The Nature of Knowledge Reliance in Source Guessing

Liliane Wulff

Inaugural Dissertation
Submitted in partial fulfillment of the requirements for the degree Doctor of Social Sciences in the Graduate School of Economic and Social Sciences at the University of Mannheim

Thesis Defense:
May 29, 2020

Supervisors:
Prof. Dr. Beatrice G. Kuhlmann
Prof. Dr. Edgar Erdfelder

Dean of the School of Social Sciences:
Prof. Dr. Michael Diehl

Thesis Reviewers:
Prof. Dr. Arndt Bröder
Prof. Dr. Edgar Erdfelder

Defense Committee:
Prof. Dr. Beatrice G. Kuhlmann
Prof. Dr. Arndt Bröder
Prof. Dr. Edgar Erdfelder

For Ansgar.

⚓

When all my dreams come true, the one I want next to me.
It's you. It's you.

from *One Tree Hill*
Season 4 - Episode 9

Table of Contents

Summary

Context matters for complex human information processing. The ability to attribute information to its origin (*source*) is not only crucial for daily social interactions (e.g., who told me something?) and impression formation about other people, but it is also essential to judge the credibility of sources and validity of the received information (e.g., where did I read something?). Even more so, failures of source attributions can have far-reaching consequences, for instance in the distribution of fake news in the recent media debate but also in eyewitness testimony where misattributions become even more severe (i.e., sentencing innocent individuals). Retrieving the origin of an episode from memory has been termed *source monitoring* (cf. M. K. Johnson et al., 1993). Alongside memory processes, source attributions can be based on guessing processes affected by general or contingency knowledge and plausibility or metacognitive beliefs in case a source's characteristic cannot be retrieved from memory (cf. M. K. Johnson et al., 1993). This thesis specifically focuses on knowledge reliance in source guessing. That is, when people infer the source due to a lack of memory traces, they rely on knowledge acquired prior to or during the learning environment. While mostly memory processes were of prime research interest in the past decades, guessing processes were given less consideration.

In this thesis, I address the underlying nature of knowledge reliance in source guessing more thoroughly as part of four manuscripts, thereby contributing to put the so far missing pieces of the source-guessing puzzle into place. In these manuscripts, I examine knowledge reliance in source guessing with regard to its cognitive dynamics, stability, resource dependence, and generalization to novel stimuli to serve the overarching objective of inferring its underlying nature. In the first manuscript, I demonstrate the utility of the process-tracing methodology mouse tracking to unpack the influence of knowledge reliance on source-monitoring processes. In the second manuscript, I quantify the extent to which individual differences in knowledge reliance in source guessing are stable across time and knowledge domains. In the third manuscript, I refine the understanding of the underlying automatic or controlled mechanisms of knowledge reliance. In the fourth manuscript, I expand the scope of knowledge reliance in source guessing to novel information contexts. In sum, this thesis provides new insights into the application of knowledge structures in judgmental processes under source uncertainty.

Manuscripts

The present cumulative thesis is targeted to foster a comprehensive framework of the underlying nature of knowledge-based influences on source guessing—a judgment process that emerges whenever memory for the original context during source attributions is absent. The reported research is conducted in the *Center for Doctoral Studies in Social Sciences* (CDSS) of the *Graduate School of Economic and Social Sciences* (GESS) at the University of Mannheim and is based on four manuscripts.

The first three manuscripts provide insights into the cognitive dynamics (Manuscript I), the state- versus trait-like characteristic (Manuscript II), and the automatic versus controlled nature (Manuscript III) of knowledge reliance in source guessing. The last manuscript (Manuscript IV) broadens the scope of knowledge-based influences on source guessing by investigating a generalization to decision contexts for new information. Manuscript I and II are published, Manuscript III has been submitted for publication, and Manuscript IV is prepared for submission.

In the main text of this thesis, I review the overarching theoretical framework for the four manuscripts, followed by a presentation of the joint statistical approach to model cognitive processes involved in source monitoring. Next, I provide a summary of each of the four manuscripts before a general discussion of the results, their strengths and limitations, theoretical and practical implications, and prospective research ideas complement the thesis. Empirical and analytical specificities of the distinct experiments are outlined in the original manuscripts appended to this thesis (in the same order as listed below).

Manuscript I

Wulff, L., & Scharf, S. E. (2020). Unpacking stereotype influences on source-monitoring processes: What mouse tracking can tell us. *Journal of Experimental Social Psychology*, *87*, 103917. https://doi.org/10.1016/j.jesp.2019.103917

Manuscript II

Wulff, L., & Kuhlmann, B. G. (2020). Is knowledge reliance in source guessing a cognitive trait? Examining stability across time and domain. *Memory & Cognition*, *48*(2), 256–276. https://doi.org/10.3758/s13421-019-01008-1

Manuscript III

Wulff, L., & Kuhlmann, B. G. (2020). *Is knowledge reliance in source guessing automatic or controlled? Evidence from divided attention and aging*. Manuscript submitted for publication.

Manuscript IV

Wulff, L., Bell, R., Mieth, L., & Kuhlmann, B. G. (2020). *Guess what?! Different source-guessing mechanisms for old versus new information*. Manuscript in preparation.

Make it count.

1 Introduction

Searching for articles on "source guessing" in the electronic literature search engine *Google Scholar* illustrates: Judgment processes in attributions about the origin of information were far less studied in empirical research on source monitoring (143 articles) compared to source memory (about 16,400 articles) which was the main focus of research in the past decades (time period from 1980 to 2020; retrieved March 18, 2020).[1] Oftentimes, guessing processes in source monitoring were not of primary interest and less in the center of attention in the scientific discourse which is why its underlying mechanisms are understudied in comparison to memory processes in source monitoring. Yet, source guessing in itself is a fascinating cognitive process that is worth dwelling on, not just from a naive perspective of a basic researcher but also from an applied point of view (as source attributions can be performed under uncertainty and distraction in everyday life). Thus, this thesis is concerned with the underlying nature of source guessing and, in particular, the influence of pre- and peri-experimental knowledge on this judgment process.

The theoretical distinction between memory and guessing processes in source monitoring (M. K. Johnson et al., 1993) has been transferred to stochastic models which made the separate measurement of cognitive processes that contribute to the performance in source attributions possible. Disentangling memory from judgment processes has revealed that source guessing is biased by knowledge (cf. Kuhlmann & Bayen, 2016, for a review). That is, previous research has demonstrated that people make inferences of the origin about information based on either generic knowledge (such as stereotypes, schemas, or plausibility beliefs) acquired prior to the test situation in (laboratory) experiments in which source monitoring is studied (e.g., W. H. Batchelder & Batchelder, 2008; Bayen et al., 2000; Bell et al., 2015; Kuhlmann et al., 2016; Spaniol & Bayen, 2002) or contingency knowledge acquired during the course of encoding in the specific learning environment (e.g., Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012; Spaniol & Bayen, 2002).

The cognitive processes involved in source monitoring can be studied in a basic paradigm. In such tasks, it is common that participants study information (e.g., words) accompanied by contextual features (e.g., different faces indicating articulation). In a subsequent memory test, participants attribute the previously-studied information (among new infor-

[1] As a matter of course, acknowledging that the terminology for source memory and guessing may be manifold in the literature (e.g., "context memory", "guessing bias", ...), summarized under the term *source monitoring*, and all relevant articles are not entirely covered by this literature data base, it may nonetheless provide a first indication for the current status of research.

mation) to the context in which the information was studied or not. Specifically, participants decide whether the information was presented before (“old”), and thereby associated with different context features (i.e., whose face was paired with the information during learning), or not (“new”), thereby implying that the information was not previously encountered. Multinomial models for categorical response data collected in such experiments (e.g., Bayen et al., 1996) enable researchers to estimate memory and guessing processes separately. Among other things, the application of such models revealed two essential findings in terms of source guessing that are significant to the present thesis: First, knowledge reliance in the source-guessing process itself is not as random as one may assume when intuitively thinking of the concept "guessing" and, second, individuals strongly differ in their systematic reliance on knowledge structures (e.g., Arnold et al., 2013; Spaniol & Bayen, 2002). This, in turn, raises the questions of what drives individuals’ source-guessing behavior and under which conditions it is most or least pronounced?

In this thesis, I take steps towards answering these questions by examining why and to what extent individuals show biased source guessing due to knowledge reliance. By this, I aim at supplementing the bigger picture of this judgment process in the absence of retrieval of contextual details. That is, I raise the overarching question of which underlying mechanisms characterize knowledge reliance in source guessing best? This thesis thereby contributes to a more profound understanding on a conceptual and theoretical level but may also serve a more applied, societal purpose at the same time. Determining why, how, and when individuals show knowledge-based source guessing is a prerequisite for well-constructed, precise interventions to overcome a knowledge reliance and to evaluate potential long-term effects of such biases. For instance, we might learn about the consequences of biased source guessing on how we process (social) information in the first place and, building on this, how we hereafter perceive others and form impressions about them.

In the following, I provide a more detailed introduction to the theoretical framework of source monitoring and its statistical implementation—followed by a review of the empirical evidence for the nature of knowledge reliance in source guessing based on four manuscripts. This thesis concludes with a discussion about the strengths, limitations, and implications of the present findings, and gives an outlook on future research.

2 Theoretical Foundation: The Source-Monitoring Framework

> Source monitoring of all types is based on characteristics of memories in combination with judgment processes.
>
> *(Johnson, Hashtroudi, & Lindsay, 1993)*

2.1 Memory and Judgment Processes

The umbrella term *source monitoring* encompasses all cognitive processes involved in the reconstructive inference process when attributing information to its origin (M. K. Johnson et al., 1993; Mitchell & Johnson, 2000, 2009). In their seminal review, Marcia K. Johnson and her colleagues (1993) proposed a superordinate theoretical framework to describe the cognitions at play during source attributions in more detail. The authors defined source monitoring as a "[...] set of processes involved when making attributions about the origins of memories, knowledge, and beliefs" (p. 3). According to M. K. Johnson et al. (1993), *sources* refer to all features of learning episodes under which memory traces are acquired. Sources are thereby not stored and retrieved as an abstract entity but rather based on detailed cues from various modalities (see also Mitchell & Johnson, 2000). Among others, this could be, for instance, perceptual, temporal, or affective information bound to the episode in which individuals form a memory representation. But on which decision criteria are source attributions based in case of inability to retrieve specific characteristics from past contextual episodes?

A second integral part of source monitoring are strategic and heuristic decision components. These are, following the theoretical source-monitoring framework (M. K. Johnson et al., 1993), biases, metacognitive awareness, plausibility beliefs but also current motivations and intentions that can have a significant impact on how source attributions are made.

In summary, source monitoring can, thus, be broken down by two essential cognitive operations that take place during attributions of information to its origin: (1) *Memory processes*, that have their roots in the specific situation in which contextual binding was experienced and a memory trace was acquired, and (2) *judgment processes*, whose evaluative character constitutes a symbolic weighting function of the strategic decision criterion (cf. M. K. Johnson et al., 1993). The source-monitoring framework highlighted the dissociation of the cognitive processes in source inference—predominantly from a theoretical perspective. Plenty of empirical support for this dissociation of memory and judgment processes was provided shortly thereafter and up until now.

But how can memory and judgment processes be disentangled and inferred from the performance collected in experimental source-monitoring tasks? A closer inspection of the data collected during source-monitoring paradigms reveals that by just looking at the categorical responses (e.g., decision between different sources) the underlying cognitive processes that lead to the behavior in the first place cannot be derived. The confound of memory and judgment processes has been addressed in the mathematical modeling of such cognitive processes involved in source monitoring as illustrated in Figure 1 and is introduced and described in detail in Section 3.

2.2 Source Guessing at a Glance

The development of multionomial processing tree (MPT) models such as the *two-high-threshold multinomial model of source monitoring* (2THSM; Bayen et al., 1996) was a milestone towards the quantification of latent cognitive processes and enabled researchers to perform hypothesis tests examining influencing factors specific to single memory and judgment parameters. Oftentimes, the source-memory parameter was of particular interest to the research community. As initially stated, a large proportion of studies mirrors that source-monitoring research primarily focused on the memory component. Nonetheless, judgment processes have attracted attention ever since the development of mathematical models to disentangle cognitive processes and are interesting for a variety of applied settings, for instance for eyewitness suggestibility and the phenomenon of inducing false memories (cf. M. K. Johnson et al., 1993; Mitchell & Johnson, 2000). Judgment processes comprise the more systematic components of source monitoring. The term "guessing" may, at first sight, imply randomness in the meaning of non-strategic source attributions under uncertainty. In fact, source guessing actually has been far from random and judgments can rather be described as the best educated guess an individual can make based on various cues such as what is already known about the sources in general. That is, individuals can initiate guessing processes to infer the source based on qualitatively distinct cues such as reasoning, plausibility beliefs, or generic knowledge (M. K. Johnson et al., 1993).

In this thesis, I will specifically focus on knowledge reliance in source guessing based on two knowledge cues: the item-source contingency and stereotypes/schemas both of which have been identified as crucial backup strategies in situations when individuals are not able to retrieve the source from memory (Kuhlmann & Bayen, 2016). Empirical evidence on the multifaceted influences of knowledge cues on the source-guessing process is reviewed hereafter.

One cue of which individuals can make use of is the item-source contingency (referred to as peri-experimental knowledge). In an experimental setting, the contingency is most frequently manipulated—with material for which no prior knowledge representation exists among participants (Ehrenberg & Klauer, 2005; Klauer & Meiser, 2000)—by varying the probability of certain pairings of item types and sources to occur during encoding (e.g., 75% of the items are paired with a specific context, for instance a certain screen position). Throughout the course of encoding, individuals can then perceive the predominant contingency such that some items are more likely accompanied by one source as by another. In an applied field setting, a contingency could be that certain information is occurs factually-justified more often in a specific situation. This on-line perception of item-source contingencies while learning is mirrored in source guessing during source attributions when memory for the source is absent and, in addition, no prior knowledge is applied. In these instances, source guessing is then biased in the direction of the contingency manipulation demonstrating that individuals indeed learn that item types are associated with specific source characteristics (e.g., Ehrenberg & Klauer, 2005).

Another essential source-monitoring cue is prior knowledge such as stereotypes and schemas. Compared to peri-experimental knowledge, this pre-experimental knowledge is acquired at some point before the source-attribution process is initiated. Mental structures such as stereotypes (i.e., beliefs about personal attributes of social groups; cf. Stroebe & Insko, 1989) and schemas (i.e., mental structures that organize and link information based on previous experience; cf. Alba & Hasher, 1983) are one such example and can then find their application in source guessing to reconstruct the origin of information. In an experimental setting, researchers usually present information that are inherently prototypical for its origin (e.g., persons, situations, or scenes). An empirical example from Kuhlmann et al. (2016) nicely illustrates the general procedure. The authors used verbal item material (statements) that was either expected to be said by young or old adults (based on a survey study by Kuhlmann et al., 2017). These expected-young and -old statements were presented by two person sources one of which was introduced as a young person and the other as a an old person (without revealing the specific assignment of the ages to the sources during encoding). Either source presented the same number of items from each expectancy type. This restriction in the assignment of statement to the sources ensured that each item was to be associated with either source with an equal probability (= null contingency). In a later memory test, participants attributed the previously-encountered statements to the sources or classified them as new. MPT model-based analyses of their categorical responses suggest that source guessing was biased such that individuals judged a statement to originate from the expected, stereotype-consistent source more frequently than simply by random chance if the origin could not be remembered—regardless that sources conformed with their age expectations only in half of the instances. Thus, individuals reconstruct the sources based on what they know about and intuitively associate with the sources acquired prior to the experiment (e.g., Bayen et al., 2000; Hicks & Cockman, 2003; Kuhlmann et al., 2016; Mather et al., 1999; Sherman & Bessenoff, 1999).

But which source-attribution cue dominates source guessing if the predictions for the source-guessing bias differ because of a mismatch in peri-experimental (contingency) knowledge and pre-experimental (stereotypes and schemas) knowledge? For instance, how do individuals attribute information to its origin in situations where information is associated with the unexpected source for most of the time and, by this, contradicting what people typically associate? The distinction between these two different kinds of cues individuals can make use of when inferring the source has been formalized in the theoretical prediction of the *probability-matching account* (Spaniol & Bayen, 2002). This account states that source guessing should be based on peri-experimental knowledge in the first place. Whenever a situation-specific contingency of item types and sources can be perceived during encoding, source guessing should be adjusted to and thus reflect this likelihood. Prior, pre-experimental, knowledge should impact source guessing mostly when knowledge about the item-source contingency is missing or cannot be perceived. The probability-matching account found empirical support in a multitude of studies since its release (e.g., Arnold et al., 2013; Bayen & Kuhlmann, 2011; Bell et al., 2020; Kuhlmann et al., 2012).

So, what experimental factors influence the supersession of pre-experimental knowledge by peri-experimental knowledge in source guessing? The following factors have been identified in the literature (Arnold et al., 2013; Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012; Spaniol & Bayen, 2002): one should (1) foster the accurate encoding of item-source pairings by providing the differentiating, individuating source information during learning (e.g., which specific person is young or old), and (2) raise awareness by intentionally instructing participants to remember the source (in addition to the items). Besides these, the source-memory performance (Kuhlmann et al., 2016) and the subjective judgments of item-source contingencies during learning (Arnold et al., 2013; Klauer & Meiser, 2000) have been shown to serve as additional experimental boundary conditions that modulate knowledge reliance in source guessing (on an individual level).

In summary, the aim of this thesis is to gain systematic insights into the nature of knowledge reliance in source guessing. Before turning to the research conducted in the four manuscripts, the basics of multinomial modeling are reviewed in the following chapter as it builds the statistical foundation throughout all manuscripts.

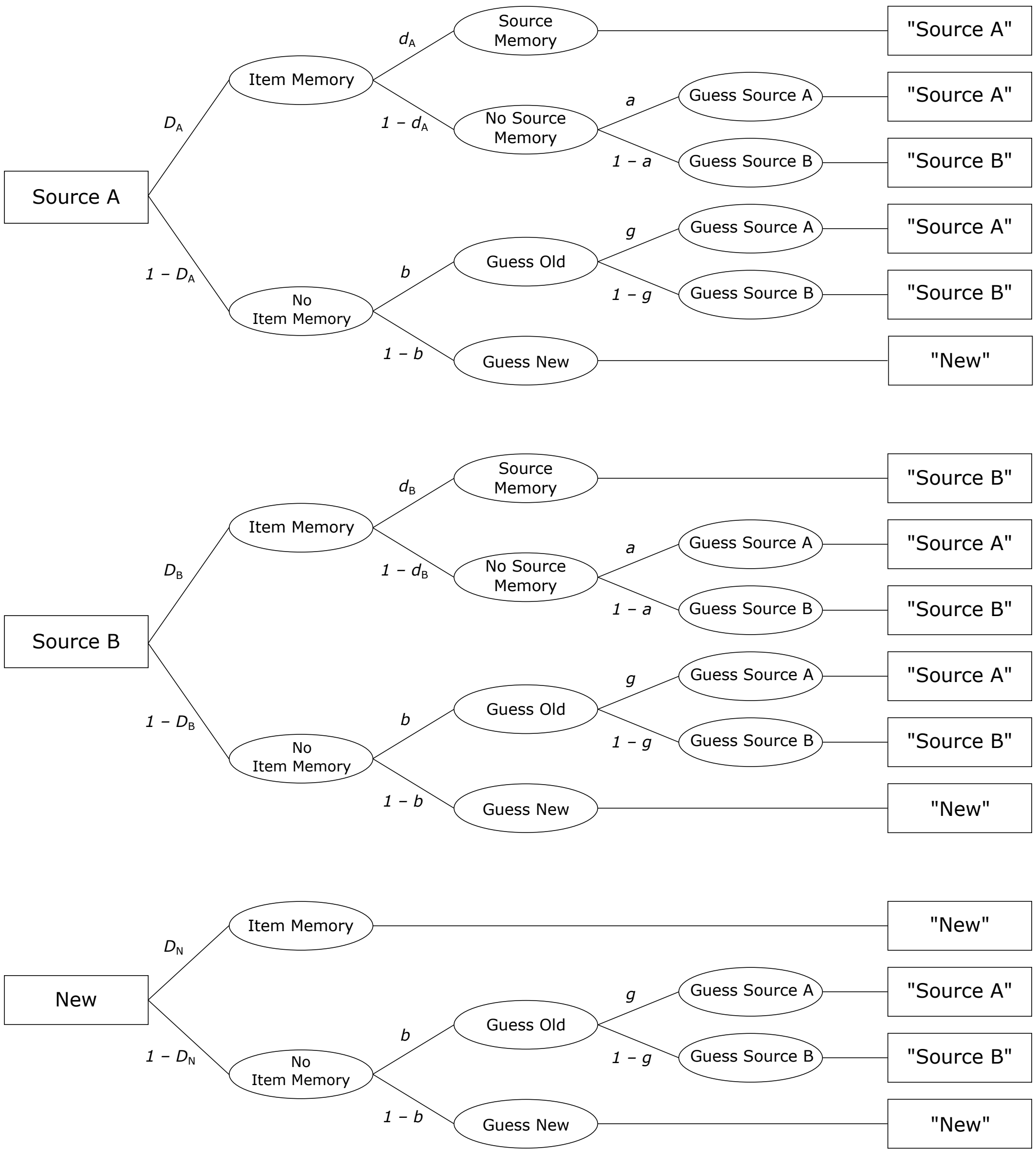

FIGURE 1: 2HTSM adapted from Bayen et al. (1996). Parameter D = probability to remember the item, parameter d = probability to remember the source, parameter b = probability to guess the old-new status of the item, parameter a and g = probability to guess the source. Subscripts of the respective tree refer to the different item types used in source-monitoring tasks—originating from source A, B, or neither source as the item is new.

3 Statistical Approach: Cognitive Modeling of Source-Monitoring Processes

> [...] estimating hypothetical parameters that represent the probabilities of unobservable events.
>
> *(Riefer & Batchelder, 1988)*

3.1 Multinomial Processing Tree (MPT) Models

Multinomial processing tree (MPT) models are widely-used instances of measurement models that make latent cognitive processes underlying behavior explicit. MPT models thereby assume that the categorical responses follow a multinomial distribution; the expected probabilities for each response category are modeled by a processing tree (Erdfelder et al., 2009; Riefer & Batchelder, 1988).

