta_p^2 = .35$. When the context repeated, probe error rates were significantly increased when participants lied in the prime (and in the probe) as compared to when participants responded honestly in the prime (and in the probe), $t(23) = 5.13$, $p < .001$, $d_z = 1.05$. When context switched there was no significant difference, $t(23) = 1.11$, $p = .279$, $d_z = 0.23$. Furthermore, the interaction of context sequence and response sequence was significant, $F(1,23) = 5.29$, $p = .031$, $\eta_p^2 = .19$. Response switches were associated with significant increases in error rates relative to response repetitions when context repeated between prime and probe, $t(23) = 2.21$, $p = .037$, $d_z = 0.45$, but not when context switched, $t(23) = 1.09$, $p = .287$, $d_z = 0.22$. The two-way interaction between prime context and response sequence and the three-way interaction were not significant, *F*s < 1.

Discussion

In Experiment 2, we found that response switches were associated with longer RTs and higher error rates in comparison to response repetitions when the context repeated, but not when the context switched. This finding demonstrates that participants formed

S-R associations during primes both when they responded honestly and dishonestly, but they automatically retrieved the motor responses during probes only when the context repeated. Thus, we conclude that irrespective of whether agents tell the truth or lie, their response (i.e., their motor output) is stored in an interdependent, hierarchical association between context, stimulus, and response. Responses were retrieved in a context-specific fashion, that is, only when the retrieval context in the probe matched the encoding context in the prime. When the context switched, neither lie-based nor truth-based S-R associations were retrieved. Crucially, the results of Experiment 2 also explain why lie-based S-R associations could not be detected in Experiment 1, in which prime lies were always associated with a context switch between prime and probe.

An important aspect to discuss is whether repetitions/switches in the semantic classifications participants provided with their responses could have affected the results. That is, whenever the context repeated, participants also had to perform the same semantic classification of a stimulus in the prime and probe. Conversely, when the context switched, the correct semantic classification of a stimulus also switched between prime and probe. Yet, there is a growing body of research suggesting that motor outputs and semantic classifications become independently associated with stimuli and are retrieved independently (e.g., Horner & Henson, 2009; Moutsopoulou et al., 2015; Pfeuffer, et al., 2017; Pfeuffer, Hosp, Kimmig, Moutsopoulou, Waszak, & Kiesel, 2018; Pfeuffer, Moutsopoulou, Waszak, & Kiesel, 2018; see also Giesen & Rothermund, 2016, for similar results for irrelevant stimuli). In these studies, participants´ responses and the (task-specific) semantic classifications they indicated were typically manipulated orthogonally. For instance, in the study of Moutsopoulou et al. (2015), participants performed an item-specific priming paradigm (lag 2-7 trials between prime and probe) in which the classification task (size vs. mechanism) and the response (left vs. right) participants had to perform independently repeated or switched between the prime and probe instance of a stimulus. Crucially, across studies, item-specific repetitions/switches in (task-specific) semantic classifications had an influence on probe performance that was independent from the effect of repetitions/switches in responses. Switches as compared to repetitions in semantic classifications were associated with longer RTs and increased error rates, but there was no indication of an interaction with S-R retrieval effects. This was the case both when participants switched to an entirely different classification task (e.g., Moutsopoulou et al., 2015) and when participants had to reverse their size classifications because of a change in the size referent (e.g., Horner & Henson, 2009).

The present experiment used a comparable design to manipulate intentional context between prime and probe, which in turn led to repetitions/switches in stimulus classification between prime and probe. Given the findings of previous studies (e.g., Horner & Henson, 2009; Moutspoulou et al., 2015) we consider it unlikely that repetitions/switches in stimulus classification played an essential role in bringing about the pattern of context-specific S-R retrieval we observed in the probe trials.

To provide further evidence against an influence of classification repetitions/switches on the basis of the current experiments, we conducted additional post-hoc analyses in which we compared prime and probe performance (see the Appendix). In a nutshell, we computed RT and error rate differences between prime and probe responses given in the same intentional context to assess how much performance improved from prime to probe. That is, per participant, for instance, mean probe honest response RTs/error rates were subtracted from mean prime honest response RTs/error rates (computed irrespective of what the probe context that had followed these prime trials in the experiment had been). We then examined the influence of context repetitions/switches from prime to probe on how much performance improved for honest and dishonest probe responses. Our reasoning here was that a context repetition automatically meant that the classification repeated, whereas a context switch meant that the classification switched. Thus, our analyses provided us with an estimate of the influence of classification repetitions/switches on honest/dishonest probe responses. We found that the influence of context/classification switches on honest and dishonest probe responses was exactly opposite. When participants were telling the truth in the probe, their performance improved more relative to the prime (i.e., the difference between prime honest responses and probe honest responses was larger) when they had also told the truth in the corresponding prime rather than lied (i.e., when the context/classification repeated rather than switched). Conversely, when participants were lying in the probe, their performance improved more relative to the prime when they had previously told the truth during the prime rather than lied. We interpret this as additional tentative evidence that classification repetitions/switches affect honest and dishonest responses differently. Yet, in the probe, we observed the same pattern of context-specific S-R retrieval for both honest and dishonest probe responses. Thus, we reason that classification repetitions/switches that differently affect overall probe performance are unlikely the origin of equivalent item-specific S-R retrieval patterns in the probe. These analyses provide additional evidence for the hypothesized difference of switching between the intentional contexts (lying and truth-telling) as compared to switching between two unrelated classification tasks (e.g., Giesen & Rothermund, 2016; Horner & Henson, 2009; Moutsopoulou et al., 2015; Pfeuffer, et al., 2017).

Furthermore, in Experiment 2, one additional alternative explanation for the pattern of results could not be ruled out. As a single colour was used to indicate the truth and lie context, context switches were always associated with colour switches, whereas context repetitions were always associated with colour repetitions. Thus, instead of forming a hierarchical association between intentional context, stimulus, and response, participants might alternatively have formed hierarchical associations between context cue (i.e., frame colour), stimulus, and response (for a similar reasoning in the task-switching paradigm see Logan & Bundesen, 2003; Mayr & Kliegl, 2003). We conducted Experiment 3 to address this potential alternative explanation.

Experiment 3

To rule out alternative explanations in terms of encoding the colour of the context cue rather than the actual intentional context, Experiment 3 was a conceptual replication of Experiment 2, but we now used four frame colours to indicate prime and probe context. Two colours were assigned to lying and two to truth-telling. One of these frame colours was only used during primes and the other was only used during probes. Thus, the item-specific transition from prime to probe always involved a switch in frame colour and prime and probe intentional context could be manipulated independently. If we found the same pattern of results as in Experiment 2, namely that item-specific response repetitions yielded performance benefits only when the context repeated but not when it switched, this would imply that participants had formed hierarchical, context-specific S-R associations that incorporated the intentional context.

Methods

Participants. An a priori sample size estimation via G*Power (Erdfelder, Faul, & Buchner, 1996) based on the effect size of Experiment 2 suggested that 44 participants were necessary to find a significant interaction between response sequence and context sequence ($\alpha < .05$) with a power of 80%. We recruited a corresponding sample (13 male, 5 left handed, mean age = 23.2 years), and participants provided written consent and received either course credit or 12€. The data of one additional participant were excluded due to error rates larger than 30%. For all remaining participants, more than 50% of their probe trials remained after trial exclusions. An additional participant was recruited in place of the excluded participant.

Stimuli, apparatus, design, and procedure. Stimuli and apparatus were the same as in Experiment 2 with the exception that four frame colours (blue – orange, red – green) were used. One of the colour pairs was assigned to the truth and lie context and one of the colours only appeared during prime trials, whereas the other colour only appeared during probe trials. Colour mappings were counterbalanced across participants. Again, participants completed 32 blocks per block type and 64 trials per condition.

Results

Prime responses. Response omissions (0.4%) were excluded from all analyses and trials with erroneous responses (12.9%) as well as outliers (1.1%) were excluded from RT analyses. RTs and error rates were then subjected to paired *t*-tests comparing honest and dishonest responses. Again, we found that lying was associated with significantly increased RTs, $t(43) = 13.18$, $p < .001$, $d_z = 1.99$, as well as error rates, $t(43) = 8.13$, $p < .001$, $d_z = 1.23$, as compared to truth-telling (see Figure 5).

Probe responses. Participants on average omitted 0.4% of the probe trials and committed errors on 10.4% of the probe trials. Response omissions were excluded from all analyses and erroneous responses as well as outliers (0.9%) were additionally excluded from RT analyses. In addition, probe trials with errors or response omissions in the corresponding prime trials were excluded from all analyses. On average, 21.7% of the probe trials were excluded from RT analysis due to these criteria.

Like in Experiment 2, RTs and error rates were subjected to 2 x 2 x 2 repeated measures ANOVAs with the factors prime context, response sequence, and context sequence. Paired *t*-tests were subsequently used to assess significant two-way interactions.