As it happens, around the same time as M. K. Johnson et al. (1993) proposed the theoretical framework, William H. Bachelder and his colleagues developed the idea of making cognitive processes, one of which source monitoring is, mathematically measurable (e.g., W. H. Batchelder et al., 1994; W. H. Batchelder & Riefer, 1990; Riefer & Batchelder, 1988; Riefer et al., 1994). The authors decisively drove the development of formalized stochastical models to account for the confounding of memory and judgment processes in categorical response data (as collected in source-monitoring tasks). A multinomial model for source-monitoring data was then provided by Bayen et al. (1996). The so-called *two-high-threshold multinomial model of source monitoring* (2THSM; Bayen et al., 1996) explicitly assumes that a threshold needs to be passed in order to detect a presented item in source-monitoring tasks as either old or new; otherwise, if the threshold is not passed, decisions are made in a state of uncertainty. In addition, a threshold for detected-old items can only be passed by old items and a threshold for detected-new items can only be passed by new items (cf. Bayen et al., 1996).

Due to its relevance for the overarching research question (variants and extensions of this model are applied to all experiments that were conducted within the scope of this thesis), a closer inspection of the model is worthwhile. Figure 1 illustrates the general structure: Items can either originate from one of two sources or are new. Four distinct latent cognitive processes underlying the attributions in a source-monitoring task shape the response behavior. Memory processes can encompass either item memory (parameter D) or source memory (parameter d), the same applies to judgment processes that can be influenced either by item guessing (parameter b) or source guessing (parameter a and g).

Researchers in the area of cognitive and social psychology have used this particular model to answer substantive research questions with regard to memory and judgment processes in source monitoring and have thereby had the benefits of its mathematical tractability and profound statistical properties (Calanchini et al., 2018; Erdfelder et al., 2009; Hütter & Klauer, 2016). For the 2HTSM, model fit can easily be quantified by goodness-of-fit tests and its (memory and guessing) parameters have been extensively validated (Bayen et al., 1996), and adapted and applied to multiple paradigms and various research questions since then.

The application of MPT models in psychological science has become even more powerful since simple hands-on software programs have become available (e.g., Moshagen, 2010; Singmann & Kellen, 2013) implementing the standard parameter estimation algorithm relying on the maximum likelihood (Hu & Batchelder, 1994). Recently, MPT estimation has been optimized for extended application purposes (e.g., to model individual or item differences, or to test single parameters and incorporate parameter correlations; cf. Hartmann et al., in press; Heck, Erdfelder, et al., 2018; Klauer, 2006, 2010; Klauer & Kellen, 2018; Matzke et al., 2015; Stahl & Klauer, 2007; Wickelmaier & Zeileis, 2018).

3.2 Bayesian-Hierarchical Extension of MPT Models

In recent years, researchers have identified Bayesian approaches to model cognitive processes. Amongst others, Klauer (2010), Lee (2011, 2018), Lee and Wagenmakers (2013), Matzke et al. (2015), and Smith and Batchelder (2010) promoted the application of Bayesian statistics to model cognitive processes including those involved in source monitoring. Traditional maximum-likelihood approaches in MPT modeling (Hu & Batchelder, 1994) infer parameter estimation from aggregated categorical responses across participants and items. This is inevitable from a mathematical point of view as parameter estimation requires a substantial data base to be precise and powerful in terms of model fit. Nonetheless, inherent assumptions which coincide with this estimation algorithm can be regarded as critical (e.g., Klauer, 2006; Smith & Batchelder, 2010)—at least when observations for participants and items are heterogeneous and thus not identically and independently distributed (i.i.d.). The beauty of current developments of Bayesian-hierarchical estimation within the scope of MPT modeling is the explicit account for this heterogeneity of participants and items by the assumption of a hierarchical group-level parameter distribution with separate estimates per individual and/or items. Parameters are based on individual and group-level information ("partial pooling"), thereby ensuring robust estimation

(Rouder & Lu, 2005).

One of the most prominent representatives of such model classes which has been applied to source-monitoring data is the *latent-trait model* proposed by Klauer (2010). The latent-trait model accounts not only for the heterogeneity but also for correlations between parameters assuming that the vector of individual parameters (probit-transformed) follow a multivariate normal distribution (prior) with group mean μ and variance-covariance matrix Σ to be estimated from the data. A graphical illustration of the Bayesian hierarchical model structure is provided in Figure 2.

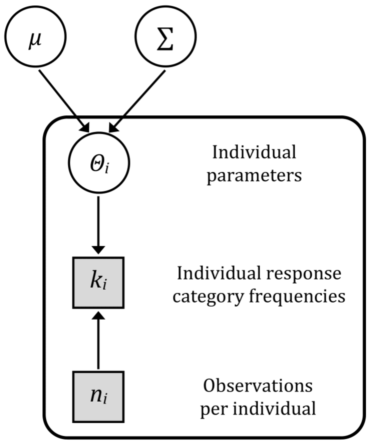

Figure 2: Conceptual illustration of Bayesian hierarchical model structure. μ = parameter group means, Σ = parameter variance-covariance matrix, θ = parameter vector for the i-th individual, k = categorical responses for the i-th individual, n = number of observations per i-th individual.

This modeling approach is certainly appealing for multiple reasons: Scientists who are interested in answering substantive research questions can benefit from it because (1) individual differences or rare populations with only small numbers of observations (such as patients or age groups), (2) parameter correlations, and (3) correlations of parameter and external variables (such as age, intelligence scores, or working-memory capacity) can be estimated. Furthermore, the latent-trait approach can be easily adapted and extended to various model specifications and is implemented in the user-friendly software program *TreeBUGS* (Heck, Arnold, et al., 2018). The statistical model that is applied throughout all manuscripts of this thesis is a latent-trait extension of the 2HTSM (Bayen et al., 1996). The memory and guessing parameters of the 2HTSM are estimated based on Bayesian inference on the posterior distribution of the parameters. That is, the prior assumptions about the parameter distributions are updated on the basis of the empirical evidence in the collected data. With the help of Markov-chain Monte Carlo (MCMC) methods samples from the posterior distribution are drawn resulting in interpretable statistics of parameter estimates

(Heck, Arnold, et al., 2018). In addition, *TreeBUGS* can be considered a powerful tool to judge the diagnosticity of the information provided by the data via parameter-recovery simulations. Researchers can simulate categorical response data for their experimental paradigm and quantify the extent to which the model is able to recover these values. Thereby, the replicability of the results found in a present sample can be assessed after the data collection (or prior to the experiment as an orientation for the planned sample size).

In general, for the sake of comparability across experiments, we follow the same step-by-step procedure for model-based analyses in each manuscript. The latent-trait model (Klauer, 2010) is set up and run in *TreeBUGS* (Heck, Arnold, et al., 2018) with the following specifications: three MCMC chains were sampled simultaneously with 20,000 iterations each, a thinning rate of 5, and an adaptation phase of 2,000 iterations (number of samples to adjust MCMC sampler in *JAGS*, Plummer, 2003). This basic model is run in loop until the pre-defined convergence criterion of $\hat{R} < 1.05$ (Gelman & Rubin, 1992) is reached. In addition, one additional sampling cycle is run to ensure convergence stability. The sampling procedure starts from scratch if the convergence criterion is not met within an upper bound of 10 resampling attempts. Model fit (i.e., whether the model and its imposed parameter restrictions) is assessed with Bayesian posterior predictive p-values for the mean (T_1 statistic; difference between observed and expected mean frequencies) and covariance structure (T_2; summed difference between observed and expected covariances; standardized by expected SD) of the categorical response data (cf. Klauer, 2010).

4 Rethinking Source Guessing: The Nature of Knowledge-Based Biases

What determines the amount of source bias that people show in source-monitoring tasks?

(Spaniol & Bayen, 2002)

The focus of the research question proposed in my thesis is to inspect the nature of knowledge reliance in source guessing to improve the understanding of reconstructive guessing biases in the source-monitoring literature. To answer this research question, the statistical tools applied to draw substantive conclusion can be best described as state of the art. By using these methods, source guessing can be examined in more depth and, at the same time, meet the increasing request to study individual differences and to combine cognitive process and their predictor variables within a joint theoretical and statistical framework. This thesis is comprised of four manuscripts that are outlined in the following sections. Note that the emphasis of the now following manuscripts is set on specific aspects of source guessing in the literature to derive the research question proposed in this thesis.

4.1 Cognitive Dynamics

Wulff, L., & Scharf, S. E. (2020). Unpacking stereotype influences on source-monitoring processes: What mouse tracking can tell us. *Journal of Experimental Social Psychology, 87*, 103917. https://doi.org/10.1016/j.jesp.2019.103917

In the first manuscript, we take an initial step in the direction of the underlying nature of knowledge reliance in source guessing by exploring the cognitive dynamics of stereotype influences on source monitoring. To do so, we conducted a source-monitoring task in which participants learned and later attributed information to its origin. While being tested for their source-monitoring performance, cursor movements were tracked—potentially revealing commitment towards or conflict between sources.

This first manuscript served as a starting point for answering the question which mechanisms may be at play and describe source-monitoring decisions more detailed. The dynamic evolution of cognition is hardly captured by static, categorical response data but can be mirrored in the cursor movements during decision scenarios. Hence, the process-tracing method of mouse tracking is well suited to reflect the ongoing decision processes in real time (cf. Hehman et al., 2015; Kieslich et al., 2019; Maldonado et al., 2019; Stillman et al., 2018; D. U. Wulff et al., 2019). Task-specific cognitive processes are mapped to motor responses reflecting the latter (Spivey & Dale, 2006) that can be quantified and interpreted as the activation of (competing) social information.

In a previous study, Freeman and Ambady (2009) tested participants' gender categorization via mouse tracking. Participants sorted faces into social categories. Precisely, they attributed female and male faces to expected-feminine (e.g., caring) and expected-masculine (e.g., aggressive) features. In typical trials, in which the facial appearance was congruent to the associated personality description, participants showed less curved mouse movements towards the competing response option reflecting the opposing gender than in atypical trials. In the latter, participants were nonetheless spatially attracted towards the stereotypical response option. Thus, Freeman and Ambady (2009)'s findings illustrated that pre-experimental knowledge tied to gender impacts the real-time dynamic of social categorization. Further evidence was provided by Koop and Criss (2016). The authors studied whether mouse tracking could mirror the sensitivity and response bias in recognition-memory processes (within the signal-detection theory framework). Participants learned words that they later had to remember as either old or new while their cursor movements were tracked during testing. And, in fact, the initial deviation towards one response option indicated the response bias: If items were mostly old and previously presented, participants adapted their response bias which was also reflected in the tendency towards old response option.

Mouse tracking has been successfully applied in various research domains, which, among other domains, include social cognition (e.g., Cassidy et al., 2017; Cloutier et al., 2014; K. L. Johnson et al., 2012; Wojnowicz et al., 2009) and memory (e.g., Abney et al., 2015; Papesh & Goldinger, 2012). While acknowledging the importance of both research domains for this thesis, we also note that, to the best of our knowledge, no study has focused on source monitoring and mouse tracking until now. Although highly related to simple recognition memory, source monitoring may encompass more complex cognitive processes that go beyond the old-new classification and that should merit attention and are worth examining with the recently developed process-tracing measures.

In this first manuscript, we therefore inspected whether decisional uncertainty between sources during source attributions is reflected in the dynamics of cursor movements in a joint statistical model with source-monitoring processes and process-tracing indices as their predictors.

Participants performed a source-monitoring task, in which they learned everyday statements as belonging either to a young or an old person (without revealing the specific assignment of the ages to the sources during encoding). The statements were normed to be expected for younger ("I rarely get sick.") or older adults ("I go to church every Sunday.") and both item types were paired with each of the sources equally often (inducing a null contingency). In a later sequential memory test, participants first classified each item among new distractor items as old or new and then assigned the items previously classified as old to their origin while their cursor movements were continuously tracked during the source-attribution decision as implemented with the *mousetrap* plug-in (Kieslich et al., 2018) in the open-source experiment builder *OpenSesame* (Mathôt et al., 2012).

The inspection of source attributions and the influence of stereotypes on such via multiple process-tracing indices (e.g., maximum deviation of trajectories, distributional analysis of prototypical trajectories, reaction times) underpinned previous observations: the stereotypical and -atypical features reflected in both the item content and source label during source attributions prompted participants to consider both sources during the course of their decision as indicated by different-curved trajectories based on the typicality of the to-be-judged item and the responses given (see Figure 3). The deviations of the cursor trajectories from the ideal path to the selected source (as measured by the *Maximum Absolute Deviation*, MAD) revealed that, indeed, the non-selected source was (at least to some extent) still taken into account when inferring the item origin based on stereotypical features (in line with e.g., Cassidy et al., 2017; Freeman & Ambady, 2009; Freeman et al., 2010).

In addition, modeling source-monitoring processes and process-tracing indices in a joint statistical framework displayed an interesting pattern: For source decisions, mouse movements were substantially related to the retrieval of sources based on contextual knowledge they had previously encoded, $\hat{\rho} = -.36\ [-.58, -.09]$ (correlation of MAD and model based source-memory parameter d; BCI adjusted for sampling error). That is, decisional uncertainty between sources, measured as the maximum deviation from the idealized line towards the selected response option, was less pronounced in case participants had actual memory for the source and, therefore, experienced less cognitive conflict in their choice. Mouse movements, however, were not reliably indicative of source guessing. Uncertainty inherent to the source attribution which is based on guessing due to lacking source memory was not mapped onto mouse movements, $\hat{\rho} = .21\ [-.10, .46]$ (correlation of MAD and model based source-guessing parameter g; BCI adjusted for sampling error). In sum, cognitive conflict did not vary as a function of the strength of knowledge reliance (i.e., stereotype application) in source guessing but did so with regard to participants' source-memory performance.

Beyond that, bimodality assessment of trajectories and the distribution of prototypical

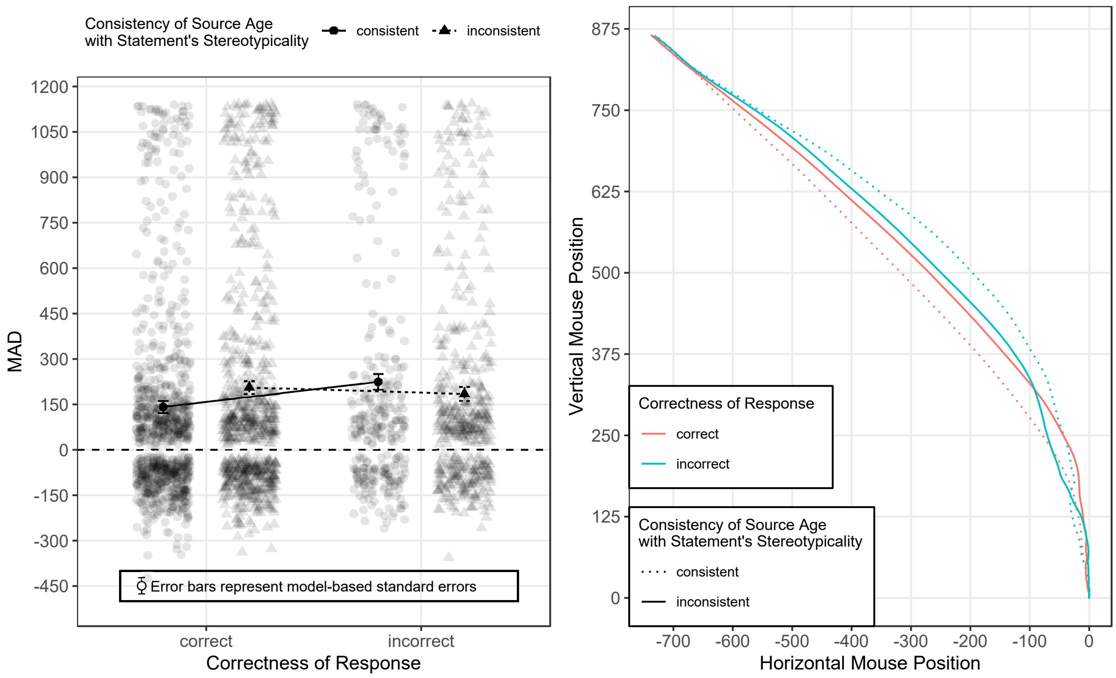

Figure 3: Manuscript I – Mouse-tracking index *Maximum Absolute Deviation* (MAD) as a function of the correctness of the source attribution and consistency of the age reflected in the item and the source's age aggregated across items/participants. Left panel: Positive values of MAD indicate spatial attraction towards the non-selected source, 0 (dashed line) indicates the idealized straight line towards the selected source, negative values indicate spatial attraction towards the selected source. Corresponding error bars represent standard errors (displayed in black), individual values per participant are displayed in grey (points = age-consistent item-source combinations, triangle = age-inconsistent item-source combinations). Right panel: Temporal dynamic of mouse trajectories based on the horizontal and vertical cursor position (in pixels) for correct (orange) and incorrect (blue) source attributions separate for age-consistent item-source combinations (dashed line) and age-inconsistent item-source combinations (solid line).

trajectory shapes highlighted another striking feature of the present data. Trajectories could be classified in two main prototypes: curved and discrete change-of-mind trajectories. So, rather smooth or abrupt shifts in the spatial attraction shaped cursor movements predominantly. This dual trajectory-type pattern may tentatively be interpreted as an indicator of underlying dual processes (Freeman et al., 2008; Freeman & Dale, 2013) that reflects consistency in the initial and later response tendency for straight trajectories and a correction of the initial categorization for change-of-mind trajectories. Of course, the derivation of underlying distinct process classes based on trajectory distributions needs to be treated with caution. Nonetheless, our results provide a comprehensive picture of stereotype influences on source monitoring in general and serve as a first indication of the underlying nature. Certainly, source memory is singled out as a crucial precondition for unbiased source guessing through pre-experimental knowledge in the form of stereotypes.

This first manuscript can be considered as an essential first step to unpack the underlying nature of knowledge reliance in source guessing and its findings provided important diagnostic information towards a more systematic and targeted examination of source guessing in isolation. We demonstrated that mouse tracking is a promising avenue to follow up on in the future if combined with recent advances in cognitive modeling (e.g., Heck, Erdfelder, et al., 2018) and its application should be considered to augment traditional analyses in a stimulating manner if the psychological mechanisms behind cognitive processes are of interest.

4.2 State versus Trait

Wulff, L., & Kuhlmann, B. G. (2020). Is knowledge reliance in source guessing a cognitive trait? Examining stability across time and domain. *Memory & Cognition*, *48*(2), 256–276. https://doi.org/10.3758/s13421-019-01008-1

In the second manuscript of this thesis, we test whether knowledge reliance in source guessing could be characterized as a *cognitive trait* (cf. Kantner & Lindsay, 2012, 2014)—indicated by parameter stability across time and knowledge domain. Derived from the literature on variables in judgment and decision making of trait-like nature (Glöckner & Pachur, 2012; Odum, 2011; Scheibehenne & Pachur, 2015), stability on decision strategies (e.g., old-new response bias) across temporal and situational contexts may be hypothesized for source guessing as well due to the its similarity to other processes for which stability has been reported (Kantner & Lindsay, 2012, 2014; Michalkiewicz & Erdfelder, 2016). Thus, a considerable body of research has focused on response biases and individual differences in the criterion sensitivity. As source guessing has been referred to as a "response bias" in the literature as well (cf. W. H. Batchelder & Batchelder, 2008), the question arises to what extend knowledge reliance in itself can be described as trait-like stable which has not undergone a systematic examination yet.

In two experiments, participants therefore performed two study-test cycles of a source-monitoring task—either separated by a certain time interval, to test stability across time (Experiment 1), or varied in the knowledge domain depicted in the stimulus material, to test stability across knowledge domains (Experiment 2). Both source-monitoring tasks were comparable with regard to their general procedure.

In Experiment 1, participants performed both study-test cycles processing item and source information based on the knowledge domain of age stereotypes. Statements (= items) were highly comparable in terms of their expectancy ratings for either young or old adults (Kuhlmann et al., 2017). We manipulated between-subjects whether both tasks needed to be performed in the same session (separated by 10 minutes) or in single sessions held seven days apart. Using this manipulation, we were able to test the temporal stability of knowledge reliance in source guessing across a short- and rather long-term time period.

In Experiment 2, participants performed the first study-test cycle processing item and source information based on the knowledge domain of age stereotypes and the second one based on gender stereotypes (separated by 10 minutes). Albeit slightly different in the overall expectancy ratings for age groups (Kuhlmann et al., 2017) and gender groups (statements normed in Experiment 1), both study-text cycles were nonetheless highly comparable in terms of their stimulus material. Using this manipulation, we were able to test the content-independent stability of knowledge reliance in source guessing across knowledge domains.

In addition, personality traits measured with the Big-Five inventory (Danner et al., 2016; Rammstedt et al., 2020; Soto & John, 2017), cognitive processing styles measured with the Rational-Experiential Inventory (Epstein et al., 1996; Keller et al., 2000), and the subjective contingency judgments assessed at the end of the experiments were incorporated as external variables in each experiment's respective model-based analysis. We thereby tested whether knowledge reliance in source guessing can be (partially) explained by participants' character facets.

The data were analyzed in a joint MPT model for both tasks with respective parameters for each study-test cycle (i.e., model equations were duplicated) to allow for parameter correlations of which the correlation of both source-guessing parameters (i.e., one for each study-test cycle) was of most interest to test our research question. Apart from the correlation-based definition of stability (which largely hinges on the variability in source guessing in the total sample), a more rigid definition of stability could, furthermore, imply consistency in source guessing in absolute terms (i.e., whether source guessing varied within participants across tasks—irrespective of the overall relative group-level tendencies as covered by parameter correlations).

The relative stability of knowledge reliance in source guessing for both experiments is illustrated graphically in the upper panels of Figure 4, the absolute stability is displayed in the lower panels. Correlation-based stability was not revealed by the data of both cross-task time conditions of Experiment 1 ($\hat{\rho} < .27$; BCI included 0), the absolute difference measure, however, suggested cross-task stability (at least to some extent). A closer inspection of participants' source-guessing biases demonstrated that the majority of individuals relied on their pre-experimental knowledge about ages in source guessing that was somewhat comparable across time (10 minutes and seven days). In Experiment 2, a cross-task correlation of credible but small size was obtained. This result indicated that knowledge reliance on age stereotypes in source guessing in the first task predicted the strength of

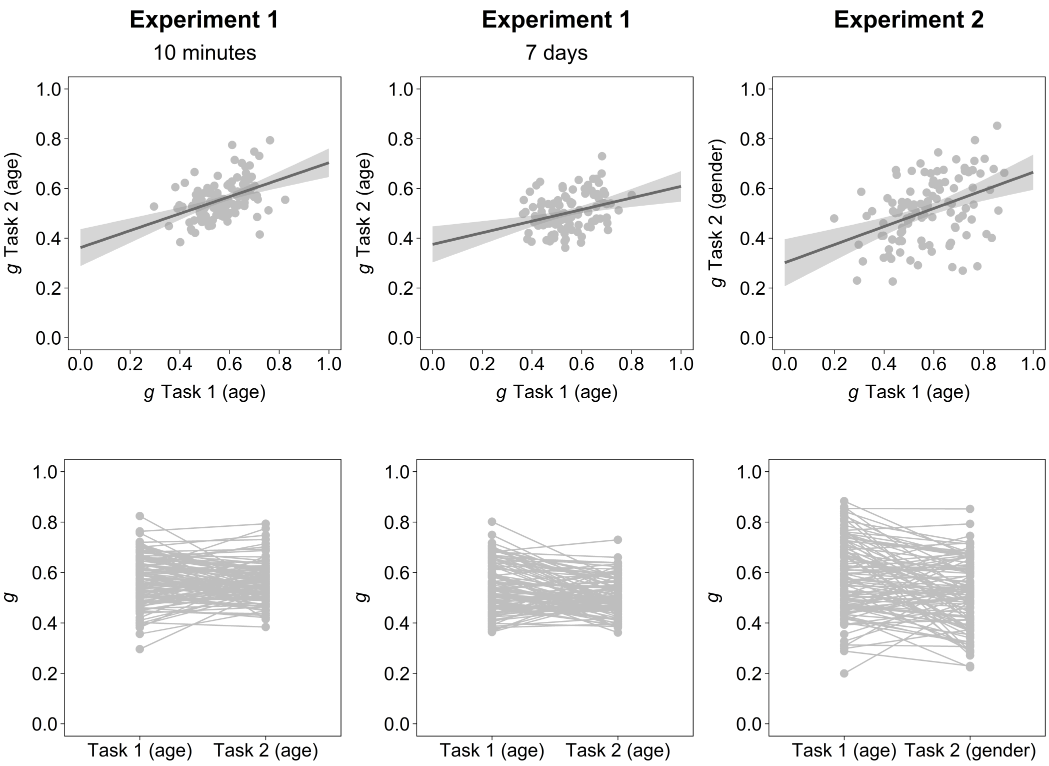

Figure 4: Manuscript II – Stability in source guessing (parameter g) across two study-test cycles in two experiments. Upper panel: relative measure of stability based on parameter correlation; lower panel: absolute measure of stability based on parameter differences. Grey points represent individual parameter estimates, linear trend line with error bars represent standard errors.

knowledge reliance on gender stereotypes in the second task ($\hat{\rho}$ = .30 [.07, .53]). The absolute difference supplemented the results for the relative stability: Most participants showed a fairly constant level of knowledge reliance on both age and gender stereotypes in source guessing across tasks.

Both experiments uncovered, if at all, weak positive associations of knowledge reliance across two source-monitoring tasks when measured with a correlation-based approach. The results added more evidence in terms of stability using an absolute measure of constant guessing behavior—regardless of the time between study-test cycles or the reflected knowledge domain. But still, a substantial portion of the interindividual variance in knowledge reliance in source guessing could not be explained by stable judgment processes across time and knowledge domain. Other processes involved in source monitoring, most of which was old-new item recognition (parameter b), indeed reflected stronger trait-like stability. What is more, none of the personality traits or cognitive processing styles were related to source guessing in a systematic manner in both experiments; the subjective contingency judgment partially did so (as indicated by positive correlations with source guessing). Whereas source guessing clearly encompasses trait-like features, it could not be described as fully trait-like as other closely related cognitive processes (Kantner & Lindsay, 2012, 2014; Michalkiewicz & Erdfelder, 2016). It rather appears to be an adaptive tool used in specific situations and for which knowledge reliance varies (to a certain extent) within and between individuals.

4.3 Automatic versus Controlled

Wulff, L., & Kuhlmann, B. G. (2020). *Is knowledge reliance in source guessing automatic or controlled? Evidence from divided attention and aging*. Manuscript submitted for publication.

In the third manuscript of this thesis, we consider whether knowledge reliance in source guessing depends on cognitive resources (as an indicator of an automatic versus controlled cognitive process; following Bargh, 1994) during source attributions. For the preregistered experiment as described below, both types of knowledge reliance are studied more closely: peri-experimental contingency knowledge and pre-experimental stereotype knowledge. The automatic (i.e., resource-independent) or controlled (i.e., resource-dependent) nature of both source cues is further examined within younger and older adults thereby obtaining a comprehensive view of knowledge reliance in source guessing from multiple perspectives.

The question to what extent (pre-experimental) knowledge reliance in source guessing is rather automatic or controlled has been raised and empirically tested in the past (Bayen & Kuhlmann, 2011; Bröder et al., 2007; Ehrenberg & Klauer, 2005; Marsh et al., 2006; Sherman et al., 2003; Spaniol & Bayen, 2002). The majority of the few existing studies on the pre-experimental knowledge reliance have characterized its nature as rather controlled (e.g., resource-dependent and conscious strategy) but evidence from Marsh et al. (2006) and Klauer and Ehrenberg (2005) suggest that an automatic reliance may be also likely. Empirical evidence on the nature of stereotypes (see, for instance, Bodenhausen et al., 1999; Hamilton & Sherman, 1994), and age stereotypes in particular (Gonsalkorale et al., 2014; Lepore & Brown, 1997; Perdue & Gurtman, 1990), underpins this assumption in addition. Furthermore, the nature of contingency influences on source guessing has not yet been well examined and the question whether the stereotype counteraction through contingency knowledge is resource dependent thus remained unanswered in the past.

For these reasons, the nature of both forms of knowledge reliance as source cues were studied. We did so by experimentally inducing cognitive load at retrieval and testing naturally-emerging load due to cognitive declines in older age. Importantly, none of the aforementioned studies examined the nature of knowledge reliance based on (age) stereotypes in both younger and older adults.

In a source-monitoring task, participants of both age groups learned age-stereotypical statements from a young and an old person (as used in Manuscript I and II; see respective sections above). Half of the younger-participants sample (18−26 years) performed a secondary, parallel (tone-monitoring) task during source attributions, the other half did not. Crossed with this manipulation, the specific source information (ages) was provided either before encoding or before testing (between-subjects) as information during encoding facilitates contingency detection (e.g., Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012). This encoding manipulation was also applied to the older-adults sample (61−80 years).

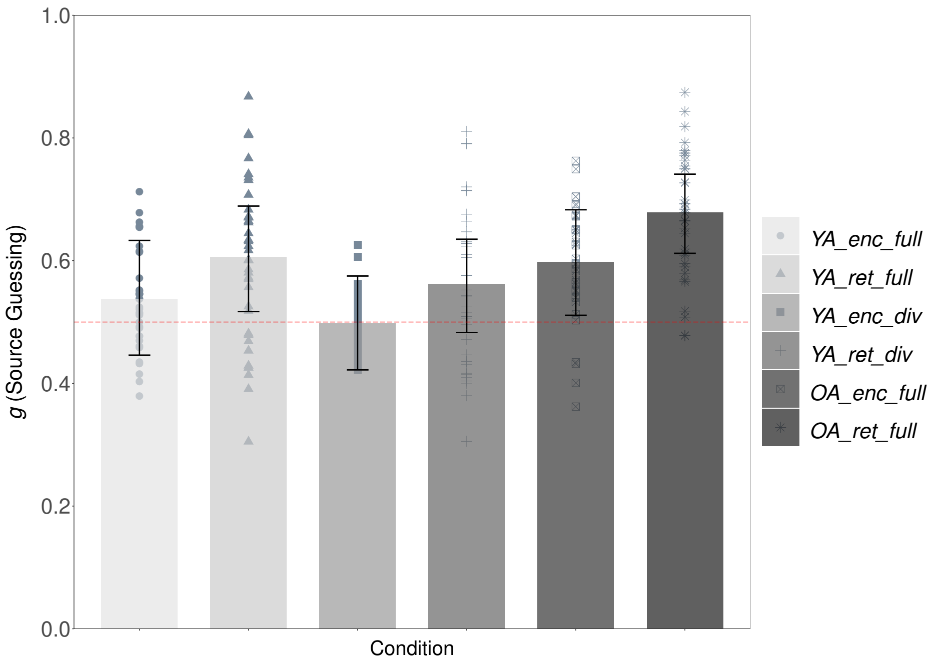

Figure 5: Manuscript III – Estimates of source guessing (parameter g) for the following experimental conditions: *YA* (younger adults) and *OA* (older adults) refer to the participants' age group, *enc* (encoding) and *ret* (retrieval) refer to the encoding-condition manipulation (i.e., whether the specific source information was revealed before encoding or at test), and *full* and *div* refer to the dual-task manipulation at test (i.e., whether participants performed a parallel tone-monitoring task or not; only manipulated in younger adults). Red dashed line = chance-level, unbiased source guessing, parameters $>$ .50 indicate stereotype-based source guessing, parameters $<$ .50 indicate counter-stereotype-based source guessing. Group-level parameter means are displayed as bars, individual parameter estimates are displayed as dots. Corresponding error bars represent the 95%-BCI.

Figure 5 illustrates the model-based results for both age groups. For younger adults, the probability-matching account (Spaniol & Bayen, 2002) was replicated: In case the accurate perception of item-source contingencies was prevented (i.e., specific ages of the sources only provided at test), their source guessing was based on stereotype knowledge whereas facilitating the detection of the contingency (specific ages of the sources already provided at encoding) led to a reliance on contingency knowledge in source guessing. A division of attention based on the dual-task implementation during the source-monitoring test neither affected the strength of contingency-based source guessing (resulting from specific source information at encoding) nor stereotype-based source guessing (resulting from specific source information only at test). The use of a dual-task manipulation at retrieval thereby helped to unpack the underlying automatic nature.

Whenever participants were able to perceive the item-source contingency in an accurate manner (i.e., source ages presented already at encoding), they based their source guessing on this contingency knowledge. Otherwise, they drew on their stereotype knowledge when guessing the source of an item. Importantly, they applied both knowledge cues independent of the cognitive resources available at test, implying the automatic nature of the judgment process. Older adults, however, strongly relied on stereotype knowledge in source guessing and were less successful in counteracting stereotype influences through contingency knowledge than younger adults. Hence, older adults did not apply contingency knowledge in a fully automatic manner to their guessing behavior when they actually could have relied on the available source information at encoding.

With this third manuscript, we stressed the independence of knowledge reliance in source guessing from cognitive capacities at the time of retrieval—at least in younger adults who could make use of the item-source contingency perceived during encoding to "unbias" their source guessing. By this, we demonstrated that stereotype reliance can be overcome by enhancing attention to actual behaviors and attitudes of different-aged target persons (at least in younger adults). Older adults whose general reliance on stereotype knowledge was already more pronounced than for younger adults were not able to decrease their guessing bias to accurate chance level based on the perceived item-source contingency. The present findings reinforce the value of unaffected encoding of information and context for source guessing to be reliant on individualized knowledge acquired in the moment of encoding and not on generalized knowledge acquired prior to the experiment that, in turn, potentially misleads the source-attribution process.

4.4 Generalization to Novel Stimuli

Wulff, L., Bell, R., Mieth, L., & Kuhlmann, B. G. (2020). *Guess what?! Different source-guessing mechanisms for old versus new information.* Manuscript in preparation.

In the fourth and last manuscript of this thesis, we investigate whether the item-source contingency encountered during the course of encoding can be transferred to source decisions for new information that has not been processed and, thus, accompanied by any source in the course of the experiment.

Albeit Spaniol and Bayen (2002) proposed a theoretical account on the use of different cues for source guessing—that was confirmed by empirical evidence, for instance from Bayen and Kuhlmann (2011) and Kuhlmann et al. (2012)—up until recently, no assumptions were formulated and tested about source guessing for new information for which a source needs to be inferred. Bell et al. (2020) recently proposed a theoretical extension of the probability-matching account (Spaniol & Bayen, 2002) for source guessing for detected-new items. In their studies, Bell et al. (2020) presented face images with profession labels of a farmer or a lawyer (as items) that were associated with expected-farmer, expected-lawyer, or unexpected behavioral statements for either profession (as sources). The authors adapted the subsequent memory test so that participants indicated the item status first ("old" versus "new") followed by a source attribution for both old *and* new items—irrespective of their decision about the item status. Their data modeled with an adapted three-sources model variant of the 2HTSM (Bayen et al., 1996; Keefe et al., 2002) with separate source-guessing parameters for detected-old and detected-new items (Bell et al., 2020) showed that participants considered the item-recognition status in their source-guessing behavior. That is, they applied contingency knowledge only to those items that they detected as old (i.e., probability matching) but relied on the schematic knowledge about the professions as a default guessing strategy for detected-new items.

What remained unanswered is whether the pre-experimental knowledge was applied to novel stimuli, because forced to make a source attribution for these items, as individuals could not build an appropriate perception of the item-source contingency due to the large number of unique pairings of faces and behaviors without specific and constant individuating source information during encoding that could, in turn, be transferred and applied to new items.

To rule out this explanation, we conducted a source-monitoring task using the well-established doctor-lawyer paradigm (Bayen et al., 2000) in which participants learned and later attributed expected-doctor and expected-lawyer information to either of three constant sources (the doctor, the lawyer, and an unnamed third person). By presenting information from constant and known sources, we aimed to test the generalization of the probability-matching account to novel decision contexts with distinct sources—given that we usually interact with recurring, specific (and oftentimes known) individuals in real life. That is, is knowledge reliance in source guessing based on peri- experimental contingency knowledge or pre-experimental stereotype knowledge in source attributions for novel, context-independent, information? In a source-monitoring task, participants learned everyday statements as belonging to either of three sources: the doctor (= physician), the lawyer, or a third unknown person who was not described in any further detail. The doctor and lawyer were accompanied by black-and-white images showing two similarly-looking adult men (see Bayen et al., 1996), a name, and profession label; the third unknown source was presented without an image, a name, and profession label. The statements were normed to be expected either for the profession of a doctor ("Your blood pressure is too high.") or a lawyer ("I have to be in court at 9am.") in a survey study by Kuhlmann et al. (2012). Both item types were paired with each of the three sources equally often (inducing a null contingency). In a later memory test, participants attributed each item among new distractor items to either of the three sources. One feature of the source-attribution test is of particular importance as it differed from the paradigms used throughout the first three manuscripts: Whenever an item was classified as new, the procedure forced participants nonetheless to attribute the novel information (either expected-doctor or expected-lawyer items) to either source. Even though participants could not draw on contextual source memory for these instances (as the information has never been presented with any source before), they should make their decision based on a simple guess for the source from which they nonetheless thought that the statement would originate.

In order to infer the cognitive processes underlying participants' categorical responses for old and new items in the source-monitoring task, we also analyzed the data with the extended three-sources model variant of the 2HTSM (Bayen et al., 1996; Keefe et al., 2002) by Bell et al. (2020). This model extension explicitly assumes separate source-guessing parameters for old and new items (detected or guessed as such), therefore, enabled us to estimate source guessing for both item-recognition states in particular.

As can be seen in Figure 6, the status of the item recognition was incorporated into source guessing: For those items that were detected as old (parameter a_{E}), participants based their guessing behavior on the item-source contingency reflected in the stimulus material (i.e., probability matching). For those items that were detected as new (e_{E}), participants based their guessing behavior on pre-experimental knowledge about the professions of a doctor and a lawyer. Thus, the learning history of the equal presentation of both item types by specific source exemplars could not be used to generalize its application to decision contexts for unstudied information. Instead, the profession schema mostly biased source guessing for detected-new items. This finding was also observed for items that were not remembered and their old-new recognition status therefore initially needed to be guessed (parameters g_{OE} and g_{NE}).

The last manuscript offers new opportunities to rethink and question the external validity of knowledge reliance in source guessing which is the generalization of the probability-matching account to new information when a source attribution is required. Replicating recent findings from Bell et al. (2020), this study demonstrated that generic, pre-

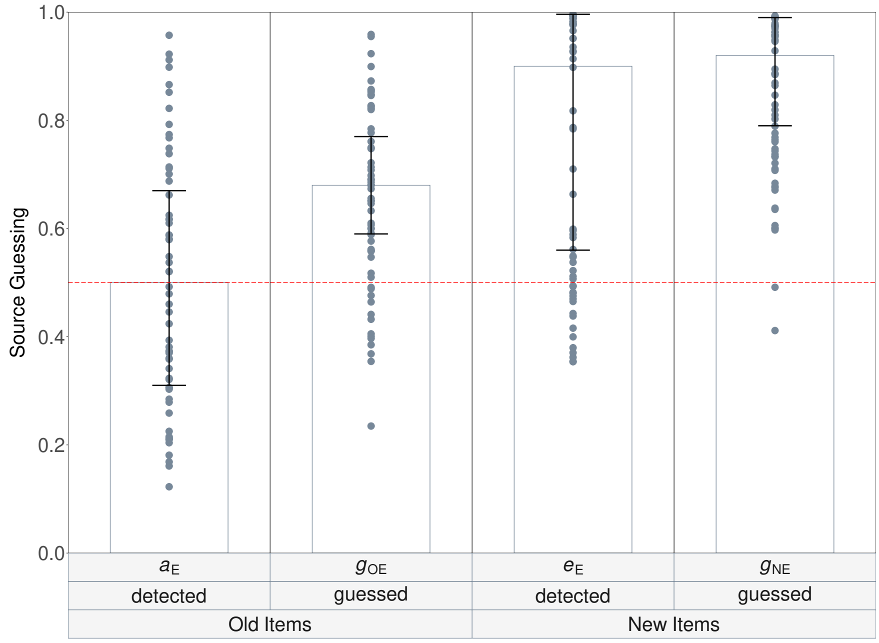

Figure 6: Manuscript IV – Source-guessing parameters of the source monitoring MPT model for detected-old and detected-new items (adapted from Bell et al., 2020). Estimates represent source guessing for the expected (E) source depending on the recognition status (old vs. new) of items that were either detected or guessed. a_{E} = probability of guessing the expected source given that the item was recognized in the first place; $g_{\mathrm{O\,E}}$ = probability of guessing the expected source given that the item was guessed to be old; e_{E} = probability of guessing the expected source given that the item was detected to be new; $g_{\mathrm{N\,E}}$ = probability of guessing the expected source given that the item was guessed to be new. Red dashed line = chance-level, unbiased source guessing, parameters $>$ .50 indicate schema-based source guessing, parameters $<$ .50 indicate counter-schema-based source guessing. Group-level parameter means are displayed as bars, individual parameter estimates are displayed as dots. The error bars represent the 95%-BCI.

experimental knowledge predominantly biased source guessing in situations where participants were not able to use previously-encountered contextual cues for their source attributions. The contingency knowledge was not even transferred to source decisions for novel information from known source exemplars for which participants have built an association between item types and the sources' characteristic that, in principle, could have provided the precondition to rely on contingency knowledge in source guessing also for new decision contexts. This last manuscript thereby disclosed a restriction of the scope of application for the probability-matching account to old items.

5 General Discussion

> [...] Knowledge structures [...] provide invaluable backup to episodic memory.
>
> *(Sherman & Bessenoff, 1999)*

In this dissertation thesis, I endeavor to address the nature of knowledge reliance in source guessing in a systematic way within a joint overarching theoretical framework. By means of multifaceted methodological approaches and recent statistical advances applied throughout the manuscripts, this thesis contributes to creating a holistic picture of the specificities of knowledge reliance in source guessing and broadens the understanding of its underlying mechanisms. Whereas judgment processes in source monitoring have often been modeled as, more or less, nuisance parameters with limited informational content to substantive research questions about source-monitoring processes (cf. W. H. Batchelder & Batchelder, 2008), we demonstrate the high degree of adaptivity of knowledge reliance across a wide range of application fields and influencing factors—thereby further qualifying guessing as “important inferential processes in reconstructive memory decisions” (Meiser et al., 2007, p. 1037). Moreover, the use of state of the art analytical tools, unmask notable interindividual and intraindividual differences in knowledge reliance in source guessing and their potential predictors.

In Manuscript I (L. Wulff & Scharf, 2020), we explore the influence of pre-experimental knowledge in forms of age stereotypes on source attributions with the increasingly popular process-tracing method of mouse tracking. We test whether cursor trajectories mirror the decisional uncertainty between sources due to cognitive conflict arising from the opposing social categories of "prototypically young" and "prototypically old". The research conducted in Manuscript I confirmed that the strength of decisional uncertainty is weakened due to individuals’ memory performance but can be less described as a function of individuals’ knowledge reliance in source-guessing behavior.

In Manuscript II (L. Wulff & Kuhlmann, 2020a), we examine whether source guessing can count as a trait-like response bias and, by this, implies stability in the extent to which individuals rely on knowledge structures in case source memory fails. We quantify the stability of participants’ knowledge reliance in source guessing across two study-test cycles of source-monitoring tasks varying the time lag or knowledge domain in between. Further, derived from the literature on trait-like variables in judgment and decision making, we take into account that source guessing can be potentially explained by personality traits and cognitive processing styles. The research conducted in Manuscript II confirmed that source guessing encompasses somewhat stable features but can by far not be characterized as trait-like to the same extent as other variables (e.g., old-new response bias) have been in the past.

In Manuscript III (L. Wulff & Kuhlmann, 2020b), we broaden the scope of source cues which individuals can presumably make use of from pre-experimental knowledge to peri-experimental knowledge (item-source contingency perceived during encoding). We look into the underlying automatic or controlled nature of both pre- and peri-experimental knowledge reliance in source guessing with regard to its resource dependence. We do so by implementing a dual task during source attributions at retrieval and, in addition, focusing on aging as a natural "dual-task" condition. Further, we assess whether participants benefit from specific source information provided already at encoding to counteract influences of knowledge acquired prior to the experiment on source guessing. The research conducted in Manuscript III confirmed that the application of knowledge in the situation of attributing information to its origin is mostly driven by automatic, resource-independent, cognitive operations. This is true for both contingency-based and stereotype-based source cues. Contingency knowledge can counteract stereotype knowledge in younger adults if preconditions are met (specific source information given at encoding) to accurately perceive the item-source contingency. This finding, however, is not entirely applicable to older adults whose source guessing does not benefit from the contingency perception to the same extent as source guessing for younger adults does.

In Manuscript IV (L. Wulff et al., 2020), we test source guessing for new information for which no contextual cues are available emerging from the item-source contingency in the learning history of the experiment's encoding phase. By setting out the conditions to build an accurate item-source contingency during encoding processing various information from constant and specific source exemplars, we investigate the generalization of the probability-matching account (Spaniol & Bayen, 2002) to novel stimuli. That is, we study whether participants base their source guessing on contingency rather than stereotype knowledge when a source attribution is required also for new information of already known sources. The research conducted in Manuscript IV confirmed that participants consider the recognition status of the item in their source-guessing behavior: They guess mostly based on contingency knowledge for detected-old items but draw on stereotype knowledge for detected-new items revealing that they do not abstract and transfer the contingency knowledge associated with known sources and acquired during the course of the experiment to novel decision contexts.