RTs. We replicated the main effect of prime context, $F(1,43) = 68.27$, $p < .001$, $\eta_p^2 = .61$, with participants exhibiting longer RTs when having lied during prime trials rather than having told the truth during prime trials (see Figure 5A). The main effect of context sequence reached significance, $F(1,43) = 26.68$, $p < .001$, $\eta_p^2 = .38$. Context switches were associated with longer RTs in comparison to context repetitions. The main effect of response sequence failed to reach significance, $F(1,43) = 3.66$, $p = .062$, $\eta_p^2 = .08$. Moreover, prime context and context sequence significantly interacted, $F(1,43) = 179.05$, $p < .001$, $\eta_p^2 = .81$. Subsequent paired *t*-tests showed that, when the context repeated, participants responded faster when they had told the truth rather than lied during the prime, $t(43) = 13.61$, $p < .001$, $d_z = 2.05$. Conversely, however, after a context switch, participants were faster when they had lied rather than told the truth during the prime, $t(43) = -9.29$, $p < .001$, $d_z = 1.40$. The interaction between context sequence and response sequence reached significance, $F(1,43) = 14.63$, $p < .001$, $\eta_p^2 = .25$. Paired *t*-tests conducted separately for context repetitions and context switches revealed that response switches were associated with increased RTs as compared to response repetitions only when the context repeated, $t(43) = 4.01$, $p < .001$, $d_z = 0.60$, but not when the context switched, $t(43) = -1.41$, $p = .167$, $d_z = 0.21$. Finally, the interaction of prime context and response sequence, $F(1,43) = 2.91$, $p = .095$, $\eta_p^2 = .06$, as well as the three-way interaction between prime context, context sequence, and response sequence, $F < 1$, failed to reach significance.

PEs. The main effects of prime context, $F(1,43) = 14.64$, $p < .001$, $\eta_p^2 = .25$, and context sequence, $F(1,43) = 10.63$, $p = .002$, $\eta_p^2 = .20$, reached significance (see Figure 5B). Participants committed more errors when they had lied during prime trials rather than told the truth. Furthermore, participants committed more errors after context switches as compared to context repetitions. The main effect of response sequence did not reach significance, $F < 1$. Furthermore, prime context and context sequence significantly interacted, $F(1,43) = 35.53$, $p < .001$, $\eta_p^2 = .45$. When the context repeated, participants committed more errors when having lied rather than told the truth during the corresponding prime, $t(43) = 7.15$, $p < .001$, $d_z = 1.08$. When the context switched, the pattern was reversed and participants committed more errors when they had told the truth rather than lied during the prime, $t(43) = -2.58$, $p = .013$, $d_z = 0.39$. Additionally, the interaction of prime context and response sequence was significant, $F(1,43) = 9.17$, $p = .004$, $\eta_p^2 = .18$. When participants had lied in the corresponding prime, they committed significantly more errors when responses switched between item-specific prime and probe, $t(43) = 2.29$, $p = .027$, $d_z = 0.34$. In contrast, when participants had responded honestly during primes, there was no significant difference in the error rates between response repetitions and response switches, $t(43) = -1.37$, $p = .179$, $d_z = 0.21$. The interaction of context sequence and response sequence, $F(1,43) = 3.63$, $p = .063$, $\eta_p^2 = .08$, and the three-way interaction of prime context, context sequence, and response sequence, $F < 1$, did not reach significance.

Discussion

Experiment 3 replicated the finding of Experiment 2 that response repetitions as compared to response switches were only associated with performance benefits when the intentional context repeated from prime to probe. This rules out the alternative explanation that participants had associated the frame colours indicating the intentional context with the stimuli and responses. Thus, we conclude that participants, in a hierarchical fashion, associated stimuli with an intentional context and only via the intentional context with the motor response (i.e., stimulus – context – response association). Response retrieval took place only when the intentional context repeated between the prime and probe instance of an item, but not when the intentional context switched.

General Discussion

We used an item-specific priming paradigm to investigate the contribution of context-specific S-R associations to repeated truth-telling and lying. In line with theories of behavioral automatization (e.g., Logan, 1988; Hommel, 2004), single-trial co-occurrence of stimuli and responses in close temporal proximity was sufficient to bind responses (i.e., motor outputs that signaled a specific honest or dishonest semantic content) to stimuli, allowing for automatic response retrieval upon re-encountering the stimulus. Importantly, S-R associations were retrieved automatically only when the context repeated but not when the context switched between prime and probe.

In Experiment 1, we only varied the prime context and had participants respond honestly in the probe. Thus, participants might alternatively not have formed S-R associations when lying. This alternative explanation could be ruled out with Experiments 2 and 3 that systematically varied prime and probe context and showed the same pattern of context-specific S-R retrieval both when participants had lied during the prime trial and when participants had told the truth during the prime trial. Furthermore, Experiment 3 ruled out that participants used visual features indicating the intentional contexts by introducing a change in frame colour both for context repetitions and for context switches.

Top-down control over S-R retrieval via intentional contexts

These findings suggest that top-down processes such as the intention to lie or tell the truth in response to a stimulus modulate automatic, bottom-up retrieval of S-R associations[2]. Our findings indicate that top-down intentional sets can create a context that is incorporated into an interdependent stimulus-context-response association. This association allows for context-specific S-R retrieval and thus increases behavioral flexibility by allowing humans to benefit from prior learning history with one

[2] Please note that we could not implement all aspects of real-life lying in the present experiment. As such, the central aspect of lying that mainly determines the intentional context of lying in the present experiments is the negation of the honest response. We will discuss the implications of the present findings for research on lying in a later section.

intentional context without incurring costs when later responding to the same stimulus in another intentional context.

The reported findings fit well with recent ideas on how top-down processes affect automatic retrieval (Waszak et al., 2013). Re-instructed stimuli (that were subsequently only used as distractors) in contrast to stimuli that just did not appear as targets any more did not yield task-rule congruency effects. One possible interpretation of this finding is that top-down processes such as intentionally formed task sets or intentional task set negations may control automatic retrieval (cf. Henson et al., 2014). However, the study of Waszak et al. (2013), cannot directly assess whether S-R associations were only retrieved when they concurred with currently valid S-R mappings. That is, Waszak et al. (2013) cannot directly examine whether supraordinate intentions inhibited or possibly reversed previously formed S-R associations. Alternatively, the absence of task-rule congruency effects for re-instructed S-R mappings could have occurred, because the previously practiced S-R mapping and the currently instructed S-R mapping (i.e., a corresponding prepared reflex) exerted opposing effect, yielding an overall influence of zero. In this case, the effect would have been caused by bottom-up processes alone - except for the intention to encode and apply the instructed S-R mapping. That is, if one assumes that this intention is necessary for instruction-based effects to occur (see e.g., Meiran, Cole, & Braver, 2012; but see e.g., Liefooghe & De Houwer, 2018; Pfeuffer et al., 2017, for some evidence of the contrary).

In contrast to Waszak et al. (2013), we did not re-instruct S-R mappings after a number of blocks to indirectly assess top-down processes. Instead we directly manipulated participants´ supraordinate task-related intentions regarding stimulus classification. If one were to describe our study in terms of the framing of Waszak et al. (2013), intentional contexts made participants use either the usual stimulus-classification mapping (e.g., apple – small) or its negation during dishonest responding (e.g., apple – large). S-R mappings in our study in contrast to the study of Waszak et al. (2013) were not instructed to be applied consistently within a part of the experiment, but varied from trial to trial with changes in the task cue. That is, S-R mappings could not be predicted. This further supported the application of supraordinate intentional contexts (i.e., top-down processes) instead of bottom-up processes, as participants could not prepare responses item-specifically, but had to infer them on a trial-by-trial basis. Thus, the present study is the first to provide direct evidence for a top-down control over S-R retrieval based on intentional contexts. Additionally, whereas Waszak et al. (2013) assume top-down control over the retrieval or non-retrieval of S-R associations themselves, we suggest that intentional context that specify how to perform the same classification task also determine the retrieval of S-R (i.e., stimulus-motor output) associations independent from the provided classification (i.e., small/large).

A possible challenge for such a top-down modulation of automatic S-R retrieval may arise, however, when there are more than two contexts that need to be associated to a certain stimulus. As indicated by previous findings (Koranyi et al., 2015), participants readily classify a context as either honest or dishonest and can retrieve this information later on. This mechanism will perform well for closed questions as used in the present design, because a dishonest context also comes with a specific response (i.e., the opposite of the honest response in the present setting). For dishonest responses to open questions, by contrast, it is conceivable that there are different lies that have been told on different occasions. For instance, the question “What did you do last night?” can be responded to with the honest answer (e.g., “I’ve been playing computer games.”) and, importantly, with different dishonest answers (e.g., “I’ve been reading a book.”, “’I’ve been to the gym.”, “I was visiting some friends.”). These situations likely pose stronger difficulty to context-dependent retrieval of S-R associations than the situations tested in the present experiment and thus represent an informative scenario for future investigation.