In summary, this dissertation thesis contributes a crucial piece in the puzzle of how knowledge reliance affects source guessing. Based on the empirical evidence reported in the four manuscripts, knowledge reliance in source guessing is rather (1) unrelated to decisional uncertainty caused by stereotype knowledge tied to age categories, (2) to some extent stable across time and knowledge domains, (3) used in an automatic manner detached from cognitive capacities, and (4) not generalizable to source attributions for novel information from specific, known source exemplars.

5.1 Strengths, Limitations, and Future Directions

Besides the substantive gain in learning about the nature of source guessing, this thesis provides a decisive methodological foundation by means of cognitive modeling of source-monitoring processes in addition; namely estimating hierarchical MPT models based on Bayesian statistics. Source-monitoring processes, from which source guessing was of highest priority for the present thesis, were modeled (1) separately for each individual, (2) accounting for parameter correlations, and (3) accounting for correlations with external variables such as process-tracing indices (Manuscript I), personality traits and cognitive processing styles (Manuscript II), or subjective contingency judgments (Manuscript II, III, IV). One by-product of this thesis which should be acknowledged as a strength is that knowledge reliance in source guessing within the domain of age stereotypes is pronounced to a comparable extent across manuscripts (Manuscript I, II, and III). Whereas the reproducibility of an age bias in source guessing is not the main focus of this thesis, it nonetheless demonstrates its robustness of knowledge reliance in source guessing—despite individual differences. The latter display an additional advantage of the methodological approach followed in this thesis. We are able to study source guessing also on an individual level, for which data of the respective manuscripts are made publicly available on the Open Science Framework (OSF). These data may serve the interest of the scientific community and the purpose of reproducible research—particularly important in times of a replication crisis in psychological science (Open Science Collaboration, 2015).

In the same vein, another favorable characteristic of the statistical approach in combination with the user-friendly software tool *TreeBUGS* (Heck, Arnold, et al., 2018) became apparent: parameter-recovery simulations. As power analyses in the Bayesian MPT framework neither exist in the classical sense nor may they be appropriate for each individual case (see Wagenmakers et al., 2015, for a discussion), simulation-based analyses are of great significance. Of course, simulations are informative to determine the required sample size prior to data collection given the true parameters of interest are known. These, however, cannot be taken for granted—particularly not if researchers suffer from a lack of reference concerning previous studies as an orientation for parameter values. In these cases and due to potentially inconclusive results, as demonstrated in the first two manuscripts of this thesis, post hoc parameter-recovery simulations can be a valuable instrument to quantify the extent to which the obtained empirical evidence is considered robust and reliable—both from a perspective of a basic researcher who seeks to answer a research question to the best of their knowledge and from a perspective of trustworthy and openly-communicated results (Open Science Collaboration, 2015). This thesis can be seen as an application example of such analyses to explicitly account for potential limitations in the precise measurement and estimation of cognitive processes and source guessing in particular.

Albeit this thesis lays a solid foundation for a more profound understanding of knowledge reliance in source guessing, the research conducted within the scope of the four manuscripts can only be considered as a starting point for many more, prospective studies that can adapt the study's experimental manipulations, procedure, or stimulus material, to be compared with the results outlined here and presented in the respective manuscripts in greater details. In the following, I address three notable challenges that have emerged in this thesis and which solving should be incorporated in future research.

First, the present thesis is mostly restricted to the knowledge domain of age stereotypes. This inherently poses the question of the generalization of the present findings to other domains. However, based on the source-monitoring literature, there is not any indication why the proposed underlying mechanisms in source guessing should necessarily differ between knowledge domains. Thus, the findings should not be limited to the special case of age stereotypes (which, nonetheless, needs to await to be backed up by appropriate empirical evidence). Rather, I am convinced that the use of this item material is beneficial at least for two reasons. Notably, the verbal stimuli (i.e., everyday statements) come closely to how we process information in our social environment and is therefore more tailored to transfer the results to actual behavioral consequences with regard to source attributions outside of the laboratory setting. I am aware, however, that there are of course knowledge domains (e.g., scene or profession schema) that, in general, trigger more pronounced biases (Bayen et al., 2000; Ehrenberg & Klauer, 2005; Küppers & Bayen, 2014; Schaper et al., 2019) for which it would be also of great interest to look at source guessing more closely. Moreover, this thesis primarily focused on age stereotypes for the sake of their content and societal relevance in everyday contexts. I elaborate on the implications of this knowledge domain in Section 5.2.

Second, modeling source-monitoring data with more advanced tools (than the ones applied in this thesis) enables prospective research to draw conclusions about the nature of the involved cognitive processes on a more fine-grained level. Recently, promising approaches have been proposed by Hartmann et al. (in press), Heck, Erdfelder, et al. (2018), and Klauer and Kellen (2018). In particular, supplementing the approaches used in in Manuscript I (L. Wulff & Scharf, 2020), in which we studied the cognitive dynamics of source-monitoring processes with mouse tracking, with more sophisticated analyses to examine the influence of pre-experimental knowledge on source guessing even more detailed certainly would increase the validity of the results. For instance, with *generalized processing tree models* (Heck, Erdfelder, et al., 2018) discrete, categorical data can be modeled jointly with continuous variables (such as process-tracing indices) given a suitable task set up and, hence, data structure (which did not apply to Manuscript I). Another example of how the outlined research could benefit from model-based extensions in upcoming studies is to incorporate reaction times into MPT models as recommended by Klauer and Kellen (2018) and Hartmann et al. (in press). This method is highly valuable because the order in which processes occur can be modeled explicitly (Klauer & Kellen, 2018). In the particular case of resource dependence of knowledge reliance in source guessing (Manuscript III; L. Wulff & Kuhlmann, 2020b), the temporal order of source-monitoring processes is particularly interesting with regard to their rather automatic (e.g., fast) or controlled (e.g., slow; Hasher & Zacks, 1979; Schneider & Shiffrin, 1977) nature (as already highlighted by Spaniol & Bayen, 2002).

Third, due to the novelty of the research questions proposed in the four projects, a null-contingency manipulation (50%:50% distribution of item types to either source) is introduced as baseline. This necessity to study unbalanced item-source contingencies (i.e., 75% of each item type to either source) can be put up for discussion. More specifically, the application of such a less pronounced stereotype-based item-source contingency during encoding could trigger counter-stereotype guessing which, in turn, should be studied with disruption at retrieval once again (as done in Manuscript III; L. Wulff & Kuhlmann, 2020b). The variation in the distribution of item types to sources would be—from my point of view—of most interest to be studied in the following. The empirical evidence suggesting an automatic nature of knowledge reliance could be supplemented by load manipulations. In fact, an intriguing additional question would be to examine whether individuals adapt their source-guessing behavior to the ratio of counter-stereotype exemplars presented by each source in a first step replicating previous research (e.g., Bayen & Kuhlmann, 2011; Bayen et al., 2000; Ehrenberg & Klauer, 2005). But what would add on the already-acquired evidence on the nature of source guessing, is to investigate the impact of such counter-stereotype contingencies on source guessing under divided attention. Do individuals in the specific situation of knowledge application during retrieval attempt to implement the counter stereotype into their source-guessing behavior or, nonetheless, stick to their pre-experimental stereotypes? Answering this question in future studies is sensible from a perspective of a basic researchers but would be also highly informative by means of potential real-life implications. If individuals fall back on pre-experimental knowledge in source guessing under limited cognitive capacities (be it induced through experimental or more natural settings such as aging)—irrespective of the encoded counter-stereotype contingency—it would challenge the effectiveness of potential interventions counteracting the reliance on pre-experimental knowledge (as, for instance, successfully demonstrated by an explicit stereotype negation instruction; Marsh et al., 2006).

5.2 Implications of Knowledge Reliance

As raised before, a potential danger is linked to biased source guessing in real life, especially in situations of uncertainty in source attributions where the stereotype or schema is applied as default guessing strategy. This may be rational and adaptive given that these knowledge structures oftentimes lead to correct source attributions. But stereotypes (and schemas) can be also referred to as simplification of our social life (cf. Bordalo et al., 2016) neglecting individual differences and context dependence. By this, they can crucially be considered a misleading cue for source attributions to a certain extent. Namely, to stick to the knowledge domain of age stereotype as an example, whenever individuals do not behave as commonly expected for their age-group membership, stereotypes and schemas become an invalid an potentially harmful source-decision cue. Those individuals are mistakenly associated with behaviors and attitudes that do not fully or in any way at all account for their true personality features. Remarkably, even if individualized information from specific sources is acquired in the past, future inferences for these sources are nonetheless most likely based on generic, prototypical knowledge about the social categories that this person belongs to (L. Wulff et al., 2020). This lack of transfer from knowledge based on previously-learned information to applied to novel information is undoubtedly critical. Misattributions, even innocently acquired and applied, can then have severe implications in various areas—reaching beyond the setting or situation in which a source attribution is made. For instance, biased source attributions can harm future impression formation about people because "if perceivers cannot attribute behaviors to their proper source, then they cannot form accurate impressions of other people. [...] Moreover, perceivers might not have the time or resources to sift through the details of their memories to disentangle the contexts" (Sherman & Bessenoff, 1999, p. 109). In the longer run, resistant misattributions may then shape general (implicit and explicit) attitudes resulting in a maintenance or even reinforcement of such stereotypes in society with "obvious negative implications for the individuals who are stereotyped" (Sherman & Bessenoff, 1999, p. 110). For age stereotypes, significant implications were reported for the stereotyped age groups in the literature (e.g., wrongful convictions; Lamont et al., 2015; Levy, 2009; Swift et al., 2017; Swift et al., 2012). The severity of source misattributions due to uncertainty while attributing information to its original context is even more critical when it comes to the legal decision-making context. In case of eyewitness testimony in which marginal details in the progression of events can be decisive at court, misattributions based on general knowledge can have serious consequences to the affected person (groups; Lindsay, 1990, 1994, 2014; Lindsay & Johnson, 1989; Mitchell & Johnson, 2000; Sherman et al., 2003). Sherman and Bessenoff

(1999) aptly described the use of stereotype cues during source attributions as a “double-edged sword that trades off efficiency and accuracy” (p. 110). Nonetheless, the costs of knowledge reliance in source guessing due to missing accuracy are mostly unstudied and one can only speculate about the long-term implications of misattribution in society (e.g., Levy et al., 2015).

This thesis brings light into the darkness and provides a remedy for potential costs of knowledge reliance for various reasons. Throughout all manuscripts, and thus knowledge domains, the extent to which participants relied on pre-experimental knowledge varied considerably between individuals. This means of course that some individuals strongly guess based on stereotypes but others do not. They are not biased in either direction or even show counter-stereotype guessing. Moreover, whereas we observed some trait-like stability of source guessing across time and knowledge domains in Manuscript II (L. Wulff & Kuhlmann, 2020a), knowledge reliance could not be undoubtedly characterized as highly stable indicating variation also within individuals. Once again, biased source guessing is not an inevitable backup for (source-)memory failure whose pronounced state-like, situational, determinants need to be addressed in future assessments. Finally, we identified the accurate encoding of item-source contingencies as a crucial factor for unbiased source guessing as most evident from Manuscript III (L. Wulff & Kuhlmann, 2020b) and in line with previous studies (e.g., Arnold et al., 2013; Bayen & Kuhlmann, 2011; Kuhlmann et al., 2016; Kuhlmann et al., 2012; Spaniol & Bayen, 2002). The relevance of item-source contingencies acquired during the learning phase is emphasized arising from the automatic counteraction of stereotype influences through contingencies. If preconditions are met for their accurate perception, source guessing is mostly based on the item-source contingency (a tendency that is observed even in older age) but if they are not met (or the contingency knowledge is not transferred to novel stimuli without a contextual representation), pre-experimental knowledge influences step in. This underpins the subtle role of unbiased encoding in source guessing and lays, together with other predictor variables, the foundation for effectively designing counter-knowledge based interventions targeted to these multifold influences bearing in mind the pronounced individual differences.

5.3 Conclusion

In this thesis, I contribute to the current status quo of judgment processes in source monitoring by uncovering the nature of knowledge reliance in source guessing within a superordinate theoretical and statistical framework. This thesis herein is an important step towards the understanding of biases in judgment processes during source attributions. It can thereby serve as an inspiration for future research on reconstructive inference processes in the memory and decision-making context, and in source monitoring in particular—both from a basic research perspective but also in terms of applied settings—and initiates how to effectively rule out these biases in judgment processes if ultimately leading to misattributions.

6 Bibliography

Abney, D. H., McBride, D. M., Conte, A. M., & Vinson, D. W. (2015). Response dynamics in prospective memory. *Psychonomic Bulletin & Review*, *22*(4), 1020–1028. https://doi.org/10.3758/s13423-014-0771-6

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203

Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin & Review*, *20*(2), 326–333. https://doi.org/10.3758/s13423-012-0342-7

Bargh, J. A. (1994). The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition. In R. S. Wyer & T. K. Skrull (Eds.), *Handbook of social cognition: Basic processes; applications* (pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. https://doi.org/10.1007/s00572-005-0022-9

Batchelder, W. H., & Batchelder, E. (2008). Metacognitive guessing strategies in source monitoring. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 211–244). New York, NY: Psychology Press. https://doi.org/10.4324/9780203805503.ch12

Batchelder, W. H., Hu, X., & Riefer, D. M. (1994). Analysis of a model for source monitoring. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 51–65). New York, NY: Springer.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review*, *97*(4), 548–564. https://doi.org/10.1037/0033-295X.97.4.548

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language*, *64*(1), 1–17. https://doi.org/10.1016/j.jml.2010.09.001

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*(3), 480–500. https://doi.org/10.3758/BF03198562

Bell, R., Mieth, L., & Buchner, A. (2015). Appearance-based first impressions and person memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 456–472. https://doi.org/10.1037/xlm0000034

Bell, R., Mieth, L., & Buchner, A. (2020). Source attributions for detected new items: Persistent evidence for schematic guessing. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021820911004

Bodenhausen, G. V., Macrae, C. N., & Sherman, J. W. (1999). On the dialectics of discrimination: Dual processes in social stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 271–290). New York, NY: Guilford Press.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics*, *131*(4), 1753–1794. https://doi.org/10.1093/qje/qjw029

Bröder, A., Noethen, D., Schütz, J., & Bay, P. (2007). Utilization of covariation knowledge in source monitoring: No evidence for implicit processes. *Psychological Research*, *71*(5), 524–538. https://doi.org/10.1007/s00426-006-0047-5

Calanchini, J., Rivers, A. M., Klauer, K. C., & Sherman, J. W. (2018). Multinomial processing trees as theoretical bridges between cognitive and social psychology. *Psychology of Learning and Motivation*, *69*, 39–65. https://doi.org/10.1016/bs.plm.2018.09.002

Cassidy, B. S., Sprout, G. T., Freeman, J. B., & Krendl, A. C. (2017). Looking the part (to me): effects of racial prototypicality on race perception vary by prejudice. *Social Cognitive and Affective Neuroscience*, *12*(4), 685–694. https://doi.org/10.1093/scan/nsw177

Cloutier, J., Freeman, J. B., & Ambady, N. (2014). Investigating the early stages of person perception: The asymmetry of social categorization by sex vs. age. *PLoS ONE*, *9*(1), 1–7. https://doi.org/10.1371/journal.pone.0084677

Danner, D., Rammstedt, B., Bluemke, M., Treiber, L., Berres, S., Soto, C. J., & John, O. P. (2016). Die deutsche Version des Big Five Inventory 2 (BFI-2) [The German version of the Big Five Inventory 2]. *Zusammenstellung Sozialwissenschaftlicher Items und Skalen (ZIS)*. https://doi.org/10.6102/zis247

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology*, *41*(4), 369–387. https://doi.org/10.1016/j.jesp.2004.08.001

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Person-*

ality and Social Psychology, *71*(2), 390–405. https://doi.org/10.1037/0022-3514.71.2.390

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie / Journal of Psychology*, *217*, 108–124. https://doi.org/10.1027/0044-3409.217.3.108

Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science*, *20*(10), 1183–1188. https://doi.org/10.1111/j.1467-9280.2009.02422.x

Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General*, *137*(4), 673–690. https://doi.org/10.1037/a0013875

Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, *45*(1), 83–97. https://doi.org/10.3758/s13428-012-0225-x

Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology*, *46*(1), 179–185. https://doi.org/10.1016/j.jesp.2009.10.002

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–511. https://doi.org/10.1214/ss/1177011136

Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, *123*(1), 21–32. https://doi.org/10.1016/j.cognition.2011.12.002

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2014). Measures of implicit attitudes may conceal differences in implicit associations: The case of antiaging bias. *Social Psychological and Personality Science*, *5*(3), 271–278. https://doi.org/10.1177/1948550613499239

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. S. Wyer & T. K. Skrull (Eds.), *Handbook of social cognition: Basic processes; applications* (pp. 1–68). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Hartmann, R., Johannsen, L., & Klauer, K. C. (in press). rtmpt: An R package for fitting response-time extended multinomial processing tree models. *Behavior Research Methods*.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General*, *108*(3), 356–388. https://doi.org/10.1037/0096-3445.108.3.356

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7

Heck, D. W., Erdfelder, E., & Kieslich, P. J. (2018). Generalized processing tree models: Jointly modeling discrete and continuous variables. *Psychometrika*, *83*(4), 893–918. https://doi.org/10.1007/s11336-018-9622-0

Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes and Intergroup Relations*, *18*(3), 384–401. https://doi.org/10.1177/1368430214538325

Hicks, J. L., & Cockman, D. W. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language*, *48*(3), 489–501. https://doi.org/10.1016/S0749-596X(02)00537-5

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the em algorithm. *Psychometrika*, *59*(1), 21–47. https://doi.org/10.1007/BF02294263

Hütter, M., & Klauer, K. C. (2016). Applying processing trees in social psychology. *European Review of Social Psychology*, *27*(1), 116–159. https://doi.org/10.1080/10463283.2016.1212966

Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology*, *102*(1), 116–131. https://doi.org/10.1037/a0025335

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition*, *40*(8), 1163–1177. https://doi.org/10.3758/s13421-012-0226-0

Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review*, *21*(5), 1272–1280. https://doi.org/10.3758/s13423-014-0608-3

Keefe, R. S. E., Arnold, M. C., Bayen, U. J., McEvoy, J. P., Wilson, W. H., Larkin, W., Read, J., Service, E. I., Mill, D., Orellana, G., Alvarado, L., Muñoz-Neira, C., Ávila, R., Méndez, M. F., Slachevsky, A., Read, J., Fosse, R., Moskowitz, A., Perry, B., . . . Diforio, D. (2002). Source-monitoring deficits for self-generated stimuli in schizophrenia: Multinomial modeling of data from three sources. *Schizophrenia Research*, *57*(1), 51–67. https://doi.org/10.1016/S0920-9964(01)00306-1

Keller, J., Bohner, G., & Erb, H.-P. (2000). Intuitive und heuristische Urteilsbildung–verschiedene Prozesse? Präsentation einer deutschen Fassung des "Rational-Experiential Inventory" sowie neuer Selbstberichtskalen zur Heuristiknutzung [Intuitive and heuristic judgment–different processes? Presentation of a German version of the Rational-Experiential Inventory and of new self-report scales of heuristic uses]. *Zeitschrift für Sozialpsychologie*, *31*(2), 87–101. https://doi.org/10.1024//0044-3514.31.2.87

Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods* (pp. 111–130). New York, NY: Routledge.

Kieslich, P. J., Wulff, D. U., Henninger, F., & Haslbeck, J. M. B. (2018). *mousetrap: Process and analyze mouse-tracking data* (Version 3.1.1)[Computer software]. https://CRAN.R-project.org/package=mousetrap

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika*, *71*, 7–31. https://doi.org/10.1007/s11336-004-1188-3

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klauer, K. C., & Ehrenberg, K. (2005). Social categorization and fit detection under cognitive load: Efficient or effortful? *European Journal of Social Psychology*, *35*(4), 493–516. https://doi.org/10.1002/ejsp.266

Klauer, K. C., & Kellen, D. (2018). RT-MPTs: Process models for response-time distributions based on multinomial processing trees with applications to recognition memory. *Journal of Mathematical Psychology*, *82*, 111–130. https://doi.org/10.1016/j.jmp.2017.12.003

Klauer, K. C., & Meiser, T. (2000). A source-monitoring analysis of illusory correlations. *Personality and Social Psychology Bulletin*, *26*(9), 1074–1093. https://doi.org/10.1177/01461672002611005

Koop, G. J., & Criss, A. H. (2016). The response dynamics of recognition memory: Sensitivity and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 671–685. https://doi.org/10.1037/xlm0000202

Kuhlmann, B. G., & Bayen, U. J. (2016). Metacognitive aspects of source monitoring. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 149–168). New York, NY: Oxford University Press.

Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. *Psychology and Aging*, *31*(8), 875–889. https://doi.org/10.1037/pag0000140

Kuhlmann, B. G., Kornadt, A. E., Bayen, U. J., Meuser, K., & Wulff, L. (2017). Multidimensionality of younger and older adults' age stereotypes: The interaction of life domain and adjective dimension. *Journals of Gerontology, Series B: Psychological*

Sciences and Social Sciences, *72*(3), 436–440. https://doi.org/10.1093/geronb/gbv049

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1365–1376. https://doi.org/10.1037/a0028147

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *Quarterly Journal of Experimental Psychology*, *67*(10), 2042–2059. https://doi.org/10.1080/17470218.2014.904914

Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and Aging*, *30*(1), 180–193. https://doi.org/10.1037/a0038586

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7. https://doi.org/10.1016/j.jmp.2010.08.013

Lee, M. D. (2018). Bayesian methods in cognitive modeling. In J. Wixted & E.-J. Wagenmakers (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (pp. 1–48). Hoboken, NJ: John Wiley & Sons. https://doi.org/10.1002/9781119170174.epcn502

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge University Press.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*(2), 275–287. https://doi.org/10.1037/0022-3514.72.2.275

Levy, B. R. (2009). Stereotype embodiment: A psychosocial approach to aging. *Current Directions in Psychological Science*, *18*(6), 332–336. https://doi.org/10.1111/j.1467-8721.2009.01662.x

Levy, B. R., Slade, M. D., Chung, P. H., & Gill, T. M. (2015). Resiliency over time of elders' age stereotypes after encountering stressful events. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *70*(6), 886–890. https://doi.org/10.1093/geronb/gbu082

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1077–1083. https://doi.org/10.1037/0278-7393.16.6.1077

Lindsay, D. S. (1994). Memory source monitoring and eyewitness testimony. In D. F. Ross, J. D. Reed, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 27–55). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511759192.003

Lindsay, D. S. (2014). Memory source monitoring applied. In T. Perfect & D. S. Lindsay (Eds.), *The sage handbook of applied memory* (pp. 59–75). London, UK: Sage. https://doi.org/10.1016/j.jelechem.2018.02.006

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*(3), 349–358. https://doi.org/10.3758/BF03198473

Maldonado, M., Dunbar, E., & Chemla, E. (2019). Mouse tracking as a window into decision making. *Behavior Research Methods*, *51*(3), 1085–1101. https://doi.org/10.3758/s13428-018-01194-x

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). Gender and orientation stereotypes bias source-monitoring attributions. *Memory*, *14*(2), 148–160. https://doi.org/10.1080/09658210544000015

Mather, M., Johnson, M. K., & De Leonardis, D. M. (1999). Stereotype reliance in source monitoring: Age differences and neuropsychological test correlates. *Cognitive Neuropsychology*, *16*(3-5), 437–458. https://doi.org/10.1080/026432999380870

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324.

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205–235. https://doi.org/10.1007/s11336-013-9374-9

Meiser, T., Sattler, C., & von Hecker, U. (2007). Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *The Quarterly Journal of Experimental Psychology*, *60*(7), 1015–1040. https://doi.org/10.1080/17470210600875215

Michalkiewicz, M., & Erdfelder, E. (2016). Individual differences in use of the recognition heuristic are stable across time, choice objects, domains, and presentation formats. *Memory & Cognition*, *44*(3), 454–468. https://doi.org/10.3758/s13421-015-0567-6

Mitchell, K. J., & Johnson, M. K. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The oxford handbook of memory* (pp. 179–195). New York, NY: Oxford University Press.

Mitchell, K. J., & Johnson, M. K. (2009). Source monitoring 15 years later: What have we learned from fMRI about the neural mechanisms of source memory? *Psychological Bulletin*, *135*(4), 638–677. https://doi.org/10.1037/a0015849

Moshagen, M. (2010). multiTree : A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Odum, A. L. (2011). Delay discounting: Trait variable? *Behavioural Processes*, *87*(1), 1–9. https://doi.org/10.1016/j.beproc.2011.02.007

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716. https://doi.org/10.1126/science.aac4716

Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin and Review*, *19*(5), 906–913. https://doi.org/10.3758/s13423-012-0281-3

Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology*, *26*(3), 199–216. https://doi.org/10.1016/0022-1031(90)90035-K

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling* [Computer software]. http://mcmc-jags.sourceforge.net

Rammstedt, B., Danner, D., Soto, C. J., & John, O. P. (2020). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment*, *36*(1), 149–161. https://doi.org/10.1027/1015-5759/a000481

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318–339. https://doi.org/10.1037/0033-295X.95.3.318

Riefer, D. M., Hu, X., & Batchelder, W. H. (1994). Response strategies in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 680–693. https://doi.org/10.1037/0278-7393.20.3.680

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*(4), 573–604. https://doi.org/10.3758/BF03196750

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metamemory expectancy illusion and schema-consistent guessing in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(3), 470–496. https://doi.org/10.1037/xlm0000602

Scheibehenne, B., & Pachur, T. (2015). Using Bayesian hierarchical parameter estimation to assess the generalizability of cognitive models of choice. *Psychonomic Bulletin & Review*, *22*(2), 391–407. https://doi.org/10.3758/s13423-014-0684-4

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Sherman, J. W., & Bessenoff, G. R. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science*, *10*(2), 106–110. https://doi.org/10.1111/1467-9280.00116

Sherman, J. W., Groom, C. J., Ehrenberg, K., & Klauer, K. C. (2003). Bearing false witness under pressure: Implicit and explicit components of stereotype-driven memory distortions. *Social Cognition*, *21*(3), 213–246. https://doi.org/10.1521/soco.21.3.213.25340

Singmann, H., & Kellen, D. (2013). MPTinR: Analysis of multinomial processing tree models in R. *Behavior Research Methods*, *45*(2), 560–575. https://doi.org/10.3758/s13428-012-0259-0

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*(1), 167–183. https://doi.org/10.1016/j.jmp.2009.06.007

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-s and BFI-2-xs. *Journal of Research in Personality*, *68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 631–651. https://doi.org/10.1037/0278-7393.28.4.631

Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211. https://doi.org/10.1111/j.1467-8721.2006.00437.x

Stahl, C., & Klauer, K. C. (2007). HMMTree: A computer program for latent-class hierarchical multinomial processing tree models. *Behavior Research Methods*, *39*(2), 267–273. https://doi.org/10.3758/BF03193157

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Stroebe, W., & Insko, C. A. (1989). Stereotype, prejudice, and discrimination: Changing conceptions in theory and research. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and prejudice* (pp. 3–34). New York, NY: Springer Series in Social Psychology. https://doi.org/10.1007/978-1-4612-3582-8_1

Swift, H. J., Abrams, D., Lamont, R. A., & Drury, L. (2017). The risks of ageism model: How ageism and negative attitudes toward age can be a barrier to active aging. *Social Issues and Policy Review*, *11*(1), 195–231. https://doi.org/10.1111/sipr.12031

Swift, H. J., Lamont, R. A., & Abrams, D. (2012). Are they half as strong as they used to be? An experiment testing whether age-related social comparisons impair older people's hand grip strength and persistence. *BMJ Open*, *2*, e001064. https://doi.org/10.1136/bmjopen-2012-001064

Wagenmakers, E.-J., Verhagen, J., Ly, A., Bakker, M., Lee, M. D., Matzke, D., Rouder, J. N., & Morey, R. D. (2015). A power fallacy. *Behavior Research Methods*, *47*(4), 913–917. https://doi.org/10.3758/s13428-014-0517-4

Wickelmaier, F., & Zeileis, A. (2018). Using recursive partitioning to account for parameter heterogeneity in multinomial processing tree models. *Behavior Research Methods*, *50*(3), 1217–1233. https://doi.org/10.3758/s13428-017-0937-z

Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science*, *20*(11), 1428–1435. https://doi.org/10.1111/j.1467-9280.2009.02448.x

Wulff, D. U., Haslbeck, J. M. B., Kieslich, P. J., Henninger, F., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: Detecting types in movement trajectories. In M. Schulte-Mecklenbeck, A. Kühberger, & J. G. Johnson (Eds.), *A handbook of process tracing methods* (pp. 131–145). New York, NY: Routledge.

Wulff, L., & Kuhlmann, B. G. (2020a). Is knowledge reliance in source guessing a cognitive trait? Examining stability across time and domain. *Memory & Cognition*, *48*(2), 256–276. https://doi.org/10.3758/s13421-019-01008-1

Wulff, L., & Kuhlmann, B. G. (2020b). *Is knowledge reliance in source guessing automatic or controlled? Evidence from divided attention and aging*. Manuscript submitted for publication.

Wulff, L., Kuhlmann, B. G., Bell, R., & Mieth, L. (2020). *Guess what?! Different source-guessing mechanisms for old versus new information*. Manuscript in preparation.

Wulff, L., & Scharf, S. E. (2020). Unpacking stereotype influences on source-monitoring processes: What mouse tracking can tell us. *Journal of Experimental Social Psychology*, *87*, 103917.

A Acknowledgements

This thesis has become reality merely due to the guidance and support of so many outstanding individuals. I feel incredibly privileged and it fills my heart with joy that you accompanied me on this journey. I would like to express my sincere gratitude to each and everyone one you, but I want to extend my particular thanks to some of you who were most significant to me over the last few years—both on a professional and a personal level.

I am deeply grateful to ...

... Bea, who triggered my interest in science during my Bachelor's degree in Düsseldorf and considered me as her very first doctoral student. If I sketch my PhD as a processing tree for which I could have had the possibility to decide on each day of the past 1,307 days whether I would like to pursue this route or not, my response certainly would have been "yes" every single day. I owe this to you as you supported me and my work reaching such a great extent. You guided me to find my way not only by means of projects related to this thesis but also when it came to side projects and I am so much looking forward to our continued collaborative work. You are such a shining example of a successful scientist and have been and still are a true mentor for me but, what is more, a valuable person I am happy to share memories with—reaching far beyond our professional life in academia.

... Edgar and Arndt, who challenged my thesis project when required and always provided me with extremely helpful and motivating feedback that made me iteratively reassess my research—thereby steadily improving my scientific thinking.

... Daniel and Pascal, who equipped me with all the knowledge necessary to transfer my research questions into good scientific practice by teaching me *R*, *TreeBUGS*, and *OpenSesame*.

... Sophie, for whom I could tell from the very first moment that her presence would be essential for me within the next years. I have learned so much from you and still do.

I am genuinely impressed by all yours and deeply fulfilled having you in my life—both as a research colleague and even more so as a mutual emotional counterbalance factor.

... Emelie, who has enlightened my life immediately since we first met in the office, and whose company is nothing less than a gift. We became "acquainted" by watching only one, but a truly life-changing, episode of a TV series. Ever after, you were and will always mean so much more than that to me.

... my friends and colleagues from the third floor in L13, with whom I have enjoyed life in all its facets during the past years.

- ... Tine, Michi, and Letti, my love, who are three fabulous reasons to believe it's true: "A journey is best measured in friends not in miles". You offered me a save harbor in turbulent times open 24 hours a day and reminded me to never forget to play hard in times of hard work.
- ... Philipp, who was the best barista and even better conversational and cooking partner. Thank you for always bringing a sparkle to our floor.
- ... Hilal, who supported me during the last months of my PhD (morally and practically with the best breakfast).

... Franzi, Jana, Kathi, Lena, David, Malte, and Martin, for sharing such great times during numerous lunch breaks, countless boozy nights at the river, or at the Uniclub.

... my fellow CDSS PhD students, most importantly Mirka, Stella, and Sebastian, the best welcome committee in Mannheim, who made my arrival feel like coming home.

... all our student assistants who have helped me during the past years: Alina, Franzi, Katja, Jana, Jenny, Jule, Michelle, Paula, Pia, Selina, and Theresa.

... every single younger and older adult who came to the lab and participated in our experiments. None of this would have been possible without you!

... Annette, Rahel, and Milanka, for their moral support and great assistance in organizational matters.

... my friends Franzi, Lea, Magda, Anke, Nina, Vanessa, and Alessa from Düsseldorf, and Nadine, Nina, and Christian from Paderborn, with whom I had marvelous times (as if nothing had changed) when leaving the PhD bubble every once in a while.

... my father, who was a secret source of motivation from afar. Without you knowing, you taught me to always strive for the best and to make the most out of all true to "I figure life is a gift and I don't intend on wasting it. You don't know what hand you're gonna get dealt next. You learn to take life as it comes at you... to make each day count." In loving memory, I promise I always will.

... my mother and my sister, whose unconditional love and fundamental trust brought me here in the first place. You are both extraordinary examples of the most courageous women I ever met and empowered me to follow this path, knowing that this is what my heart is beating for. You always cheered me up, shared my excitement, and served as a driving force and an oasis of a calm in unsettled times. How fortunate can an individual be to be raised and accompanied by you two?

But foremost and from the bottom of my heart, I would like to thank Ansgar to whom I felt so close yet being so far away during the past years. It was you who encouraged me to begin my PhD and brought me back on track when I tend to get lost. This thesis is dedicated to you. What a coincidence that this challenging time for both of us also marked the most precious one and brought us, albeit physically separated, even closer together, as we began to spend the rest of our lives together. —"At this moment, there are almost 7.8 billion people in the world. [...] And sometimes, all you need is one."

Liliane Wulff
Mannheim, March 2020

B Statement of Originality

1. I hereby declare that the presented doctoral dissertation with the title *The Nature of Knowledge Reliance in Source Guessing* is my own work.
2. I did not seek unauthorized assistance of a third party and I have employed no other sources or means except the ones listed. I clearly marked any quotations derived from the works of others.
3. I did not present this doctoral dissertation or parts of it at any other higher education institution in Germany or abroad.
4. I hereby conform the accuracy of the declaration above.
5. I am aware of the significance of this declaration and the legal consequences in case of untrue or incomplete statements.

I affirm in lieu of oath that the statements above are to the best of my knowledge true and complete.

Signature:

Date:

C Co-Authors' Statements

Co-Author: Sophie E. Scharf

It is hereby confirmed that the following manuscript included in the thesis *The Nature of Knowledge Reliance in Source Guessing* was primarily conceived and written by Liliane Wulff, PhD candidate at the at the *Center for Doctoral Studies in Social and Behavioral Sciences* of the *Graduate School of Economic and Social Sciences* at the University of Mannheim:

Wulff, L., & Scharf, S. E. (2020). Unpacking stereotype influences on source-monitoring processes: What mouse tracking can tell us. *Journal of Experimental Social Psychology*, *87*, 103917. https://doi.org/10.1016/j.jesp.2019.103917

I sign this statement to the effect that Liliane Wulff is credited as the primary source of the ideas and the main author of the manuscript as she derived the theoretical and methodological background, collected the data, implemented the statistical MPT analyses/simulations, wrote the first drafts, and contributed to improving and revising the manuscript. I contributed to developing and refining the theoretical background, I performed the statistical analyses of linear mixed models and prototype clusters, designed the graphical illustrations of results, and provided recommendations for structuring and improving the manuscript.

M.Sc. Sophie E. Scharf
Mannheim, March 2020

Co-Author: Beatrice G. Kuhlmann

It is hereby confirmed that the following manuscripts included in the thesis *The Nature of Knowledge Reliance in Source Guessing* were primarily conceived and written by Liliane Wulff, PhD candidate at the at the *Center for Doctoral Studies in Social and Behavioral Sciences* of the *Graduate School of Economic and Social Sciences* at the University of Mannheim:

Wulff, L., & Kuhlmann, B. G. (2020). Is knowledge reliance in source guessing a cognitive trait? Examining stability across time and domain. *Memory & Cognition*, *48*(2), 256–276. https://doi.org/10.3758/s13421-019-01008-1

Wulff, L., & Kuhlmann, B. G. (2020). *Is knowledge reliance in source guessing automatic or controlled? Evidence from divided attention and aging*. Manuscript submitted for publication.

Wulff, L., Bell, R., Mieth, L., & Kuhlmann, B. G. (2020). *Guess what?! Different source-guessing mechanisms for old versus new information*. Manuscript in preparation.

I sign this statement to the effect that Liliane Wulff is credited as the primary source of the ideas and the main author of all three manuscripts as she derived the theoretical and methodological background, collected the data, implemented the statistical analyses/simulations, wrote the first drafts, and contributed to improving and revising the manuscripts. I contributed to developing and refining the theoretical background, suggested ideas for the analyses and their interpretations, devised prospective research questions, generated the basic concept for the experiment conducted in the last manuscript, and provided recommendations for structuring and improving the manuscripts.

Prof. Dr. Beatrice G. Kuhlmann
Mannheim, March 2020

Co-Author: Raoul Bell

It is hereby confirmed that the following manuscript included in the thesis *The Nature of Knowledge Reliance in Source Guessing* was primarily conceived and written by Liliane Wulff, PhD candidate at the at the *Center for Doctoral Studies in Social and Behavioral Sciences* of the *Graduate School of Economic and Social Sciences* at the University of Mannheim:

Wulff, L., Bell, R., Mieth, L., & Kuhlmann, B. G. (2020). *Guess what?! Different source-guessing mechanisms for old versus new information.* Manuscript in preparation.

I sign this statement to the effect that Liliane Wulff is credited as the primary source of the ideas and the main author of the manuscript as she derived the theoretical and methodological background, collected the data, implemented the statistical analyses, wrote the first draft, and contributed to improving and revising the manuscript. I contributed to refining the theoretical background, developed the idea for the experiment conducted in the manuscript, and provided recommendations for related, prospective research ideas described in the *Discussion* section of the manuscript.

PD Dr. Raoul Bell
Düsseldorf, March 2020

Co-Author: Laura Mieth

It is hereby confirmed that the following manuscript included in the thesis *The Nature of Knowledge Reliance in Source Guessing* was primarily conceived and written by Liliane Wulff, PhD candidate at the at the *Center for Doctoral Studies in Social and Behavioral Sciences* of the *Graduate School of Economic and Social Sciences* at the University of Mannheim:

Wulff, L., Bell, R., Mieth, L., & Kuhlmann, B. G. (2020). *Guess what?! Different source-guessing mechanisms for old versus new information.* Manuscript in preparation.

I sign this statement to the effect that Liliane Wulff is credited as the primary source of the ideas and the main author of the manuscript as she derived the theoretical and methodological background, collected the data, implemented the statistical analyses, wrote the first draft, and contributed to improving and revising the manuscript. I contributed to refining the theoretical background, developed the idea for the experiment conducted in the manuscript, and provided recommendations for related, prospective research ideas described in the *Discussion* section of the manuscript.

L. Mieth

Dr. Laura Mieth
Düsseldorf, March 2020

D Copies of Manuscripts

Journal of Experimental Social Psychology 87 (2020) 103917

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

Case Report

Unpacking stereotype influences on source-monitoring processes: What mouse tracking can tell us[☆]

Liliane Wulff[a,*], Sophie E. Scharf[a,b]

[a] Department of Psychology, School of Social Sciences, University of Mannheim, Germany
[b] Social Cognition Center Cologne, University of Cologne, Germany

A R T I C L E I N F O

Keywords:
Process tracing
Mouse tracking
Source monitoring
Social categorization

A B S T R A C T

The goal of this study was to understand the cognitive dynamics of stereotype influences on source monitoring employing mouse tracking. By continuously recording cursor movements, we examined how stereotypical knowledge influences decision uncertainty when processing and later remembering stereotype-consistent and -*in*consistent exemplars of the age categories of "young" and "old". In a source-monitoring task, participants (N = 60) learned age-stereotype consistent or -*in*consistent statements from two different-aged sources (young vs. old person) that they attributed to their original sources via mouse clicks in a later memory test. Our results showed that individuals experienced cognitive conflict during source attributions depending on both the correctness of the source response and whether the original source was (*in*)consistent with the stereotype of the respective age group reflected in the statement. This pattern of results was supplemented by the analysis of prototypical mouse-trajectory clusters. Modeling individual source-monitoring processes revealed that individuals' experienced conflict was less pronounced when they remembered the source and was unrelated to guessing resulting from memory failure. These results highlight the benefits of combining cognitive modeling and process-tracing techniques to unpack the mechanisms behind social influences on source monitoring. The methodology of mouse tracking illuminated the role of stereotypes in the underlying cognitive processes during source attributions that is not evident from discrete categorical responses. For designed counter-stereotypical interventions, process-tracing methods may also be used to test their effectiveness on cognitive processes involved in source monitoring.

1. Introduction

In our daily life, we face a constant stream of information. When later remembering a piece of information, it is often important to know the source of the information in order to judge the quality or trustworthiness of the just-remembered information. Yet it is often hard to remember a source and it has been shown that people use prior knowledge, such as stereotypes, relevant to the piece of information to make an informed guess as to what the source of this information is (e.g., Kuhlmann, Bayen, Meuser, & Kornadt, 2016). This use of prior knowledge has been observed, for instance, in the domain of age stereotypes in that people are more likely to guess that statements such as "I am discontent with my health" were said by an older person than by a younger person based on the consistency of the stereotype reflected in the statement and the person's age (Kuhlmann et al., 2016). Although remembering a stereotype-consistent source for a piece of information seems innocent enough, a systematic misapplication of stereotypical knowledge through biased source guessing may contribute to the persistence and reinforcement of stereotypes over time (e.g., Levy, Slade, Chung, & Gill, 2015) with potentially detrimental effects on the affected

☆ The research reported in this article was supported by the University of Mannheim's Graduate School of Economic and Social Sciences. Parts of this research were presented at the 2018 59th Annual Meeting of the Psychonomic Society in New Orleans, USA. Data and analyses can be retrieved from OSF: https://osf.io/85936/?view_only=214fa149e30748cbbd07852c43fab419. We thank Pascal Kieslich for helpful advice on the analysis of mouse-tracking data, Beatrice G. Kuhlmann for detailed comments on the research idea and an earlier draft of this manuscript, and Daniel W. Heck for assistance with Bayesian hierarchical multinomial processing tree modeling. We thank Mary Frame and Britt Bolin for their diligent proofreading of the manuscript. We thank Theresa Pfeiffer and Jennifer Panitz for their assistance with the data collection. We thank two anonymous reviewers for their comments on an earlier draft of this manuscript. We affirm that in case the article is published we will share data on request for scientific purposes, and retain the raw data for five years after publication.
* Corresponding author at: Corresponding author at: Department of Psychology, School of Social Sciences, University of Mannheim, L13, 15, Room 321, Mannheim D-68161, Germany.
E-mail address: wulff@psychologie.uni-mannheim.de (L. Wulff).

https://doi.org/10.1016/j.jesp.2019.103917
Received 29 January 2019; Received in revised form 5 November 2019; Accepted 6 November 2019

person (groups). For instance, older adults perform worse when they are aware of their age-group related negative stereotypes (see Lamont, Swift, & Abrams, 2015, for a meta-analytic review). Up to now, the nature of stereotype influences on source-monitoring processes has not been fully understood. How exactly do individuals process and attribute previously acquired stereotype-consistent and -*in*consistent information to people of alternate social groups? Do they consider stereotype-consistent and -*in*consistent information during their source attribution? Do they experience decision conflict in a state of memory failure or stereotype-biased source guessing? And, if so, to what extent is decision uncertainty related to individuals' proneness for stereotype influences on their guessing behavior? Process-tracing methods may provide one possibility to study the cognitive dynamics of source attributions and may contribute to break down the underlying processes involved in it. Assuming that mouse movements reflect ongoing decision processes (Kieslich, Henninger, Wulff, Haslbeck, & Schulte-Mecklenbeck, 2019; Wulff, Haslbeck, Kieslich, Henninger, & Schulte-Mecklenbeck, 2019), real-time measurement of cognition via mouse tracking may foster the understanding of how stereotypes influence source-monitoring processes. Mouse tracking may reveal whether processing and remembering information from different-aged sources elicits conflict when these sources are associated with stereotype-consistent information for some instances and stereotype-*in*consistent for others. Keeping in mind that the biased inference of conforming attitudes and behaviors may promote the persistence or reinforcement of stereotypes, the examination of cognitive dynamics may allow researchers to design effective intervention tools to counteract stereotypical influences on source-monitoring processes in the future. In the following, we will first introduce the theoretical background of stereotype influences on source monitoring and then highlight how mouse tracking can help to get a clearer picture of the processes involved in it. We then propose combining both to study cognitive conflict associated with the reliance on stereotypes during source attributions.