Furthermore, in the present experiments stimuli were always task-relevant. It would be interesting to see whether intentional contexts also control the retrieval of associations incorporating task-irrelevant stimuli. This could be tested by assessing not only target-response bindings, but by presenting additional distractors and assessing distractor-response bindings (see e.g., Frings & Rothermund, 2011; Giesen & Rothermund, 2014) in the present paradigm.

The impact of intentional contexts in comparison to task sets

We previously discussed whether S-R retrieval might have been affected by repetitions/switches in a stimulus´ classification associated with the manipulation of repetitions/switches in intentional context. Our results clearly indicate that intentional contexts differ from task sets. Here, a number of recent studies provided evidence for independent associations between stimuli and (task-specific) classifications and stimuli and motor outputs (e.g., Horner & Henson, 2009; Moutsopoulou et al., 2015; Pfeuffer, et al., 2017). This pattern of results is the opposite of the interdependent associations between stimulus, intentional context, and response we observed. Crucially, this cannot only be attributed to previous studies using two distinct (task-specific) classifications that did not rule each other out like small and mechanic. For instance, Horner and Henson (2009) used changes in a size referent to vary size classification. As such, in their study, participants also sometimes classified the same stimulus as small in one part of the experiment and as large in another part of the experiment.

Thus, differences between previous studies and the present experiments support the idea that a switch in intentional context (i.e., from truth-telling to lying or vice versa) is distinctly different from a switch in stimulus classification (i.e., from small to mechanic). The present study therefore argues that intentions like lying and truth-telling are not the same as two different classification tasks. Whereas supraordinate intentions lead to hierarchical and interdependent S-R associations (i.e., stimulus – context – response), different classification tasks do not yield hierarchical associative structures.

Interestingly, our findings are, however, similar to item-unspecific effects observed in task switching studies in which transitions from trial N-1 to trial N were assessed. There, participants typically respond faster for response repetitions than response switches when the task repeats on two consecutive trials (i.e., from trial N-1 to trial N; e.g., Druey & Hübner, 2008 a, b; Hübner & Druey, 2006, 2008; Koch, Schuch, Vu, & Proctor, 2011; Rogers & Monsell, 1995; Schuch & Koch, 2004; Steinhauser, Hübner, & Druey, 2009; see also Kiesel, Steinhauser, Wendt, Falkenstein, Jost, Philipp, & Koch, 2010, for a review on task switching addressing this aspect). However, when the task switches there is no or even a reversed effect and participants tend to respond slower for response repetitions than for response switches. This might indicate that response repetition benefits in task switching studies could at least partly result from remaining transient activation of intentional contexts associated with preceding tasks. In turn, our findings additionally indicate that such transient activation from intentional contexts may be interdependently bound to a stimulus.

As mentioned earlier, switching tasks from lying to truth-telling from one trial to the next is more difficult than repeatedly responding with the same intention (Debey et al., 2014; Foerster et al., 2017). The present findings suggest that automatic response tendencies in terms of S-R associations do not contribute to these switch costs. Our results clearly indicate that both honest and dishonest responses are retrieved automatically only within the same intentional context. Thus, our findings converge with prior studies suggesting that a prime candidate that underlies switching costs in this context is the interference between succeeding intentional contexts (Debey et al., 2014; Foerster et al., 2017).

The Relevance of the Present Findings for Research on Lying

As previously mentioned, in the present study we could not assess all real-life aspects of lie-telling (e.g., the deceptive intention). Instead, we focused on one central aspect occurring during lying, namely the activation and inhibition of the honest response. Thus, at present we cannot determine whether other aspects of lying could also constitute what we have termed an intentional context and how these aspects would affect the pattern of results. Future studies should try to replicate our findings in more realistic lying scenarios to provide further information on the validity of the present findings for real-life lying. Additionally, it would be interesting to compare further intentional contexts to determine whether the present findings are specific to (aspects of) lying and truth-telling. As previously discussed in the context of research on task switching, we would assume that this is not the case. Lying and truth-telling are just two exceptionally clear real-life examples of distinct intentional contexts.

Although the present context-specific retrieval of S-R associations may not be specific to lying, our findings clearly indicate that it is an important aspect that enables real-life lying. Without this context-specificity, previous honest/dishonest responses would severely impair responding in the opposite context and we would be unable to effectively lie about things we usually told the truth about in the past when promoted to do so. Not only would our delayed responses provide cues to our interaction partners, but the retrieval of the previously given honest response would lead to frequent, accidental honest responses.

Apart from highlighting the context-specific nature of lying and truth-telling, our findings provide direct evidence that the honest classification is inhibited when lying (see e.g., Debey et al., 2014; Walczyk et al., 2003, 2014, for the idea that the lying entails the inhibition of the honest response). This is evident in increased performance benefits for probe lying as compared to prime lying in blocks with switches in the intentional context as compared to repetitions. Here, our findings do not only support the idea that the honest classification is inhibited in the prime, but further indicate that associations formed between stimuli and (honest and/or dishonest) classifications in the prime contain an additional inhibitory tag, impairing the retrieval of the associated classification in a later probe instance of the same stimulus. At present, on the basis of our data we cannot determine whether stimuli are associated with the honest or dishonest classification in the prime. Both an inhibition of the honest and of the dishonest classification would similarly impair probe performance in lie-lie blocks relative to truth-lie blocks. However, as our findings suggest that the respective stimulus-classification association contains an inhibitory aspect, it would be reasonable to assume that it is the honest classification that becomes associated with a stimulus during lying.

Conclusions

Overall, our findings suggest that humans do not only store information about previous intentional contexts in memory (Koranyi et al., 2015), but use it to retrieve responses context-specifically. These findings do not only reveal the associative foundations of lying, but are also informative for theories of associative learning in general. The hierarchical organization of S-R associations ensures that only those responses fitting the current intentional context are retrieved, whereas currently irrelevant and inadequate responses are not automatically retrieved upon stimulus re-encounter. Thus, the present studies highlight a central mechanism that allows for efficient intention-based, context-specific S-R retrieval. The suggested context-specific retrieval of responses based on intentional contexts might also add to the current discussion of an associative learning framework for cognitive control (Abrahamse et al., 2016), suggesting that intentional sets can be used for context-specific response retrieval. Future research should aim to further explore the role of intentional sets, that is, top-down components, in associative learning and extend the present findings to further domains.

The data of the reported experiments as well as experiment files and syntaxes are available via the Open Science Framework: https://osf.io/cx269/; DOI: 10.17605/OSF.IO/CX269

References

Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin*, *142,* 693-728.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 14325 - 14329.

Christ, S. E., van Essen, D. C., Watson, J. M., Brubaker, L. E., & McDermott, K. B. (2009). The contributions of prefrontal cortex and executive control to deception: evidence from activation likelihood estimate meta-analysis. *Cerebral Cortex, 19,*1557–1566.

Debey, E., de Houwer, J., & Verschuere, B. (2014). Lying relies on the truth. *Cognition, 132*, 324–334.

Debey, E., Liefooghe, B., De Houwer, J., & Verschuere, B. (2015). Lie, truth, lie: the role of task switching in a deception context. *Psychological Research*, *79,* 478-488.

Dennis, I., & Perfect, T. J. (2013). Do stimulus–action associations contribute to repetition priming?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39,* 85-95.

Dike, C. C., Baranoski, M., & Griffith, E. E. H. (2005). Pathological lying revisited. *Journal of the American Academy of Psychiatry and the Law Online, 33,* 342–349.

Dobbins, I. G., Schnyer, D. M., Verfaellie, M., & Schacter, D. L. (2004). Cortical activity reductions during repetition priming can result from rapid response learning. *Nature*, *428,* 316-319.

Druey, M., & Hübner, R. (2008a). Effects of stimulus features and instruction on response coding, selection, and inhibition: Evidence from repetition effects under task switching. *Quarterly Journal of Experimental Psychology*, 61, 1573-1600.

Druey, M., & Hübner, R. (2008b). Response inhibition under task switching: Its strength depends on the amount of task-irrelevant response activation. *Psychological Research*, 72, 515-527.

Duran, N. D., Dale, R., & McNamara, D. S. (2010). The action dynamics of overcoming the truth. *Psychonomic Bulletin & Review, 17*, 486–491.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28,* 1-11.

Foerster, A., Pfister, R., Schmidts, C., Dignath, D., Wirth, R., & Kunde, W. (in press). Focused cognitive control in dishonesty: evidence for predominantly transient conflict adaptation. *Journal of Experimental Psychology: Human Perception and Performance.*

Foerster, A., Wirth, R., Herbort, O., Kunde, W., & Pfister, R. (in press). Lying upside-down: Alibis reverse cognitive burdens of dishonesty. *Journal of Experimental Psychology: Applied.*

Foerster, A., Wirth, R., Kunde, W., & Pfister, R. (2017). The dishonest mind set in sequence. *Psychological Research, 81*, 878-899.

Frings, C., & Rothermund, K. (2011). To be or not to be…included in an event file: Integration and retrieval of distractors in stimulus–response episodes is influenced by perceptual grouping. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1209-1227.

Gamer, M. (2011). Detection of deception and concealed information using neuroimaging techniques (pp. 90-113). In Verschuere, Ben-Shakhar, & Meijer

(Eds.). *Memory detection: Theory and application of the concealed information test.* Cambridge, MA: Cambridge University Press.

Giesen, C., & Rothermund, K. (2013). You better stop! Binding ‘‘stop’’ tags to irrelevant stimulus features. *Quarterly Journal of Experimental Psychology, 67,* 1–24.

Giesen, C., & Rothermund, K. (2014). Distractor repetitions retrieve previous responses and previous targets: Experimental dissociations of distractor–response and distractor–target bindings. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 645-659.

Giesen, C., & Rothermund, K. (2016). Multi-level response coding in stimulus-response bindings: Irrelevant distractors retrieve both semantic and motor response codes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*, 1643-1656.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46,* 107-119.

Henson, R. N., Eckstein, D., Waszak, F., Frings, C., & Horner, A. J. (2014). Stimulus–response bindings in priming. *Trends in Cognitive Sciences*, *18,* 376-384.

Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences, 8,* 494-500.

Horner, A. J., & Henson, R. N. (2009). Bindings between stimuli and multiple response codes dominate long-lag repetition priming in speeded classification tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 757-779.

Hsu, Y. F., & Waszak, F. (2012). Stimulus-classification traces are dominant in response learning. *International Journal of Psychophysiology, 86,* 262-268.

Hübner, R., & Druey, M. D. (2006). Response execution, selection, or activation: What is sufficient for response-related repetition effects under task shifting? *Psychological Research, 70,* 245-261.

Hübner, R., & Druey, M. (2008). Multiple response codes play specific roles in response selection and inhibition under task switching. *Psychological Research,* 72, 415-424.

Johnson, R., Barnhardt, J., and Zhu, J. (2005). Differential effects of practice on the executive processes used for truthful and deceptive responses: an event-related brain potential study. *Cognitive Brain Research, 24,* 386–404.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin, 136,* 849-874.

Koch, I., Schuch, S., Vu, K. P. L., & Proctor, R. W. (2011). Response-repetition effects in task switching—Dissociating effects of anatomical and spatial response discriminability. *Acta Psychologica, 136,* 399-404.

Koranyi, N., Schreckenbach, F., & Rothermund, K. (2015). The implicit cognition of lying: Knowledge about having lied to a question is retrieved automatically. *Social Cognition, 33,* 67-84.

Liefooghe, B., & De Houwer, J. (2018). Automatic effects of instructions do not require the intention to execute these instructions. *Journal of Cognitive Psychology, 30,* 108-121.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492-527.

Logan, G. D., & Bundesen, C. (2003). Clever homunculus: Is there an endogenous act of control in the explicit task-cuing procedure? *Journal of Experimental Psychology: Human Perception and Performance, 29,* 575-599.

Mayr, U., & Kliegl, R. (2003). Differential effects of cue changes and task changes on task-set selection costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29, 362-372.

Meiran, N., Cole, M. W., & Braver, T. S. (2012). When planning results in loss of control: Intention-based reflexivity and working-memory. *Frontiers in Human Neuroscience*, *6*:104.

Moutsopoulou, K., Yang, Q., Desantis, A., & Waszak, F. (2015). Stimulus–classification and stimulus–action associations: Effects of repetition learning and durability. *The Quarterly Journal of Experimental Psychology, 68*, 1744-1757.

National Research Council (2003). *The polygraph and lie detection.* The National Academy Press: Washington, D.C.

Osman, M., Channon, S., & Fitzpatrick, S. (2009). Does the truth interfere with our ability to deceive? *Psychonomic Bulletin & Review, 16*, 901-906.

Pfeuffer, C. U., Hosp, T., Kimmig, E., Moutsopoulou, K., Waszak, F., & Kiesel, A. (2018). Defining stimulus representation in stimulus-response associations formed on the basis of task execution and verbal codes, *Psychological Research, 82,* 744-758.

Pfeuffer, C. U., Moutsopoulou, K., Pfister, R., Waszak, F., & Kiesel, A. (2017). The Power of Words: On item-specific stimulus-response associations in the absence of action. *Journal of Experimental Psychology: Human Perception and Performance, 43,* 328-347.

Pfeuffer, C. U., Moutsopoulou, K., Waszak, F., & Kiesel, A. (2018). Multiple priming instances increase the impact of practice-based but not verbal code-based stimulus-response associations, *Acta Psychologica, 184,* 100-109.

Pfeuffer, C. U., Pfister, R., Foerster, A., Stecher, F., & Kiesel, A. (2018). Binding Lies: Flexible retrieval of honest and dishonest behavior [data files, syntaxes, and experiments]. Retrieved from https://osf.io/cx269/

Pfister, R., Foerster, A., & Kunde, W. (2014). Pants on fire: The electrophysiological signature of telling a lie. *Social Neuroscience, 9*, 562-572.

Pfister, R., & Janczyk, M. (2013). Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology*, *9*, 74-80.

Polage, D. C. (2012). Fabrication inflation increases as source monitoring ability decreases. *Acta Psychologica, 139,*335–342.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General,* 124, 207-231.

Rothermund, K., Wentura, D., & De Houwer, J. (2005). Retrieval of incidental stimulus-response associations as a source of negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 482-495.

Schuch, *S*., & Koch, I. (2004). The costs of changing the representation of action: Response repetition and response-response compatibility in dual tasks. *Journal of Experimental Psychology: Human Perception and Performance, 30, 566-582.*

Spence, S. A., Farrow, T. F. D., Herford, A. E., Wilkinson, I. D., Zheng, Y., & Woodruff, P. W. R. (2001).Behavioural and functional anatomical correlates of deception in humans. *Neuroreport*,*12*, 2849–2853.

Steinhauser, M., Hübner, R., & Druey, M. (2009). Adaptive control of response preparedness in task switching. *Neuropsychologia,* 47, 1826-1835.

Van Bockstaele, B., Verschuere, B., Moens, T., Suchotzki, K., Debey, E., & Spruyt, A. (2012). Learning to lie: Effects of practice on the cognitive cost of lying. *Frontiers in Psychology, 3,* 177-184.

Vendemia, J. M. C., Buzan, R. F., and Green, E. P. (2005). Practice effects, workload, and reaction time in deception. *American Journal of Psychology, 5,* 413–429.

Verbruggen, F., & Logan, G.D. (2008) Long-term aftereffects of response inhibition: memory retrieval, task goals, and cognitive control. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1229–1235.

Verschuere, B., Spruyt, A., Meijer, E. H., & Otgaar, H. (2011). The ease of lying. *Consciousness and Cognition, 20,* 908–911.

Vrij, A., Fisher, R., Mann, S., & Leal, S. (2006). Detecting deception by manipulating cognitive load. *Trends in Cognitive Sciences, 10*, 141–142.

Vrij, A., Granhag, P. A., Mann, S., & Leal, S. (2011). Outsmarting the liars: toward a cognitive lie detection approach. *Psychological Science, 20,* 28–32.

Walczyk, J. J., Griffith, D. A., Yates, R., Visconte, S. R., Simoneaux, B., & Harris, L. L. (2012). Lie detection by inducing cognitive load: Eye movements and other cues to the false answers of "witnesses" to crimes. *Criminal Justice and Behavior, 39,* 887–909.

Walczyk, J. J., Harris, L. L., Duck, T. K., & Mulay, D. (2014). A social-cognitive framework for understanding serious lies: Activation-decision-construction-action theory. *New Ideas in Psychology, 34*, 22–36.

Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive lie detection: Response time and consistency of answers as cues to deception. *Journal of Business and Psychology, 24*, 33–49.

Walczyk, J. J., Roper, K. S., Seemann, E., & Humphrey, A. M. (2003). Cognitive mechanisms underlying lying to questions:Response time as a cue to deception. *Applied Cognitive Psychology, 17*, 755–774.

Waszak, F., Pfister, R., & Kiesel, A. (2013). Top-down vs. bottom-up: When instructions overcome automatic retrieval. *Psychological Research, 77*, 611-617.