1.1. Source monitoring

The cognitive processes that are at play when attributing information to sources (e.g., who told you something?, where did you read something?) are summarized under the term *source monitoring* (M. K. Johnson, Hashtroudi, & Lindsay, 1993). According to the source-monitoring framework (M. K. Johnson et al., 1993), source attributions can be based on memorized contextual details (e.g., of spatial or temporal nature) or on prior knowledge such as stereotypes (beliefs about personal attributes of social groups; cf. Stroebe & Insko, 1989) and schemas (mental structures that organize and link information based on previous experience; cf. Alba & Hasher, 1983). Diverse studies have demonstrated that source attributions are indeed based on stereotypes/schemas (e.g., age stereotypes: Kuhlmann et al., 2016; gender stereotypes: Marsh, Cook, & Hicks, 2006; social attitudes: Ehrenberg & Klauer, 2005; political attitudes: Klauer & Ehrenberg, 2005; and profession schemas: Bayen, Nakamura, Dupuis, & Yang, 2000). Individuals make use of these stereotypes/schemas to attribute information to sources by eliminating and, thus, reducing the potential candidates for the origin of the information (cf. Marsh et al., 2006). Thinking back to the introductory example of remembering the person who told you "I am discontent with my health", may lead you to guess that it must have been your grandparents rather than your colleague in his 20s. Stereotypical knowledge, such as older people generally talking more about health-related issues (e.g., medication), makes it more likely to attribute the information to the typical and stereotype-consistent source. A common finding is that source attributions are better for stereotype/schema-consistent information (e.g., an older person talking about medicine) than for stereotype/schema-*in*consistent information (e.g., an older person talking about physical activity). The source-attribution benefit for these typical statement-source combinations, however, may be due to better memory or knowledge-based (e.g., stereotype-based) guessing of the source (M. K. Johnson et al., 1993).

To answer the question of whether actual memory or biased guessing is responsible for the benefit in attributing typical statement-source combinations, multinomial processing tree (MPT) models can be used. MPT models separately estimate these memory and guessing processes based on categorical source-attribution data (Batchelder & Riefer, 1990; Bayen, Murnane, & Erdfelder, 1996). These categorical data can be collected in an experimental paradigm (such as the one used by Kuhlmann et al., 2016) in which participants learn everyday statements that are either typical for younger or older adults presented by two person sources. After the study phase, the specific ages (e.g., 23 vs. 70 years old) of the sources are revealed before, in a later memory test, participants attribute the statements to either source or classify them as new (described in more detail in Kuhlmann et al., 2016). An MPT model-based analysis showed that the beneficial performance for typical statement-source combinations was due to stereotype-biased source guessing rather than better source memory (Kuhlmann et al., 2016).[1] That is, if participants did not remember the original source, they were more likely to guess the stereotype-consistent one. For an overview of experimental evidence on stereotype and schema-based typicality effects in source monitoring refer to Kuhlmann and Bayen (2016). Most of the few existing studies point in the direction of source guessing being characterized by more systematic/controlled processing (e.g., defined as rather slow, deliberate, resource-dependent; Bargh, 1994) than heuristic/automatic (e.g., defined as rather fast, unintentional, resource-independent; Bargh, 1994) use of prior knowledge (e.g., Bayen & Kuhlmann, 2011; Ehrenberg & Klauer, 2005; Spaniol & Bayen, 2002; but see also Ehrenberg & Klauer, 2005; for a theoretical overview, see M. K. Johnson et al., 1993; M. K. Johnson & Raye, 2000). For instance, Spaniol and Bayen (2002) studied the nature of source-monitoring processes focusing on the time course of item memory and schema-based source guessing using a response-signal method. The authors found that a schema bias in source guessing emerged after the onset of item memory, suggesting that the influence of knowledge takes more time to evolve and, thus, relies more strongly on systematic/controlled processing.

1.2. Mouse tracking

One approach to study the nature of cognitive processes in real time is mouse tracking, an increasingly popular process-tracing method.[2] Mouse tracking continuously maps dynamic motor responses to task-related cognitive processing (Spivey & Dale, 2006) which has been referred to as "hand in motion reveals mind in motion" (Freeman, Dale, & Farmer, 2011, *p.* 1; but see Wulff et al., 2019). Mouse tracking seeks to quantify cognitive conflict arising from *in*consistent information. This conflict or, put differently, the co-activation of the unchosen response option, is assumed to be reflected in the spatial deviation of

[1] Schemas/stereotypes can also influence source memory. If an item is very *a*typical for the source it was presented with, then attention is drawn to the item-source combination due to violation of expectations. As a result, source memory may be better for *a*typical combinations (*inconsistency effect*, e.g., Bell, Buchner, Kroneisen, & Giang, 2012; Ehrenberg & Klauer, 2005; Kroneisen & Bell, 2013), whereas the source-guessing bias nevertheless emerges.

[2] Just to name a few examples of research areas mouse tracking has been successfully applied to: *language & numerical processing*: Dale, Kehoe, & Spivey, 2007; Farmer, Cargill, Hindy, Dale, & Spivey, 2007; Faulkenberry, 2014; *social cognition*: Cloutier, Freeman, & Ambady, 2014; Freeman & Ambady, 2009, 2011, 2014; Freeman, Ambady, Rule, & Johnson, 2008; Freeman, Pauker, Apfelbaum, & Ambady, 2010; K. L. Johnson, Freeman, & Pauker, 2012; *decision making*: Dshemuchadse, Scherbaum, & Goschke, 2013; Szaszi, Palfi, Szollosi, Kieslich, & Aczel, 2018; Tabatabaeian, Dale, & Duran, 2015; *reasoning*: McKinstry, Dale, & Spivey, 2008; *perception and attention*: Huette & McMurray, 2010; *memory*: Abney, McBride, Conte, & Vinson, 2015; Koop & Criss, 2016; Papesh & Goldinger, 2012).

mouse trajectories from an idealized straight line (Stillman, Shen, & Ferguson, 2018). For instance, in a study on gender categorization by Freeman and Ambady (2009), participants attributed male and female faces (e.g., typical-male faces vs. feminized, and therefore *a*typical, male faces) to feminine and masculine-stereotype labels (e.g., caring vs. aggressive). In sex-*a*typical trials (e.g., feminized male faces) mouse trajectories were spatially more biased (curved) towards the opposite-gender stereotype label (e.g., caring) compared to sex-typical trials revealing a parallel activation of the stereotype knowledge related to the sexes. This real-time dynamic of social categorization has been reported for explicit stereotypical attitudes (e.g., Cloutier et al., 2014; Freeman & Ambady, 2011; Freeman et al., 2010; Freeman, Pauker, & Sanchez, 2016; Wojnowicz, Ferguson, Dale, & Spivey, 2009), also varying as a function of personally-held prejudices (Cassidy, Sprout, Freeman, & Krendl, 2017), and implicit attitudes (e.g., Yu, Wang, Wang, & Bastin, 2012). Based on stereotype-induced cognitive conflict, mouse movements also have been used to predict actual consequential behavior (e.g., voting behavior: Hehman, Carpinella, Johnson, Leitner, & Freeman, 2014; trust: Freeman et al., 2016).

2. Overview of the current experiment and research questions

As outlined above, the influence of prior knowledge on cognitive processes in source monitoring is one approach to study stereotyping on a memory-based level. Starting from the assumption that the influence of age stereotypes on source monitoring may be disclosed through cognitive conflict that is tracked by mouse movements, we conducted an experiment including a source-monitoring task with age-stereotypical item material. The real-time recording of cursor movements opens up the possibility to inspect the spatial attraction towards the unchosen source. Additionally, we model the relationship of mouse trajectories and the cognitive processes involved in source attributions. **Hypothesis 1.** As described earlier, Freeman and Ambady (2009) reported that individuals considered both gender-stereotypical and *a*typical knowledge while categorizing faces to character traits as indicated by more curved mouse trajectories towards the opposing sex when processing *in*consistent information. In line with Freeman and Ambady (2009), we expect to find a comparable result transferred to our experimental paradigm using a source-monitoring task based on age stereotypes. First, we hypothesize that the cognitive conflict when processing stereotype-*in*consistent statement-source combinations also maps onto mouse movements. That is, when indicating which source presented a statement, the experienced cognitive conflict and, thus, the curvature of mouse trajectories should be more pronounced if the age depicted in the item material is *in*consistent with the originally presenting source's age. We assume that this increased cognitive conflict stems from the simultaneous activation of stereotypical knowledge tied to the age categories.

Hypothesis 2. Current evidence points towards source guessing using prior knowledge under uncertainty in a systematic/controlled way (e.g., Spaniol & Bayen, 2002). Therefore, we predict that a stronger influence of stereotypes on source guessing reflects a more extensive integration of underlying knowledge in source attributions and should lead to more experienced conflict when faced with stereotype-consistent and -*in*consistent information. Due to this assumed cognitive conflict, individuals with stronger stereotype-biased source guessing should show an increased spatial attraction towards the stereotype-consistent source. Applying a Bayesian-hierarchical MPT model (Klauer, 2010), we, thus, cautiously predict that biased guessing should be positively correlated with curved mouse movements but are aware of the exploratory character of this research question when interpreting the results.

3. Methods

3.1. Materials

For the source-monitoring task, we chose age-stereotypical statements from a previous survey study by Kuhlmann, Kornadt, Bayen, Meuser, and Wulff (2017), in which the authors examined the multidimensionality of age stereotypes held by younger and older adults based on typicality ratings for these statements. The latter were generated based on life domains that reflect different behaviors and attitudes (e.g., “I am financially independent” reflecting the life domain “finances” and the adjective dimension “autonomy”). Given that we only tested young participants, we used statements that were rated as typical for either young or old people by the 69 younger adults from the survey study. We defined a statement as “typical” when the mean statement rating (on a scale from 1 = *very atypical* to 5 = *very typical*) was > 3.3 for one (target) age group (e.g., young person) and, at the same time, < 2.8 for the other (target) age group (e.g., old person). For more information on the procedure and the creation of these typicality ratings, see Kuhlmann et al. (2017). Applying this cut-off criterion, the remaining item set consisted of 118 statements, from which we randomly selected 90 statements. Sixty served as to-be-learned in the study phase and 30 as distractors in the test phase. Out of these 90 statements, we created three lists - comparable in their mean ratings for both (target) age groups (pairwise comparisons of the latter for each of the three lists; all p-values > .11). Each list consisted of 30 statements, 15 statements reflecting typically-young behavior and 15 statements reflecting typically-old behavior. For each participant, two lists served as to-be-learned in the study phase (one for each source), the remaining list served as distractors in the test phase. Analogous to Kuhlmann et al. (2016), “Christian” and “Michael,” two middle-aged German first names (norms retrieved from Rudolph, Böhm, & Lummer, 2007), served as source labels.

3.2. Design and procedure

The experimental design was a 2 (source age: young vs. old) × 2 (age stereotypicality: typicality reflected in statement consistent vs. *in*consistent with respective source age) within-subjects design. Source label (Christian vs. Michael), source age (23 vs. 70 years old), color of the source label (yellow vs. green), position of the yes-no response and the sources in the test phase (left vs. right), and assignment of the three statement lists as two study and one distractor list(s) were counterbalanced between subjects. The experiment was programmed with OpenSesame (Mathôt, Schreij, & Theeuwes, 2012) and run on computers with a screen resolution of 1920 × 1080 pixels. We tracked participant's mouse movements using the open-source mouse-tracking package *mousetrap* added to OpenSesame as plug-in (for a detailed overview of the *mousetrap* package, see Kieslich & Henninger, 2017). The mouse set-up was adjusted to default settings on each computer (acceleration turned on, medium speed). Position of the mouse cursor was recorded every 10 ms. Participants were tested individually in groups up to 10, and initially provided informed consent before computerized instructions explicitly informed them about a subsequent memory test.[3] In the study phase, we presented the 60 statements, 30 typical for a young person (e.g., “I rarely get sick”), and 30 typical for an old person (e.g., “I go to church every Sunday”). Half of the typically-young and typically-old statements were presented with Christian as source and the other half with Michael, respectively (= zero contingency between sources and statement typicality). After studying, instructions for the source-memory test followed and participants

[3] We did not provide participants with an intentional source-memory instruction because this has been shown to reduce guessing bias (e.g., Kuhlmann, Vaterrodt, & Bayen, 2012).

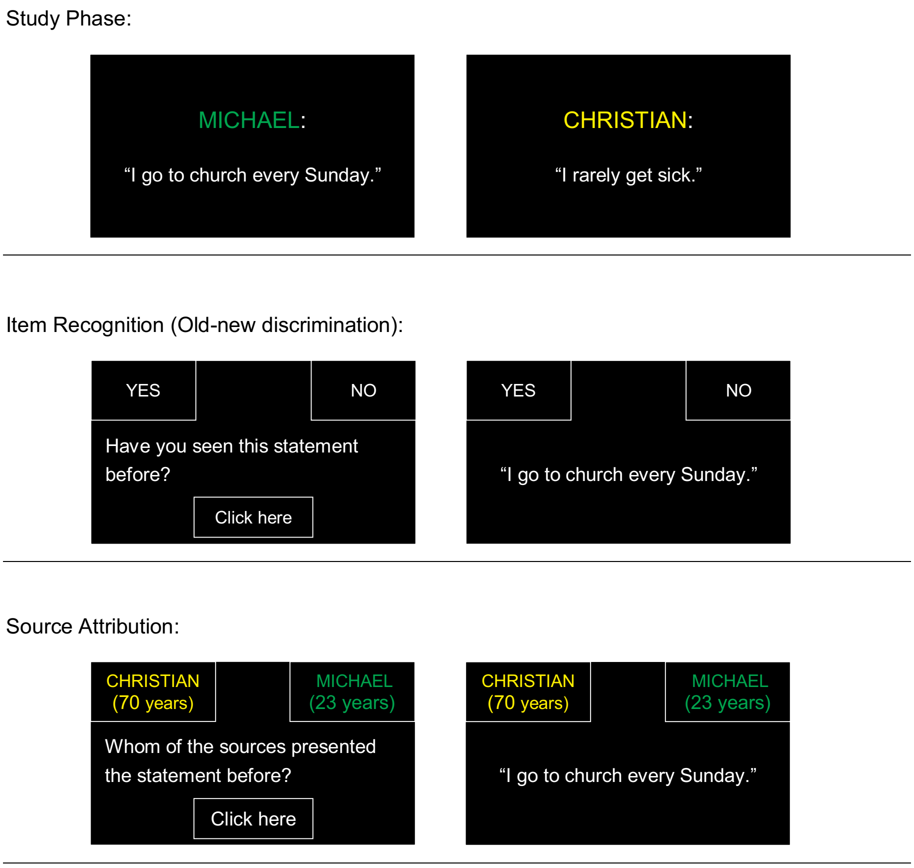

Fig. 1. Example visualizations of the source-monitoring task set-up. On the top, an example of the two types of sources in the study phase is included. In the second row, an example of the recognition stage is presented. The bottom row shows exemplar screens of the source-attribution stage. We included three practice statements at the beginning and the end of the study phase to control for primacy (Anderson & Barrios, 1961) and recency effects (Greene, 1986). Statements appeared on the screen for 4000 ms with an inter-stimulus interval of 250 ms.

learned the specific source age (young: 23 years/old: 70 years). We used a two-stage response format and presented statements blocked for each stage. In a recognition stage, participants first had to decide whether a statement had been presented before or not. The response options "YES" (indicating that a statement was old) and "NO" (indicating that a statement was new) were constantly shown on the top left and right corners of the screen. Thus, participants could reach them without overshooting the option. Responses were indicated via mouse click on one of the two options. In each trial, participants first had to click on the question "Have you seen this statement before?" in the bottom of the screen, then the statement appeared. The reset of the mouse to a starting position after each trial ensured that the cursor had the same distance from both response options in each trial.

In a second stage, we informed participants that they now had to remember the source of each statement they had classified with "YES" before. Participants initially clicked on the question "Whom of the sources presented the statement before?" to see the statement and indicated their source attribution immediately. The response options "CHRISTIAN" and "MICHAEL," and their respective age were again constantly shown in the upper corners of the screen. A graphical illustration of the procedure is displayed in Fig. 1. We presented statements in randomized order. The number of statements recognized as "old" thereby predefined the trial number of the source-attribution stage. Participants were instructed to react fast and did not receive error feedback. They completed the entire source-monitoring task approximately within 30 min, filled out a demographic questionnaire and were debriefed. In exchange for their participation, participants received either course credit or a monetary compensation.

3.3. Measures and analyses

3.3.1. Mouse tracking

Mouse-tracking data were preprocessed as recommended by Kieslich and Henninger (2017) and Kieslich et al. (2019). We computed the *Maximum Absolute Deviation* (MAD; e.g., Freeman & Ambady, 2010) of all trajectories, which is the maximum distance between the observed trajectories and the idealized straight line from the start button to the response option. The MAD is used as an index of the spatial attraction towards the unchosen response option (Kieslich et al., 2019), and is assumed to indicate the difference in activation of competing response alternatives (e.g., Spivey, Dale, Knoblich, & Grosjean, 2010). In this interpretation, large MAD (i.e., strong curvature) would indicate that both response alternatives were substantially co-activated and, therefore, cognitive conflict was present.

For all mouse-tracking analyses, we compared trials in which the age of the originally presenting source was consistent or *in*consistent with the age reflected in the statement.[4] We further took the correctness of the source attribution (correct vs. incorrect attribution of the statement to either source) into account.[5] In an exploratory manner, we tested whether participants' mouse trajectories were generally more or less biased in correct and incorrect source attributions and whether both factors (consistency and correctness of response) would interact. As there were multiple observations per participant and per statement

[4] We removed distractor statements that were incorrectly recognized as old (corresponded to 9.60% of all trials) from these analyses because these statements were not presented by any source and, thus, have systematically missing values on the consistency factor.

[5] Even though it is common practice in the literature to only analyze correct trials as it simplifies the interpretation of curved mouse movements (i.e., attraction towards the "typical" category; e.g., Dale et al., 2007; Dignath, Pfister, Eder, Kiesel, & Kunde, 2014; Freeman & Ambady, 2011), we decided to not discard incorrect trials for the following reasons: Source monitoring is a complex memory task, we, therefore, expect a considerable number of incorrect source attributions compared to, for instance, semantic or facial categorizations.

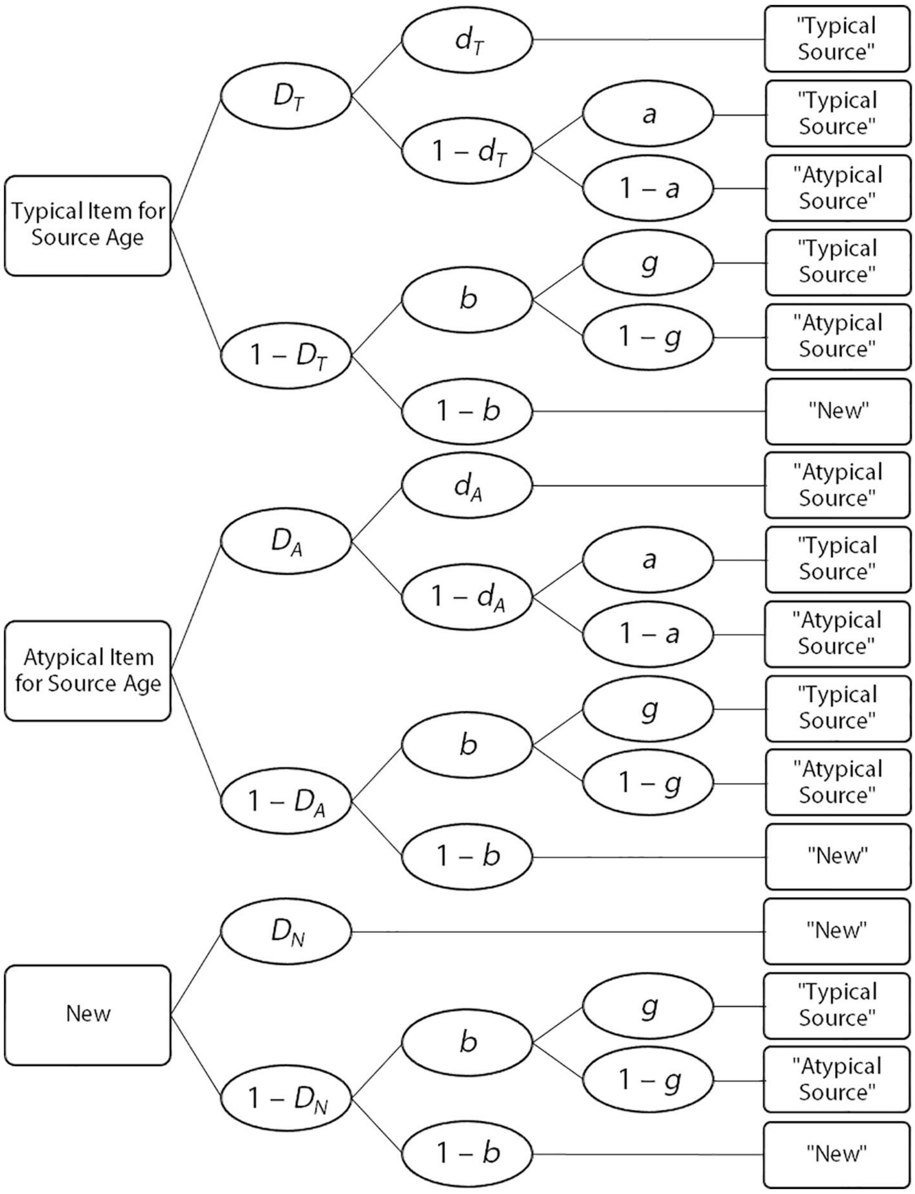

Fig. 2. Two-high-threshold multinomial model of source monitoring. D_T = probability of recognizing a statement that had been presented by the typical, stereotype-consistent source; D_A = probability of recognizing a statement that had been presented by the *a*typical, stereotype-*in*consistent source; D_N = probability of knowing a statement is new; d_T = probability of correctly remembering the source of a statement that had been presented by the typical source; d_A = probability of correctly remembering the source of a statement that had been presented by the *a*typical source; b = probability of guessing that an unrecognized statement is old; g/a = probability of guessing that a (un)recognized statement had been presented by the typical source. Source: Adapted from Bayen et al. (1996).

from the source phase for each of the dependent variables in both types of observations (consistent vs. *in*consistent statement-source combinations and correct vs. incorrect responses), we used linear mixed models for analyses with consistency, correctness of the response, and their interaction as predictors.[6]

3.3.2. Multinomial modeling of source monitoring

We applied the *Two-high-threshold multinomial model of source monitoring* (2HTSM) from Bayen et al. (1996) to disentangle memory and guessing parameters. For a visualization of the model-tree structure adapted to our study, see Fig. 2. The most parsimonious identifiable 2HTSM submodel (for an overview of all possible submodels, see Bayen et al., 1996) explains the observed categorical responses with only four latent parameters that reflect distinct cognitive processes. Parameter D (= item memory) measures the probability of recognizing a statement as old or as new, knowing that the presented statement is a distractor. With the complementary probability ($1 - D$), the statement is not recognized as old or new. In this case, the status of a statement is guessed to be either old (parameter b) or new ($1 - b$) because the source can also not be remembered without item memory (e.g., Malejka & Bröder, 2016). Parameter d (= source memory) measures the probability of remembering the source of a statement that has been recognized as old. If the source is not remembered ($1 - d$) or the statement has been guessed to be old (b), the status of the source is guessed to be either typical (g) or *a*typical ($1 - g$) for the respective statement-source combination. Summarizing, item memory is assumed to be equal for statements that are presented with the age-consistent source (i.e., aggregate of typically-young statements presented by the "young" source & typically-old statements presented by the "old" source), with the age-

[6] Both, consistency and correctness, were effect-coded with +1/−1 (+1: correct/consistent; −1: incorrect/*in*consistent) in all linear mixed model analyses. To account for inter-trial dependencies, we included random intercepts for participants as well as random intercepts for statements. To determine whether the addition of random slopes improved model fit, we implemented the maximum random-effects structure for each model first and then reduced the complexity of the model by removing the random slopes with the least variance until the model converged. We then tested whether including the random slopes improved model fit with Likelihood Ratio Tests for nested models. For the models reported here, including random slopes did not improve model fit. More details on the used procedure can be found in the uploaded analyses scripts: https://osf.io/85936/?view_only=214fa149e30748cbbd07852c43fab419.