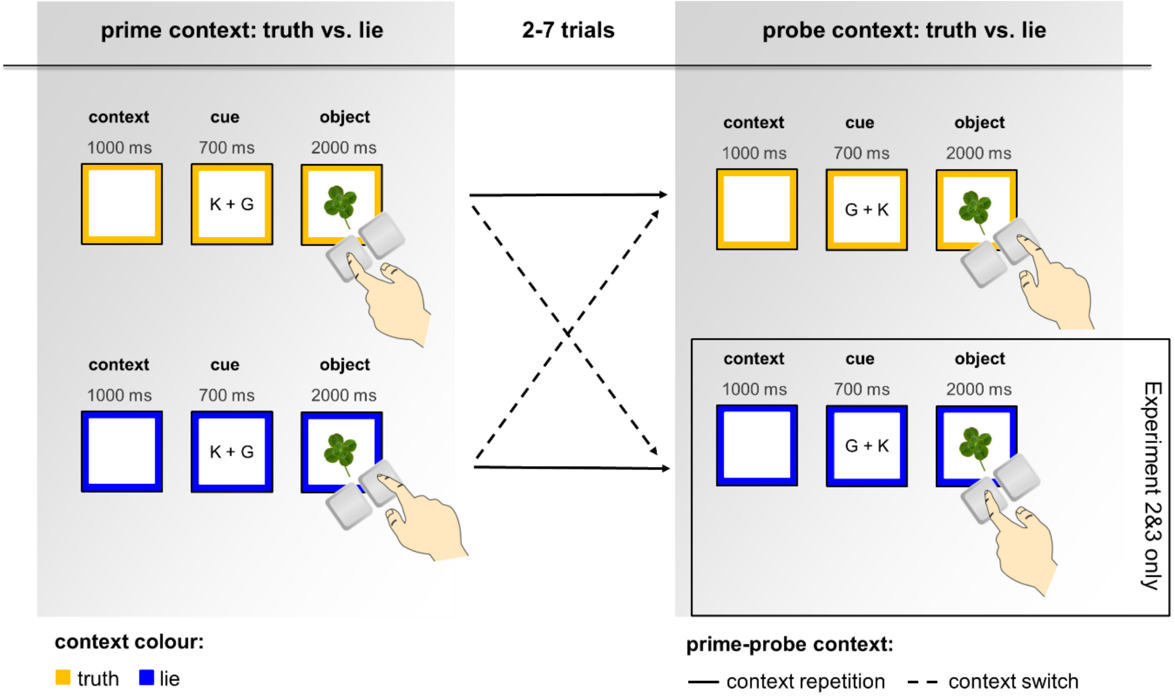

Figure 1. Different combinations of prime context and probe context in Experiment 1 (truth-truth vs. lie-truth) and Experiment 2 and 3 (truth-truth vs. lie-truth vs. lie-lie vs. lie-truth) as well as the structure of individual (truth or lie) trials. A colored frame (1000 ms; two colours per intentional context in Experiment 3) indicated the current truth-lie context and a cue (700 ms) instructed the classification-response mapping. The cue was followed by the object image (until response, maximum 2000 ms). Inaccurate classifications (inappropriate classifications during truth trials or accidentally appropriate classification during lie trials) as well as response omissions were followed by appropriate feedback (500 ms). Each item appeared once as a prime and once as a probe (2-7 trials after the prime).

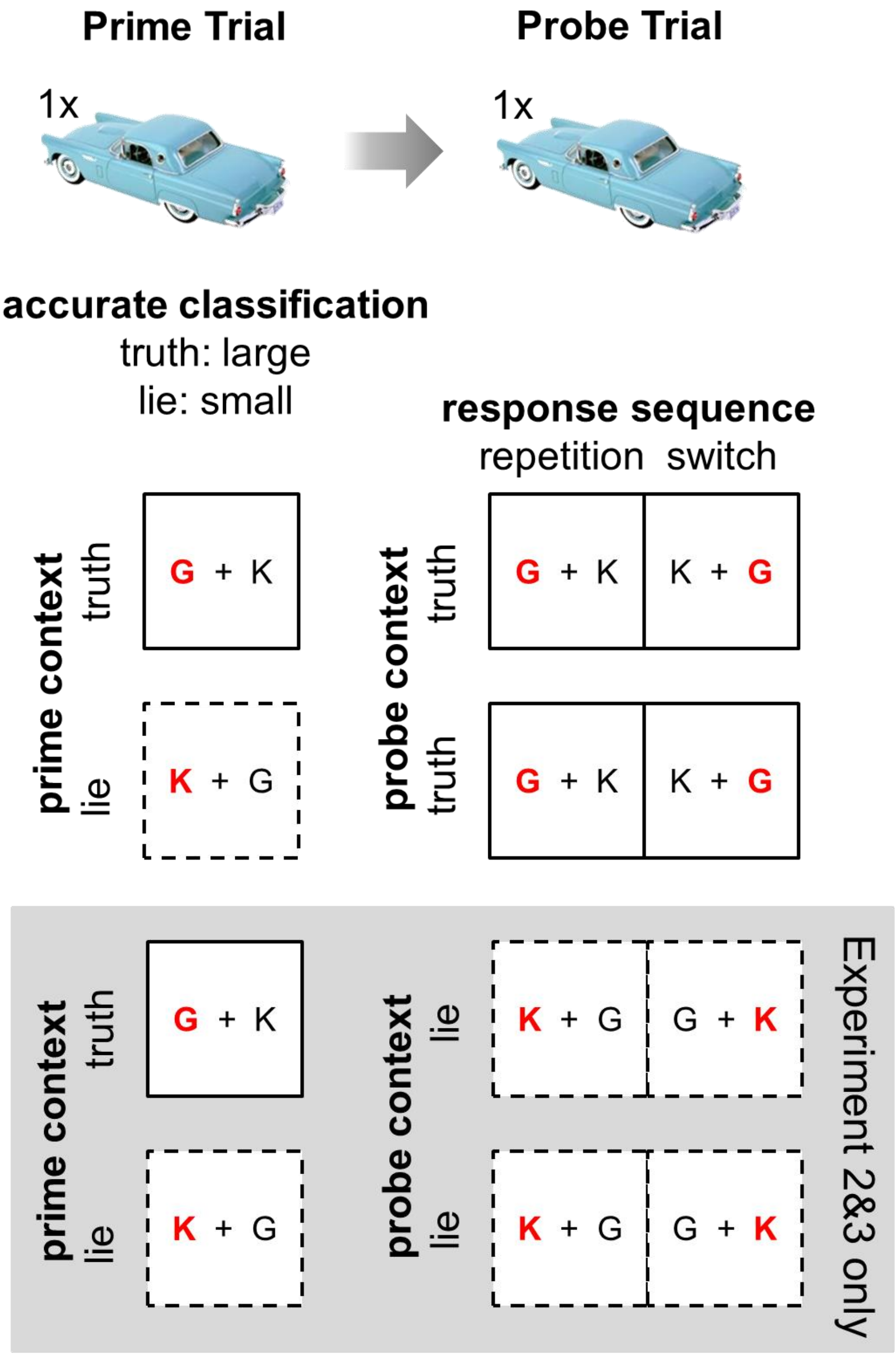

Figure 2. Overview of the prime-probe response sequence in Experiments 1 and Experiments 2 and 3. Independent of the current context (truth vs. lie), response mappings could either repeated or switch between the prime of an object and its corresponding probe (2-7 trials later). Accurate responses are marked as red and bold.

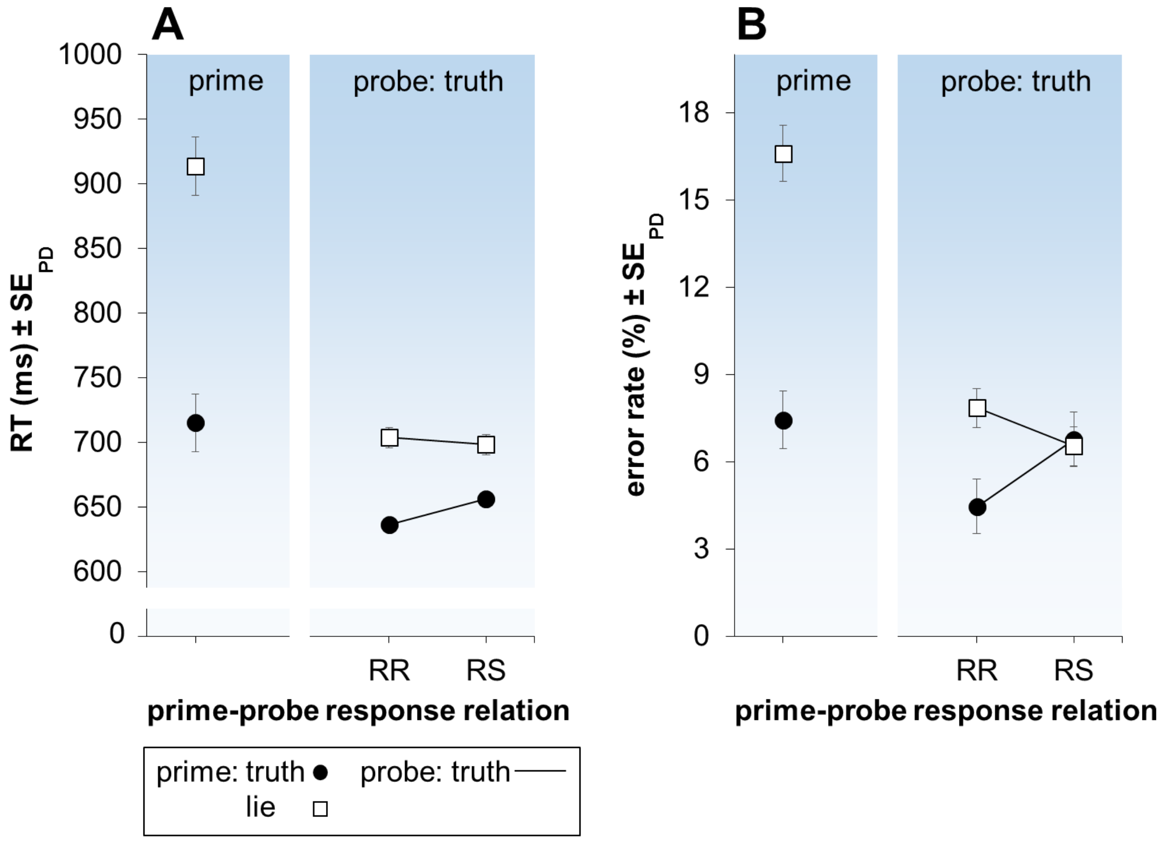

Figure 3. Prime and probe trial A) RTs and B) error rates of Experiment 1. Probe trial RTs and error rates are plotted as a function of the factors prime context (truth vs. lie), probe context (truth vs. lie), and response sequence (response repetition, RR, vs. response switch, RS). Error bars indicate standard errors of paired differences computed separately for each comparison of response repetition and response switch trials (Pfister & Janczyk, 2013).

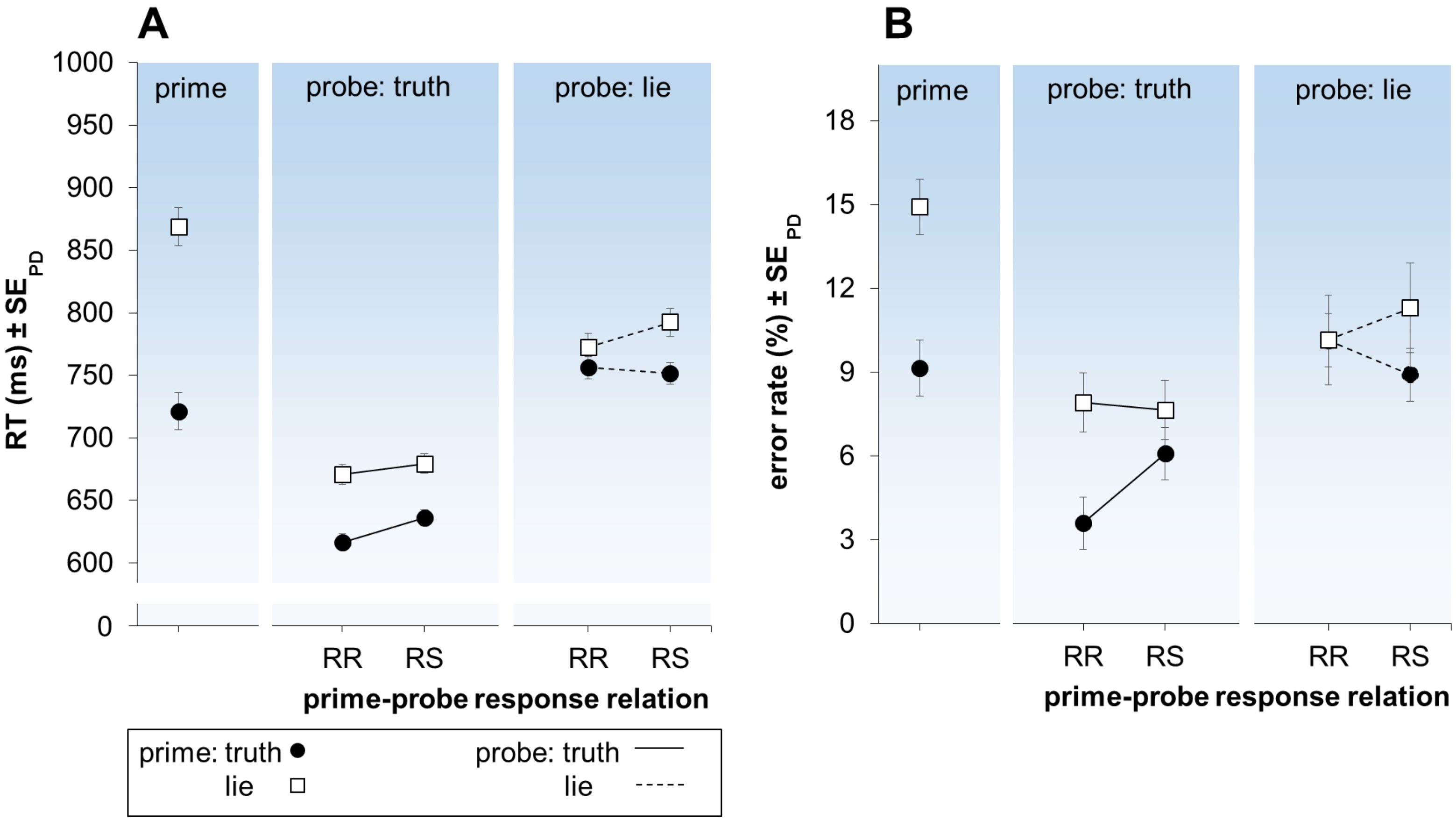

Figure 4. Prime and probe trial A) RTs and B) error rates of Experiment 2. Probe trial RTs and error rates are plotted as a function of the factors prime context, probe context, and response sequence. Error bars indicate standard errors of paired differences computed separately for each comparison of response repetition and response switch trials (Pfister & Janczyk, 2013).

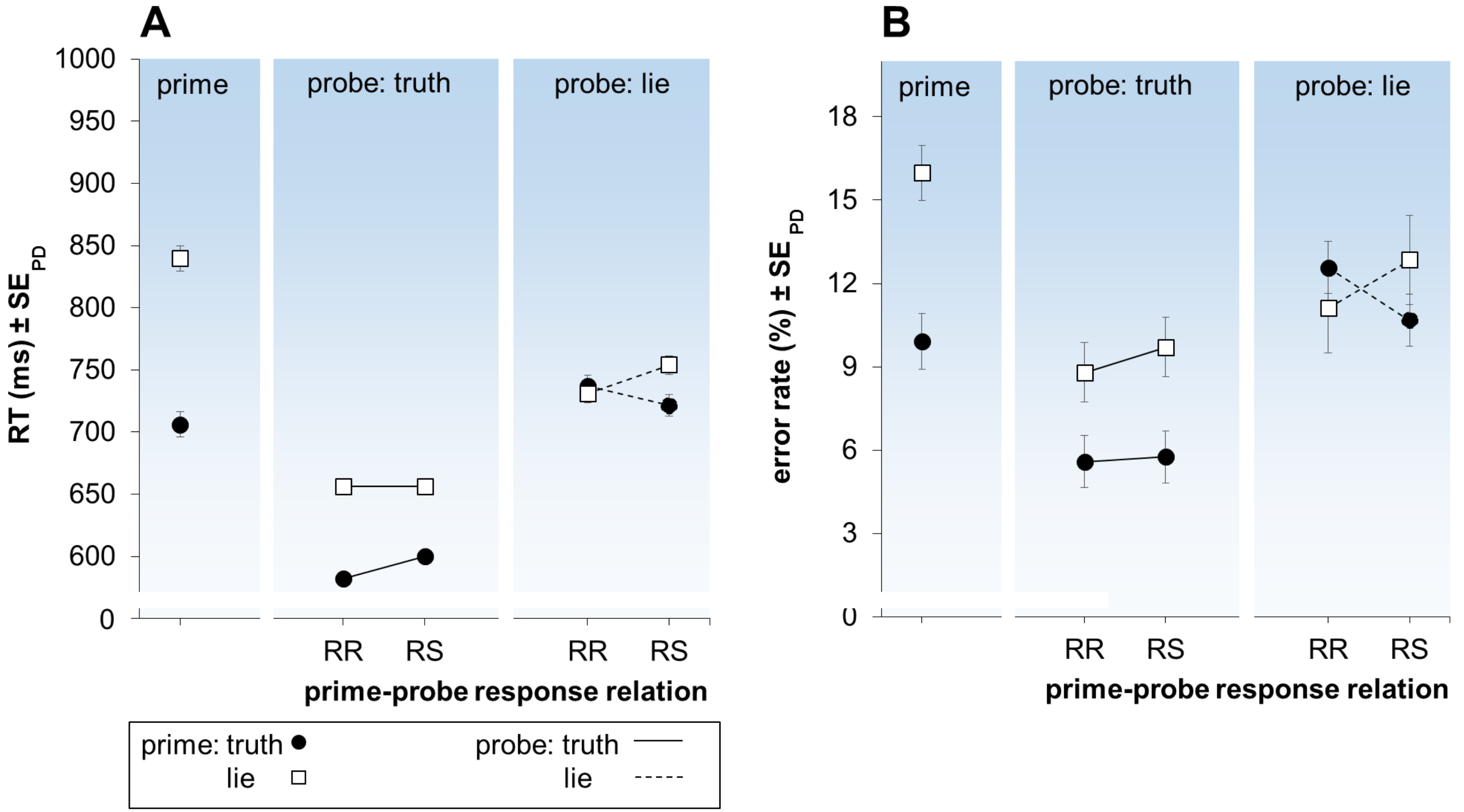

Figure 5. Prime and probe trial A) RTs and B) error rates of Experiment 3. Probe trial RTs and error rates are plotted as a function of the factors prime context, probe context, and response sequence. Error bars indicate standard errors of paired differences computed separately for each comparison of response repetition and response switch trials (Pfister & Janczyk, 2013).