*in*consistent source (i.e., aggregate of typically-old statements presented by the "young" source & typically-young statements presented by the "old" source) and those that are new (aggregate of typically-young and -old statement distractors). Source memory for age-consistent and age-*in*consistent statements, as well as source guessing for remembered and unremembered statements, is assumed to be equal, too ($D_T = D_A = D_N$; $d_T = d_A$; $g = a$). This submodel has been used elsewhere with similar experimental paradigms (e.g., Kuhlmann et al., 2016; Spaniol & Bayen, 2002) and for analyses aggregated across item types (e.g., Ehrenberg & Klauer, 2005; Kuhlmann et al., 2016; Schaper, Kuhlmann, & Bayen, 2019; Spaniol & Bayen, 2002).

To account for individual differences, we applied the latent-trait MPT model from Klauer (2010) - a Bayesian-hierarchical approach of multinomial modeling.[7] This approach treats individual parameters as random effects. The separate source-monitoring parameter estimates per individual are constrained by population-level parameters assumed to follow a multivariate normal distribution of probit-transformed parameters as prior distribution with a mean and covariance matrix to be estimated from the data (a conceptual illustration of the general model structure is displayed in Appendix A but see Klauer, 2010, for a theoretical foundation and statistical details).[8] Based on this multivariate prior, the latent-trait approach considers individual parameters and their correlations jointly in one model. Additionally, correlations with continuous, external predictors, such as mouse-tracking indices, can be easily modeled. Based on the Markov-chain Monte Carlo (MCMC) algorithm, samples from the posterior distribution of parameters can be drawn. The latent-trait MPT model with the submodel described above applied to our data showed a good model fit (as indicated by non-significant Bayesian posterior predictive *p*-values > .05; see Appendix B for model-fit details) and outperformed a more differentiated model without aggregating across statement types (see also Appendix B for a model comparison and model-based results for the more differentiated model).

3.4. Participants

Determining effect sizes and appropriate power analyses for linear mixed models is not trivial (Brysbaert & Stevens, 2018). Thus, we opted to determine the required sample size for our first hypothesis by considering a roughly equivalent, non-trial based analysis, namely a paired *t*-test for the effect of statement-source consistency on MAD. Compared to Freeman and Ambady (2009), who found an effect of (stereo-)typicality on mouse trajectories of $d_z = 0.71$, we proposed a more conservative effect size of $d_z = 0.50$, due to the complexity of the item material. Thus, when setting equal Type I and Type II error rates of $\alpha = \beta = .05$, the required sample size is $N = 45$ (as determined with G*Power; Faul, Erdfelder, Lang, & Buchner, 2007). For our second hypothesis, a rough equivalent (to the correlation in the hierarchical Bayesian MPT model) is a frequentist (Pearson) correlation analysis. If we assume a moderate true correlation of $r = .3$ ($\hat{\rho}$ in the Bayesian framework) for the average MAD and source guessing with $\alpha = .05$ and a power = .80, the required sample size is $N = 64$.[9] We decided to orient our aspired sample size to the required larger *N*. Due to time constraints, we were only able to collect data from a total of 60 undergraduate students (41 females, $M_{age} = 21.62$ years, age range: 18–28). The obtained sample size yielded high statistical power ($1 - \beta = .98$) to detect a medium effect of $d_z = 0.50$ in the paired *t*-test and high statistical power ($1 - \beta = .78$) to detect a moderate correlation of source-monitoring processes and covariate(s).

4. Results

All analyses were conducted using *R* (R Core Team, 2018). The mouse tracking related analyses were done using the *mousetrap* package (Kieslich et al., 2019; Wulff et al., 2019). Linear mixed model analyses were conducted with the *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). Posthoc tests for the mixed models were conducted using the *emmeans* (Lenth, 2019). The hierarchical MPT model analyses were conducted with the *TreeBUGS* package (Heck, Arnold, & Arnold, 2018). All plots were based on *ggplot2* (Wickham, 2016) as well as on the *mousetrap* (Kieslich & Henninger, 2017) and *afex* package (Singmann, Bolker, Westfall, & Aust, 2018).

4.1. Do age stereotypes elicit decision uncertainty in source attributions?

On average, participants' accuracy for the old-new discrimination was high. They classified 81.11% (*SD* = 6.66) of the statements correctly as old or new and attributed 61.38% (*SD* = 16.79) of the statements to the correct source.

We ran a linear mixed model to test our first hypothesis and expected to find a significant, negative regression coefficient for the predictor consistency. For an overview of the linear mixed model results, see Table 1. Contrary to our prediction, consistency did not show a significant effect on MAD. However, there was a significant interaction of the factors correctness and consistency, b = -25.95, $t(2634.83) = -3.34$, $p < .001$ (see Fig. 3). To further investigate this interaction effect, we ran post-hoc pairwise comparisons with Bonferroni-Holm corrected *p*-values (tested two-sided). MAD for consistent statement-source combinations were significantly smaller for correct than for incorrect source attributions, indicating less experienced conflict, $t(2631) = -3.57$, $p = .002$. Furthermore, consistent and *in*consistent statement-source combinations differed significantly when the source attribution was correct, such that trajectories were more curved for *in*consistent than for consistent statement-source combinations, $t(2578) = -3.80$, $p = .001$ (please refer to Fig. 4 for a schematic breakdown of the interaction pattern).

4.2. Does stereotype consistency influence the shape of mouse trajectories?

When using MAD as a dependent variable for mouse tracking, the information from the trajectory is condensed into one single index. Parts of the information, such as the concrete shape of the trajectory, are, thus, lost in this type of analysis. In the past, one way to deal with this property of MAD was to look at their distribution - if the MAD distribution revealed bimodality, it would be interpreted as evidence

[7] Traditional MPT models are estimated on the group level, aggregating the categorical response data across individuals and items, using Maximum likelihood (e.g., Hu & Batchelder, 1994). Aggregating data always bears the risk of neglecting individual differences or dependencies between cognitive processes (Batchelder & Riefer, 1999; Lee, 2011; Matzke, Dolan, Batchelder, & Wagenmakers, 2015). This inherent assumption of homogeneity can lead to misspecified MPT models and biased parameter estimates (Klauer, 2010; Matzke et al., 2015; Smith & Batchelder, 2010).

[8] We specified weakly informative prior distributions (following Klauer, 2010; Matzke et al., 2015; see Heck, Erdfelder, & Kieslich, 2018, for further details) which are implemented as default settings in *TreeBUGS* for the group-level means (prior: standard normal distributions) and covariance matrix (prior: scaled inverse-Wishart uniform distribution) that were updated by the incorporated data. The algorithm cycled through three MCMC chains with 20,000 iterations each until all parameters reached the desired convergence-fit criterion as indicated by $\hat{R} < 1.05$ (Gelman & Rubin, 1992) plus additional 20,000 iterations (to verify convergence stability).

[9] We reduced the aspired level of power for the second analysis compared to the first for the following reason: Due to monetary constraints, we did not have enough funds to collect data from $N = 111$ participants, the required sample size when setting the aspired power level to .95. As the correlational analysis with the MPT parameters was exploratory in its nature, we decided to be less conservative with regard to power and reduced the aspired power level to .80.

Table 1
Linear mixed model with maximum absolute deviations as dependent variable.

Predictors	*b*	*se*	*t*	*df*	*p*
Intercept	189.21	18.11	10.45	74.28	< .001
Correctness	−15.47	7.95	−1.95	2649.82	.052
Consistency	−6.16	7.53	−0.82	2552.01	.414
Correctness × Consistency	−25.95	7.77	−3.34	2634.83	< .001

Note. Linear mixed model results for Maximum Absolute Deviations (MAD). Correctness (of source attribution) and consistency (of statement-source combination) were both effect-coded with +1 (correct/consistent)/−1 (incorrect/*in*consistent). *b* = beta-weight of effect, *se* = standard error, *t* = *t*-value, *df* = degrees of freedom, *p* = *p*-value. To account for inter-trial dependencies, random intercepts for participants as well as statements were included.

that the trajectory shape was not homogeneous (Wulff et al., 2019). Therefore, before running our planned analyses, we conducted exploratory bimodality analyses of the trajectories overall as well as for the different levels of the factors correctness and consistency. Indeed, the distribution of mouse trajectories revealed substantial bimodality according to the bimodality coefficient (*BC* > 0.75 overall and in each group, SAS Institute, 1989) as well as the Hartigan's dip statistic (*ps* < .002 overall and in each group, Hartigan & Hartigan, 1985). That is, participants showed straight or slightly curved trajectories in some trials and strongly curved or change-of-mind trajectories in others (see Fig. 5 for the distribution of raw, time-normalized trajectories).

The non-homogeneous trajectory-shape distribution indicated that different trajectory types might be present in our data set. Using MAD as a dependent variable might conceal the variation of different trajectory types due to our predictors source-response correctness and statement-source consistency (e.g., different trajectory types can have the same MAD value; c.f., Wulff et al., 2019). Further, when analyzing MAD we cannot make any inferences with regard to how exactly the source response unfolded over the course of trials as the whole trajectory was reduced into one single index. By looking more closely at the trajectory shapes, we also hope to get a clearer picture of the temporal dynamics (Dale & Duran, 2011; Hehman, Stolier, & Freeman, 2015) of the source-attribution process. Therefore, we conducted an exploratory analysis of the distribution of the prototypical mouse trajectory shapes

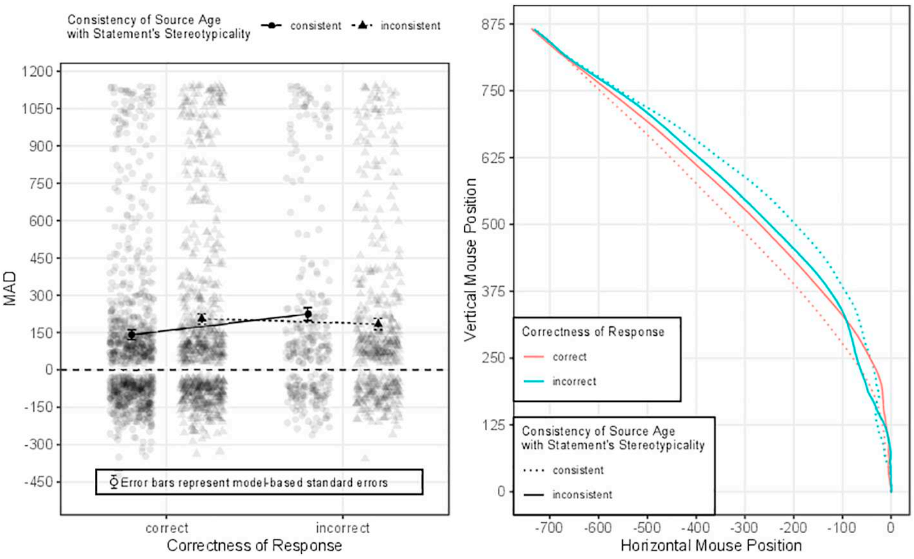

Fig. 3. Left panel: Maximum Absolute Deviation (MAD) as a function of the two factors correctness (of source attribution) and consistency (of statement-source combination). Positive values of MAD indicate spatial attraction towards the unchosen source, 0 (dashed line) indicates the idealized straight line towards the chosen source, negative values indicate curved trajectories towards the chosen source. Mean values per factor combination and corresponding error bars representing standard errors are displayed in black, individual values per participant are displayed in grey (points = consistent statement-source combinations, triangle = *in*consistent statement-source combinations). Right panel: Temporal dynamic of mouse trajectories based on the horizontal and vertical mouse position (in pixels) for correct (orange) and incorrect (blue) source attributions separate for consistent statement-source combinations (dashed line) and *in*consistent statement-source combinations (solid line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

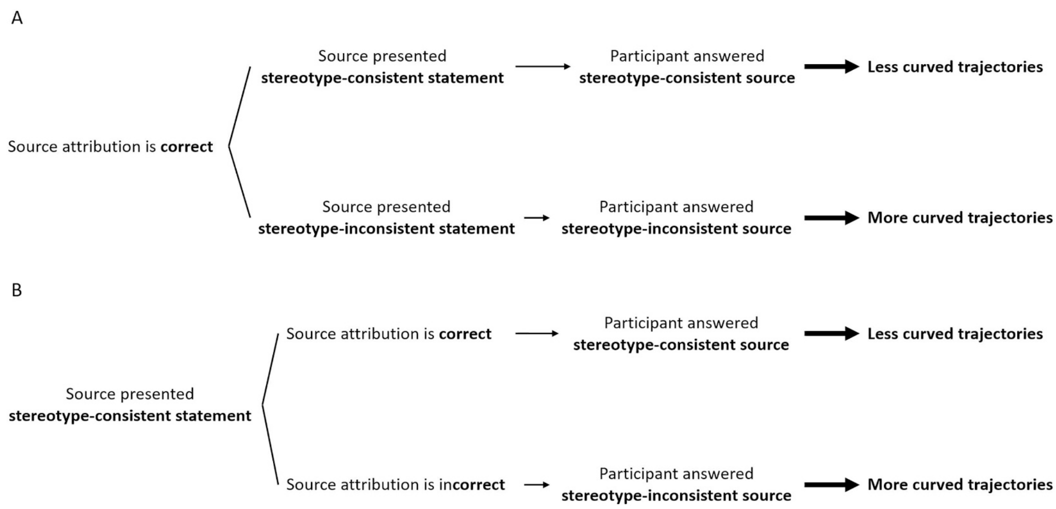

Fig. 4. Breakdown of the interaction pattern of statement-source consistency and source-attribution correctness. Panel A represents the first significant, post-hoc group comparison. In this case, trajectories were more curved when the correct source attribution was stereotype-*in*consistent with the presenting source compared to when it was stereotype-consistent with the presenting source. Panel B represents the second significant, post-hoc group comparison. Here, trajectories were more curved when participants answered the incorrect, stereotype-*in*consistent source compared to when they answered the correct, stereotype-consistent source.

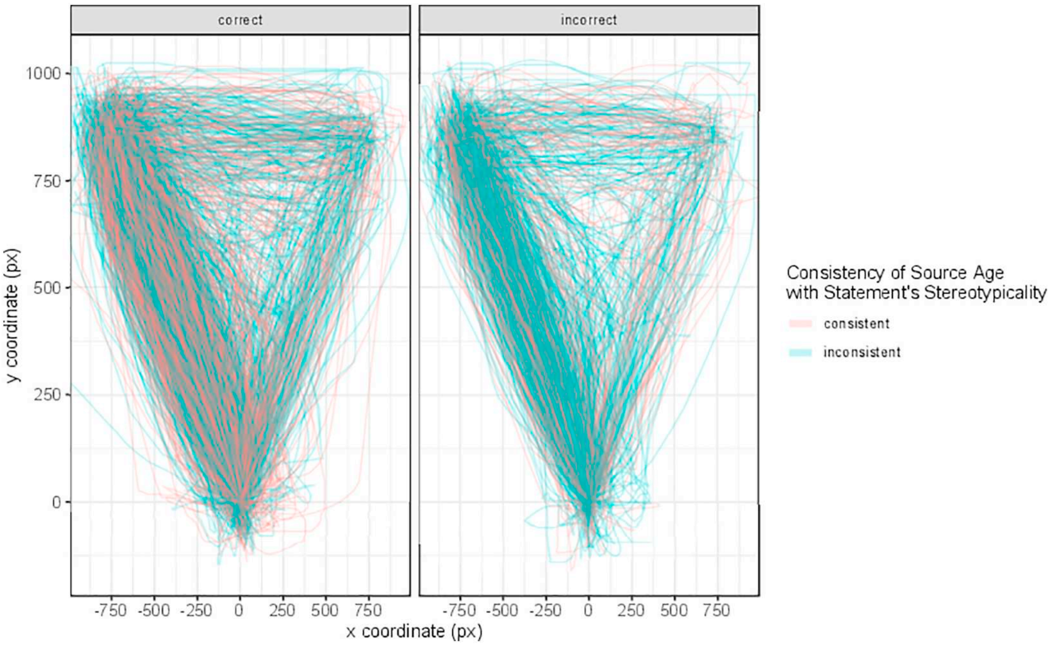

Fig. 5. Time-normalized trajectories for correct (left panel)/incorrect (right panel) source attributions and consistent (orange)/*in*consistent (blue) statement-source combinations. x coordinate (px) = position of the cursor in pixels on x-coordinate; y coordinate (px) = position of the cursor in pixels on y-coordinate.

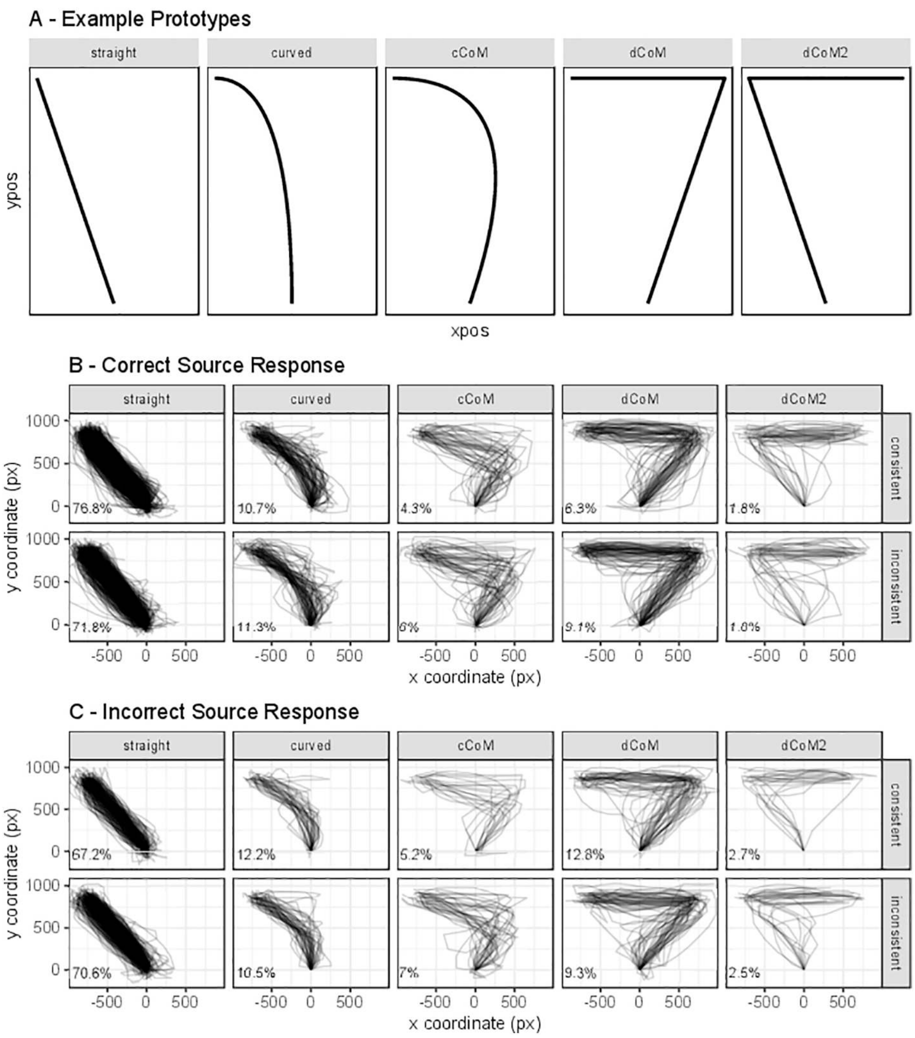

Fig. 6. Exemplar ordinal prototype cluster which trajectories can be classified into (Panel A). Trajectories are sorted according to their curvature from straight trajectories to double change-of-mind trajectories and separately shown for correct (Panel B) and incorrect (Panel C) source attributions of consistent and *in*consistent statement-source combinations. The relative frequency of each prototype (within the respective design cells) is depicted in the bottom left corner of each cell in Panel B and C. x coordinate (px) = position of the cursor in pixels on x-coordinate; y coordinate (px) = position of the cursor in pixels on y-coordinate; straight = most direct trajectory from the start button to the response option, curved = medium curved trajectory, cCoM = continuous change of mind, dCoM = discrete change of mind, dCoM2 = double discrete change of mind.

Table 2
Mixed ordinal regression for mouse trajectory prototypes.

Predictors	*b*	*se*	*z*	*p*
Correctness	−0.05	0.05	−0.95	.342
Consistency	−0.05	0.05	−0.95	.344
Correctness × Consistency	−0.15	0.05	−2.90	.004

Note. Mixed ordinal regression results for mouse trajectory prototypes. Correctness (of source attribution) and consistency (of statement-source combination) were both effect-coded with +1 (correct/consistent)/−1 (incorrect/*in*consistent). b = beta-weight of effect, se = standard error, z = z-value, p = p-value. To account for inter-trial dependencies, random intercepts for participants as well as statements were included.

to test whether our MAD results were an artifact of condensing the trajectories and to get a better picture of the underlying cognitive dynamics.