Appendix:

Prime-probe comparisons for Experiments 2 and 3

As mentioned in the Discussion of Experiment 2, when the context repeated between prime and probe, the classification participants provided also repeated, whereas it switched when the context switched. For example, a car was categorized as large in the prime trial as well as in the probe trial when the honest context repeated and two times as small when the dishonest context repeated. In contrast, the correct classification of the car switched from large to small or from small to large when the intentional context switched between prime and probe. Our previous analyses do not provide any information on whether this repetition/switch in classification might also (partly) account for the pattern of results we observed.

A comparison with experiments assessing item-specific switches between two classification tasks that also entail a switch in classification between prime and probe (e.g., Giesen & Rothermund, 2016; Horner & Henson, 2009; Moutsopoulou et al., 2015; Pfeuffer et al., 2017) suggests that repetitions/switches in classification cannot account for the present findings regarding context-specific S-R retrieval. Specifically, in these experiments, it has consistently been found that each stimulus became independently associated with the action a participant performed to classify it (S-A association) and its task-specific semantic classification (S-C association). That is, costs associated with switches in the S-C and S-A mapping were additive and there was no interaction that would have indicated an interdependent association between stimulus, classification, and action. Thus, as the retrieval of S-A associations is unaffected by repetitions/switches in classification, these studies suggest that differences in S-A retrieval effects between conditions cannot be accounted for by assuming an influence of repetitions/switches in classification.

Inspired by a study by Osman, Channon, and Fitzpatrick (2009), here we will report an additional post-hoc analysis of the data of Experiments 2 and 3 that supports the conclusion that the context-specific S-R retrieval we observed in the probe did not occur due to repetitions/switches in classification. This analysis also provides further information on the associative foundations of lying versus truth-telling.

Osman et al. (2009) found that when participants first responded dishonestly and then honestly to the same set of questions, RTs did not differ between honest responses that participants provided for questions they had initially lied to and first-time honest responses to a different set of questions. Having lied to a question thus does not seem to affect later honest responding, supporting our notion that responses are not retrieved automatically when switching between an honest and dishonest mindset. However, participants´ RTs for first-time honest responses to questions they had responded to dishonestly before could have been influenced by the change in the given answer. Thus, these findings cannot provide conclusive evidence regarding the context-specific retrieval of S-R associations.

This alternative explanation, indicating that the change of the answer participants gave could have influenced the results of Osman et al. (2009), is based on the logic that, in case a stimulus has previously been associated with a semantic classification, RTs should be slower when the semantic classification a person is supposed to provide for a stimulus changes. Applying this logic, in the present study we can use RT and error rate differences between prime and probe trials to assess the impact of repetitions/switches in classification between prime and probe on probe performance.

Additionally, such an analysis can ideally provide further evidence for the notion that the context-specific S-R (S-A) retrieval effects we observed were not influenced by repetitions/switches in classification between the prime and probe of a stimulus. That is, in our previously reported probe analyses, we found the same pattern of context-specific S-R retrieval effects for both the honest and dishonest intentional context. Regarding the additional analyses of performance differences between prime and probe trials, indicating the influence of repetitions/switches in classification per se, two patterns of results are possible for the two intentional contexts. If prime-probe RT and error rate differences for honest and dishonest responses were similarly affected by context repetitions/switches, this might indicate that classification repetitions/switches between prime and probe had an impact on S-R (S-A) retrieval. At least it would not allow us to rule out such an influence. However, if, conversely, prime-probe RT and error rate differences for honest and dishonest responses were differently affected by context repetitions/switches, this would provide further tentative support for our hypothesis that classification repetitions/switches did not affect context-specific S-R retrieval in the probe. That is, context-specific S-R retrieval patterns in the probe were equivalent for honest and dishonest contexts. Thus, if repetitions/switches in classification were (partly) responsible for this pattern of results, the effects of repetitions/switches in classification on honest and dishonest probe responses (i.e., the performance differences between prime and probe) would also have to be similar for the two intentional contexts. Conversely, if we found differences between the two intentional contexts regarding the impact of classification repetitions/switches, this would further support the notion that repetitions/switches in classification between prime and probe cannot account for the context-specific S-R retrieval effects we observed.

Results. Per participant, we computed difference measures for RTs and error rates as prime mean minus probe mean separately for the two prime/probe contexts (truth vs. lie) and the two context sequences (context repetition vs. context switch). Differences were computed by subtracting the individual RT/error rate of the probe context (truth vs. lie) from the RT/error rate of the same prime context (truth vs. lie) independent from the probe context that followed it in the experiment (e.g., $\Delta\mathrm{RT}_{\mathrm{truth,context\ repetition}}$ = $\mathrm{RT}_{\mathrm{Prime\ truth-(truth/lie)}}$ - $\mathrm{RT}_{\mathrm{Probe\ truth-truth}}$, $\Delta\mathrm{RT}_{\mathrm{truth,context\ switch}}$ = $\mathrm{RT}_{\mathrm{Prime\ truth-(truth/lie)}}$ - $\mathrm{RT}_{\mathrm{Probe\ lie-truth}}$). That is, the reported prime-probe differences always constituted differences between equivalent prime and probe contexts (i.e., between prime truth and probe truth or between prime lie and probe lie) irrespective of whether the respective honest/dishonest prime trial was presented in the same block as the probe trial. Prime truth/lie averages were computed irrespective of the probe context of a block. Prime-probe differences were then analyzed with respect to whether the probe data stemmed from a context repetition (i.e., truth-truth or lie-lie) or a context switch (i.e., truth-lie or lie-truth) block. We conducted 2 x 2 repeated measures ANOVAs with the factors probe context (truth vs. lie) and context sequence (context repetition vs. context switch) on the prime-probe RT and error rate differences of Experiments 2 and 3 (see Figure A1).

Experiment 2. In RTs, we found a main effect of probe context, $F(1,23) = 8.13$, $p = .009$, $\eta_p^2 = .26$. RT decreases from prime to probe were more pronounced for probe honest as compared to probe dishonest responses. Furthermore, the interaction between probe context and context sequence was significant, $F(1,23) = 36.67$, $p < .001$, $\eta_p^2 = .62$. Paired *t*-tests further examining this interaction showed that, when participants responded honestly in the probe, RT decreases from prime to probe were significantly smaller when the context (i.e., the stimulus classification) switched rather than repeated, $t(23) = 4.73$, $p < .001$, $d_z = 0.97$. Conversely, when participants responded dishonestly, prime-probe RT differences increased when the context (i.e., the stimulus classification) switched as compared to when the context repeated, $t(23) = -2.28$, $p = .033$, $d_z = -0.46$. The main effect of context sequence failed to reach significance, $F(1,23) = 1.26$, $p = .273$, $\eta_p^2 = .05$.

In participants´ error rates, only the interaction of probe context and context sequence was significant, $F(1,23) = 10.64$, $p = .003$, $\eta_p^2 = .32$. For honest probe responses, prime-probe error rate differences were larger when the context (i.e., the stimulus classification) repeated rather than switched, $t(23) = 3.18$, $p = .004$, $d_z = 0.65$. For dishonest probe responses, prime-probe error rate differences were larger when the context (i.e., the stimulus classification) switched rather than repeated, $t(23) = -2.16$, $p = .041$, $d_z = -0.44$. The main effects of context, $F(1,23) = 2.02$, $p = .169$, $\eta_p^2 = .08$, and context sequence, $F(1,23) = 1.56$, $p = .225$, $\eta_p^2 = .06$, failed to reach significance.

Experiment 3. Participants´ prime-probe RT differences showed significant main effects of both probe context, $F(1,43) = 7.91$, $p = .007$, $\eta_p^2 = .16$, and context sequence, $F(1,43) = 25.29$, $p < .001$, $\eta_p^2 = .37$. Participants showed larger prime-probe RT differences when responding dishonestly rather than honestly and for context (i.e., stimulus classification) repetitions rather than context (i.e., stimulus classification) switches. Most importantly, the interaction of probe context and context sequence also reached significance, $F(1,43) = 72.77$, $p < .001$, $\eta_p^2 = .63$. For honest probe responses, prime-probe RT differences were larger when the context (i.e., the stimulus classification) repeated rather than switched, $t(43) = 10.21$, $p < .001$, $d_z = 1.54$. Conversely, for dishonest probe responses, prime-probe RT differences were larger when the context (i.e., the stimulus classification) switched rather than repeated, $t(43) = -2.05$, $p = .047$, $d_z = -0.31$.

Similarly, participants´ prime-probe error rate differences showed main effects of probe context, $F(1,43) = 11.10$, $p = .002$, $\eta_p^2 = .21$, and context sequence, $F(1,43) = 13.75$, $p = .001$, $\eta_p^2 = .24$. Error rate differences were larger for dishonest rather than honest probe response and for context (i.e., the stimulus classification) repetitions rather than context switches. Again, the interaction of probe context and context sequence was significant, $F(1,43) = 14.76$, $p < .001$, $\eta_p^2 = .26$. For honest probe responses, participants´ prime-probe error rate differences were larger when the context (i.e., the stimulus classification) repeated rather than switched, $t(43) = 5.58$, $p < .001$, $d_z = 0.84$. For dishonest probe responses, there was no significant difference, $t(43) = -0.35$, $p = .731$, $d_z = -0.05$.

Discussion. Our post-hoc prime-probe comparison analysis of Experiments 2 and 3 provided further tentative evidence against an influence of repetitions/switches in classification between prime and probe on the observed context-specific retrieval of item-specific S-R associations. Context repetitions/switches similarly affected S-R retrieval effects in the probe trials irrespective of whether participants were responding honestly or dishonestly. However, prime-probe performance differences, reflecting probe performance costs associated with context (i.e., classification) switches, showed opposing patterns for honest and dishonest responses. This indicates that switches in classification affected honest and dishonest probe responses differently.[3] As repetitions/switches in classification had opposing effects on performance depending on whether participants were in an honest or dishonest intentional context, repetitions/switches in classification could not conceivably have simultaneously led to the same pattern of results for honest and dishonest responses in the probe. Thus, we conclude that switches in classification between prime and probe could not have caused the observed pattern of results in the probe.

Our analysis thus further corroborates our comparative assessment of the present findings in contrast to studies of Giesen and Rothermund (2016), Horner and Henson

[3] Please note that probe response means were subtracted from prime response means in the same intentional context to compute prime-probe performance differences. As such, smaller/larger prime-probe differences for context repetitions/switches reflect not only how the prime-probe differences were affected. They simultaneously reflect how the probe responses of one intentional context (honest/dishonest) were affected by context repetitions/switches, as the same honest/dishonest prime trials were used for reference both for context repetitions and context switches.

(2009), Moutsopoulou et al. (2015), and Pfeuffer et al. (2017). It indicates that a switch between the intentional contexts of lying and truth-telling is distinctly different from the switch between two classification tasks. Additionally, this further supports previous findings suggesting that classifications are associated with stimuli independent from responses (e.g., Giesen & Rothermund, 2016; Horner & Henson, 2009; Moutsopoulou et al., 2015; Pfeuffer et al., 2017).

Furthermore, we observed that when the context switched from prime to probe, participants´ performance improved more for the lie-truth than the truth-lie prime-probe context combination. This finding is remarkable given that lying as compared to truth-telling is associated with costs and participants should therefore have benefited less when they had to lie in the probe. Yet, it has been suggested that lying is accompanied by the inhibition of the honest response (e.g., Walczyk et al., 2014; Walczyk, Roper, Seemann, & Humphrey, 2003). Indeed, an item-specific inhibition of the honest response (i.e., the appropriate size classification) during the prime trial when S-R associations are formed could account for the observed result pattern. If participants inhibited the honest classification of an object during the prime, classifications may have been bound to the stimulus with an inhibitory tag (see e.g., Giesen & Rothermund, 2013; Verbruggen & Logan, 2008, for evidence that inhibitory STOP tags can be bound to S-R bindings/associations). That is, in case participants experienced first the lie and then the truth context, during a dishonest prime response, the honest classification would have been associated with an inhibitory tag. This could have led to increased reaction times and error rates in the subsequent honest probe trial when the honest classification had to be retrieved. Conversely, when participants classified objects in truth-lie blocks, neither the honest nor dishonest response became associated with an inhibitory tag in the prime, leading to overall larger performance benefits in the probe compared to lie-truth blocks. As such, our findings additionally provide further direct evidence for the theoretical assumption that the honest answer becomes inhibited when we lie.

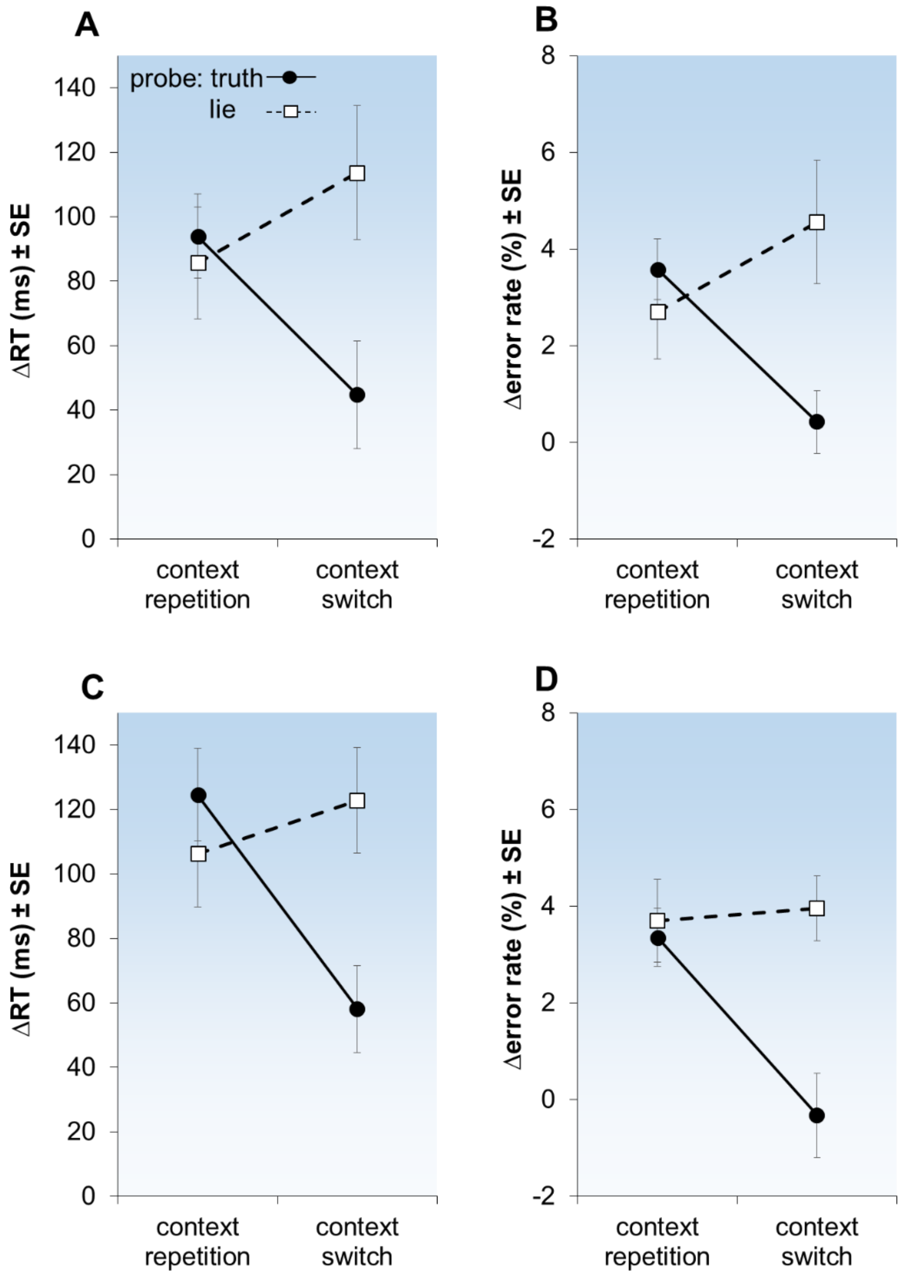

Figure A1. A and C) RT and B and D) error rate differences between prime and probe trials (ΔRT = $\mathrm{RT}_{\mathrm{Prime}}$ - $\mathrm{RT}_{\mathrm{Probe}}$) in Experiments 2 (A and B) and 3 (C and D) displayed separately for the two probe contexts (truth vs. lie) and the two context sequences (context repetition vs. context switch). Differences were computed by subtracting the RT/error rate of the probe context (truth vs. lie) from the RT/error rate of the same prime context (truth vs. lie) independent from the probe context it was associated with (e.g., $\Delta\mathrm{RT}_{\mathrm{truth,context\ repetition}}$ = $\mathrm{RT}_{\mathrm{Prime\ truth-(truth/lie)}}$ - $\mathrm{RT}_{\mathrm{Probe\ truth-truth}}$).