We followed the procedure described by Wulff et al. (2019) and classified trajectories into five prototypes: straight, curved, continuous change of mind, discrete change of mind, and double discrete change of mind (for a visualization of the prototypes and the clustered trajectories see Fig. 6). When looking at the distribution of trajectory types for correct and incorrect source attributions, one can see that there are more discrete change-of-mind trajectories for *in*consistent than consistent correct source attributions. For incorrect source attributions, one can see that there are more straight trajectories when the original statement-source combination was *in*consistent than when it was consistent (i.e., trajectories were more often classified as straight when participants responded incorrectly but stereotype-consistent than when they responded incorrectly and stereotype-*in*consistent). In order to analyze whether the probability of showing a more curved trajectory prototype changes with correctness of the source attribution and the consistency of the statement-source combination, we ran an ordinal mixed regression with the trial-based prototype categories as dependent variable. Based on the results from the mixed model with MAD as dependent variable, we expected to find the same interaction pattern of consistency and response correctness. The results showed a significant interaction of both predictors, $b = -0.15$, $z = -2.90$, $p = .004$ (see also Table 2 and Fig. 6). Again, we ran Bonferroni-Holm adjusted pairwise comparisons (tested two-sided) for each level of the predictors. The results mirrored the results of the post-hoc analyses for MAD as dependent variable: for consistent statement-source combinations, the odds of conforming with a more curved prototype category were higher for incorrect compared to correct source attributions, $z = -2.61$, $p = .045$. Similarly, within correct source attributions, the odds were higher for *in*consistent statement-source combinations to conform with a more curved prototype category than for consistent statement-source combinations, $z = -3.36$, $p = .005$.

4.3. Do individuals guess based on (age) stereotypes when their source memory fails?

To study source monitoring on a process level in addition to the recording of cursor movements, the model-based analyses are reported next. In the following, we provide the means and corresponding 95%-Bayesian Credibility Intervals (BCIs) of the posterior distribution for parameters and their correlations. On the group level, participants' item memory (D) was .65 [.61, .68]. If they recognized a statement as old, they were able to remember the source of a statement (d) with a probability of .37 [.24, .51]. If they did not recognize the status (old or new) of a statement, participants guessed the statement to be old (b) with a probability of .40 [.35, .45]. Given that they had either guessed or remembered a statement to be old (without remembering the source), participants guessed (g) the stereotype-consistent source (age reflected in the statement corresponded to the source's age) more often with a probability of .59 [.54, .63] (chance level: 50%). To test in an

Table 3
Model-based estimates of source monitoring parameter correlations.

Parameter	*D*	*d*	*b*	*g*
D	—	–	–	–
d	**.59 [.26, .84]**	–	–	–
b	−.33 [−.63, .03]	−.29 [−.62, .07]	–	–
g	−.25 [−.58, .12]	**−.52 [−.78, −.19]**	.11 [−.26, .47]	–

Note. Estimates of the Bayesian-hierarchical MPT model for correlations between parameters. D = probability of recognizing a statement as previously presented and probability of knowing that a distractor statement is new; d = probability of correctly remembering the (either stereotype-consistent or *in*consistent) source of a statement; b = probability of guessing that an unrecognized statement is old; g = probability of guessing that a statement had been presented by the stereotype-consistent source when the source is not remembered. Brackets indicate 95%-Bayesian Credibility Intervals (BCI) and substantial correlations (BCI excludes 0) are marked in bold.

exploratory manner whether it is indeed legitimate to argue in favor of an age-stereotypical source-guessing bias here, we sampled the posterior distribution for the difference in source guessing and chance-level guessing. On the group-level, source guessing was stereotype-biased, $\Delta(g - .50) = .09$ [.04, .13] to a substantial extent (as the BCI excluded 0). Exploratory parameter correlations further revealed that source memory and source guessing were linked (see Table 3 for correlations). Participants compensated for poor source memory (d) by guessing more stereotype-biased, $\hat{\rho} = -.52$ [−.78, −.19]. As the credibility intervals for the parameter correlations were rather large, we conducted a parameter-recovery simulation.[10] Results and further information on how we conducted the simulation are reported in Table C1 in Appendix C. The simulation demonstrated that parameter means and correlations were mostly recovered, which was also true for the negative correlation of source memory and guessing obtained in our sample. We, thus, replicated prior findings on source memory as a determinant of the strength of the source-guessing bias (e.g., Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013; Kuhlmann et al., 2016) which is also in line with theoretical assumption that successful source monitoring essentially relies on the quality of available (encoded) information in memory and the decision process (guessing) when making source attributions (M. K. Johnson et al., 1993).

4.4. Do source-monitoring processes map onto mouse trajectories?

To test whether individuals who show a strong source-guessing bias generally experience greater cognitive conflict during source attributions when processing stereotype-consistent and -*in*consistent statements-source combinations, we included MAD (mean-aggregated across trials per participant) as a continuous, external covariate in the hierarchical MPT model. In a first step, the correlations of MAD with the posterior values of all individual MPT parameters were computed as default in *TreeBUGS* (Heck, Arnold, & Arnold, 2018). The correlation computation was repeated for all posterior samples, thereby accounting for uncertainty in estimating the sample correlation due to parameter estimation (see Heck et al., 2018, for details in *TreeBUGS*). Our second hypothesis pertained to the relationship of the source-guessing parameter g and MAD: Source guessing was positively correlated with MAD, $\hat{\rho} = .21$ [.05, .35]. Individuals who were more likely to attribute statements to the stereotype-consistent source (and, thus, also showed more incorrect source attributions) experienced greater conflict during their source attribution (see Fig. 7). In addition, the exploratory inspection of other source monitoring and MAD correlations revealed that source memory was linked to mouse-trajectory deviations towards the

[10] We thank an anonymous reviewer for his/her suggestion to conduct a parameter-recovery simulation.

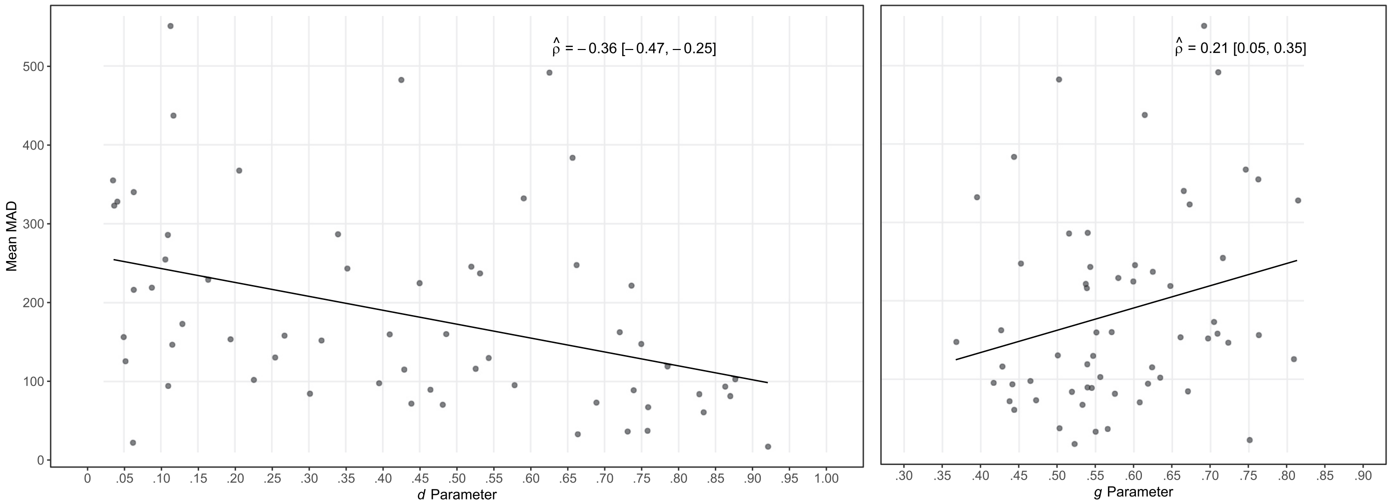

Fig. 7. Sampled population correlation coefficients $\hat{\rho}$ between the source-memory parameter (*d*, on the left; can vary between 0 and 1) and mean-aggregated Maximum Absolute Deviation (MAD) as well as the source-guessing parameter (*g*, on the right; can vary between 0 to 1) and MAD based on the Bayesian-hierarchical latent-trait MPT model (Klauer, 2010). Individual correlations are displayed in grey points as well as the linear trend line. The correlation coefficient and corresponding 95%-Bayesian Credibility Interval (BCI) are displayed in the upper right corner of each plot.

unchosen source, $\hat{\rho} = -.36\ [-.47,\ -.25]$ (see also Fig. 7). This means that individuals who were more often in a state of source uncertainty due to poor source memory also experienced greater cognitive conflict. In a second step, we analyzed whether the correlation specific to our sample can be generalized to the population level. That is, will the reported correlation be valid for a newly drawn sample from the same population (Ly et al., 2017)? The sample correlations were reused to derive the posterior of the correlation adjusted for sampling error, which are the averaged across all posterior samples estimate the population correlation (Heck, Arnold, & Arnold, 2018; Ly et al., 2017). The BCI of the posterior distribution for the population correlation of source guessing and MAD included 0 [−.10, .46], indicating that the reported correlation was restricted to the sample. The BCI for the correlation of source memory and MAD, however, indicated that the results were also valid for a new sample [−.58, −.09].

5. Discussion

The primary aim of this study was to examine the cognitive dynamics of stereotypical influences on the ability of attributing information to its origin. For this purpose, we made use of the increasingly popular process-tracing method of mouse tracking. We tracked individuals' mouse movements when deciding whether a stereotype-consistent or -*in*consistent source presented a respective statement. Mouse tracking has thus far only been applied to study recognition memory (e.g., Koop & Criss, 2016; Papesh & Goldinger, 2012), another crucial part of source monitoring, but not yet to source monitoring on its own. We endeavored to disentangle and investigate the nature specifically behind stereotypical influences on source-monitoring processes by means of response dynamics as potentially reflecting the rather complex cognition involved in these processes. The results did not support our a priori hypotheses that the stereotype consistency of the statement and the source alone induced cognitive conflict or that source guessing and cognitive conflict were substantially correlated. We could show, however, that processes of social categorization due to age stereotypes in combination with the correctness of the source response induced cognitive conflict in the source-attribution process that, in addition, varied as a function of individuals' source memory.

5.1. Age stereotypes elicit decision uncertainty in source attributions

We found a substantial interaction of response correctness and consistency of the statement-source combination: while processing consistent statement-source combinations (age reflected in statement corresponds to source age), greater trajectory curvature was observed when the source attribution was incorrect compared to correct. When looking at correct source attributions only, we replicated previous findings that the cognitive conflict (as indicated by larger MAD) was more pronounced for *in*consistent than for consistent statement-source combinations. That is, individuals were (spatially) more attracted to the stereotype-consistent source when correctly choosing the stereotype-*in*consistent source. Thus, an internal *in*consistency in the perception of typical/*a*typical statements of different-aged people was uncovered by MAD. The mouse-tracking results indicate that age stereotypes were activated during source attribution prompting individuals to consider both sources. Mouse tracking, therefore, informed us that participants were actually attracted towards and considering the unchosen source based on its stereotypical features (e.g., replicating Cassidy et al., 2017; Freeman & Ambady, 2009; Freeman et al., 2010). The results also show the added value of analyzing mouse trajectories during source attributions: Mouse movements are better able to capture this simultaneous consideration of both sources compared to, for instance, reaction times.[11]

5.2. Stereotype consistency influences prototypical mouse-trajectory shapes

The MAD results received further support when looking at the trajectory shape at the prototype level: trajectories conformed with a higher probability to a more curved prototype category when individuals responded incorrectly to an originally consistent statement-source combination (i.e., responding stereotype-*in*consistent) compared to correctly (i.e., responding stereotype-consistent). In addition, when individuals responded correctly, trajectories were more likely to belong to a more curved prototype if the source's age was *in*consistent (compared to consistent) with the respective statement typicality. This

[11] As reaction times are an established dependent variable for investigating cognitive processes (e.g., Nosofsky & Palmeri, 1997; Ratcliff, Smith, Brown, & McKoon, 2016), we also analyzed the effect of statement-source consistency on reaction times in an exploratory manner. By leveraging this analysis, we hope to get a more complete picture of the cognitive processes involved in source attributions. In addition, we also analyzed different indices of trajectory curvature to test the stability of our MAD findings. The results of both additional analyses can be found in the Supplemental Materials.

interaction pattern for correct source attributions was revealed in both MAD and prototypes but could not be found for either dependent variable for incorrect source attributions. But looking at the distribution of prototypes for incorrect source attributions, it appears that at least the descriptive pattern of prototypes shows an effect of consistency similar to the one found for correct source attributions. When participants responded incorrectly in trials in which the age of the presenting source was consistent with the stereotypicality of the statement (i.e., chose the source for which the statement was *in*consistent with the respective age), they showed less straight and more discrete change-of-mind trials than when they incorrectly chose the stereotype-consistent source. This pattern might not have been significant because participants responded to more trials correctly than incorrectly due to relatively high source-memory performance. Additionally, the correctness of the response was also confounded with the consistency: consistent statement-source combinations were remembered more often than *in*consistent ones.

Our results also highlight why it is beneficial to look at the prototypical curvature of mouse trajectories, since it enabled us to cautiously infer the underlying process behind the trajectories. For instance, it seems as if the observed (aggregated) curvature mainly stems from either curved trajectories or discrete change-of-mind trajectories. This gives reason to speculate whether there are possibly two different processes underlying the spatial attraction towards the unchosen source (Freeman et al., 2008). Dual-process accounts in social cognition (Devine, 1989), for instance, have proposed that the impression formation about others is based on a fast, non-conscious, and automatic evaluation before a second, slow, conscious, and more fine-grained modification emerges if the initial impression needs to be corrected. The same logic applies to bimodality of mouse trajectories (for a theoretical review, see Freeman et al., 2008; Freeman & Dale, 2013). Straight trajectories indicate that the initial categorization and later correction point in the same direction. Curved, and change-of-mind trajectories in particular, indicate that the initial categorization is incorrect and needs to be adjusted (Freeman et al., 2008). The bimodal distribution of mouse trajectories in our study might indicate that there is a dual process in source attributions as well. However, it cannot be ascertained with reasonable confidence that the bimodal nature of mouse trajectories during source attribution is traced back to source guessing or source memory (note that we were not able to incorporate trial-level trajectory indices as source-monitoring covariates in our model; see Limitations section).

5.3. Individuals guess based on age stereotypes when their source memory fails

We further demonstrated that processing social information is influenced by social expectancies about other (social) groups (cf. Ehrenberg & Klauer, 2005). We observed considerable individual differences in the age-stereotype reliance in source guessing but, on group level, replicated previous findings of stereotype-biased source guessing (Kuhlmann et al., 2016) and poor source memory as its determinant (Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013; Kuhlmann et al., 2016) with the latent-trait model (Klauer, 2010). Stereotypical knowledge was used as a backup for episodic-memory failures (cf. Sherman & Bessenoff, 1999), indicating that, when source memory was absent or at least not optimal for making an adequate source attribution, individuals substituted their missing contextual information with prior knowledge (cf. Hicks & Cockman, 2003). In this case, prior knowledge was based on stereotypical expectations and what is commonly associated with being young or old (at least to some extent). In the present experiment, the magnitude of source-guessing bias was comparable to Kuhlmann et al. (2016) using similar item material and an experimental procedure even if the overall guessing bias - still of credible size - may not be as pronounced as for other schema or stereotype domains. This might, however, be explained by the inherent nature of age stereotypes revealing that individuals do not have as strong of attitudes towards what is typically young and what is typically old (normed explicit typicality ratings, Kuhlmann et al., 2017) as, for instance, what is schematically expected to be said by a doctor versus a lawyer (see Bayen et al., 2000).

5.4. Source-monitoring processes partially map onto mouse trajectories

Furthermore, we examined whether the strength of source guessing was related to the curved mouse movements. The reliance on stereotypical knowledge when a participant did not remember the source was displayed in the curvature of mouse trajectories at first sight but became invalid when controlling for sampling error in the Bayesian-hierarchical MPT model. Strong claims about the underlying nature of source guessing (e.g., individuals who guessed more in line with stereotypes experienced greater conflict revealing a rather systematic/controlled process, in line with, e.g., Bayen & Kuhlmann, 2011; Spaniol & Bayen, 2002, as both age categories are at least partially considered before a final decision for one source is made) could not be derived with certainty. The integration of mouse tracking into modeled source-monitoring processes revealed that mouse movements mirror decision uncertainty due to poor source memory – not source guessing.

5.5. Limitations

While the study provides a promising insight into the complex processing in source monitoring, results should be viewed in the context of some limitations. The comparatively high complexity of statements could influence the observed effects due to relatively long cognitive processing (reading) that might contradict the prerequisite of mouse tracking, namely to react quickly and intuitively. To satisfy this need, future studies could take advantage of the less complex item material (e.g., words) commonly used in schematic source-monitoring research (e.g., Bayen et al., 2000; Kuhlmann et al., 2012; Schaper et al., 2019) to rule out this methodological concern, assuming that the cognitive basics of processing stereotypical and schematic information are not fundamentally different. Due to the mouse-tracking set-up we ensured, however, that the relevant information needed to infer the source (i.e., presentation of to-be-classified statement in the test phase) was available only shortly before the decision was made and, thus, reduced the amount of information that needs to be processed. Even when the complexity of the item material appears to be a limitation at first sight, we may also consider it an advantage from an applied point of view. In our everyday life, we commonly process highly complex (social) information, which is also true for source monitoring. Therefore, the use of rather complex, verbal item material mimics this process and generalizes to a more realistic setting on how we process social information.

The assignment of specific ages to the sources only at test can be discussed critically, too. Of course, when bearing social interactions in real life in mind, we often encode information with salient source characteristics (e.g., know the age or profession of a familiar person in advance). First, we wanted to create a test situation in which we could study stereotype influences on source-monitoring best. This has been shown to be true when no internal representation of item-source contingency can be built (e.g., Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012). Second, it may be reasonable to assume that in some situations, no specific source characteristics may be present (e.g., when interacting with a stranger via phone or communicating on the internet).

Furthermore, using a study design such as ours, the reliance on age stereotypes is an adequate, rational response strategy in the memory test under some circumstances, namely when the stereotype-consistent source leads to a correct source attribution (which is only true in half of the trials). The age stereotype is, however, misleading and results in incorrect source attributions in the other half of the trials. We are aware that the use of this null contingency between statement typicality and source age may seem artificial and does not necessarily reflect the "true" world (which is also based on the subjective perception of this reality; e.g., Augoustinos & Walker, 1998) in which people of different ages may be more likely to be associated with their respective age-stereotypical attitudes and corresponding behavior at least to some extent (Chan et al., 2012). But stereotypes also mirror a simplification of our social life (cf. Bordalo, Coffman, Gennaioli, & Shleifer, 2016) for which individual differences for members of a social group and context-dependencies (e.g., Casper, Rothermund, & Wentura, 2011; Gluth, Ebner, & Schmiedek, 2010; Kornadt & Rothermund, 2011) as well as counter-stereotypical exemplars may be (partly) neglected. We aimed to best study source monitoring, and guessing in particular, under circumstances with the largest possible *in*consistency in the perception of different-aged people. However, to what extent the cognitive bias observed in source guessing is generalizable to the impression formation about other persons outside of the lab still awaits further research and is beyond the scope of our paper.

Apart from potential limitations inherent in the experimental paradigm, an open issue concerns how source-monitoring processes and mouse movements are linked on each trial. We estimated individuals' guessing bias but aggregated trajectories across trials for the correlation analysis. Thus, to further foster the understanding of the nature of source guessing, future studies may use more advanced statistical analyses such as the ones reported by Matzke et al. (2015), Heck & Erdfelder, 2016, and Heck, Erdfelder, and Kieslich (2018) to estimate source-monitoring processes on a trial-by-trial basis (considering also item heterogeneity) covarying with the respective mouse trajectory in that distinct, single trial - given a sufficient amount of data per participant which in our case were considerably low when participants classified a statement as new and no source attribution was made. This shortage of data was due to our experimental design: To follow recommendations of two spatially-separated response options with regard to the mouse-tracking set-up (see Hehman et al., 2015), we needed to split up and implement a two-stage test phase. Due to this set-up, we faced missing values by design whenever individuals respond new in the first stage and subsequently no mouse movements for this specific statement were recorded in the source-attribution stage. Therefore, an important caveat that we would like to highlight is that, unfortunately, we were not able to jointly model categorical data and continuous variables as suggested by Heck et al., 2018.

Methodological weaknesses could be overcome when adapting the experimental paradigm to the needs of more sophisticated approaches of data analysis, for example by using a one stage source-monitoring design which offers three response alternatives (i.e., the two sources and "new" for distractor items). This design would allow more sophisticated analyses and researchers interested in source monitoring may use such a set-up in combination with mouse tracking. As we were mainly interested in the clean measurement of mouse trajectories, we opted for the two-stage process which only presents two response alternatives in every stage (i.e., "old" vs. "new" in the recognition stage and "young source" vs. "old source" in the source-attribution stage). With this procedure, mouse trajectories and MAD are easily interpretable.

5.6. Future directions for response dynamics in source monitoring

Taken together, our study was a first attempt to examine the cognitive processes behind source monitoring within the domain of age stereotypes by tracking individuals' mouse movements while attributing information to its origin. The results show us how important it is to apply new methodologies to already well-established paradigms to inform us further about the underlying mechanisms of cognitive processes. Furthermore, we have uncovered additional insights that the involvement of stereotypes in source attributions is accompanied by cognitive conflict, shedding light on the processes in the application of stereotypical knowledge. In prior studies, mouse movements have been tracked to investigate the consequential behavior of stereotypical categorizations (see Stillman et al., 2018). Researchers should continue to pursue this line of research as stereotypes influence not only how we process information but also how we remember and retrospectively associate people with attributes or behavior that did not originate from them. We demonstrated that source memory may be a crucial and adequate prerequisite (Macrae, Bodenhausen, Schloerscheidt, & Milne, 1999) for individuation as it prevents the necessity to make use of prior knowledge in source guessing. Accurate encoding, maintenance, and retrieval seem to be key to reduce stereotypical influences in social categorization. If source memory fails, source guessing can come into play (for an overview, see Kuhlmann & Bayen, 2016) which, in turn, may lead to biased information processing and inaccurate impression management about others in the future (cf. Ehrenberg & Klauer, 2005; Sherman & Bessenoff, 1999), with potential behavioral consequences (e.g., false accusations in court: Lindsay, 2014).

Thus, if the underlying nature of source guessing (e.g., rather systematic/controlled or heuristic/automatic) is studied sufficiently, interventions to overcome the reliance on stereotypes in source monitoring could be targeted to its nature. In a first step, the reliance on stereotypes in source guessing should be manipulated, prompting individuals to either guess stereotype-biased or even counter-stereotype-biased (e.g., manipulate the item-source contingency, Bayen & Kuhlmann, 2011; or negate stereotypes before testing, Marsh et al., 2006). Assuming that a repeatedly processing counter-stereotypical exemplars will modify and shape basic knowledge structures in cognition, interventions for stereotype change may be a promising avenue to follow up on (Ehrenberg & Klauer, 2005). In a second step, the effectiveness of counter-stereotypical interventions on source-monitoring processes could then be measured with mouse tracking again - a valuable tool to provide a sophisticated understanding of the underlying processes in the future.

Appendix A. Illustration of Bayesian-hierarchical MPT model

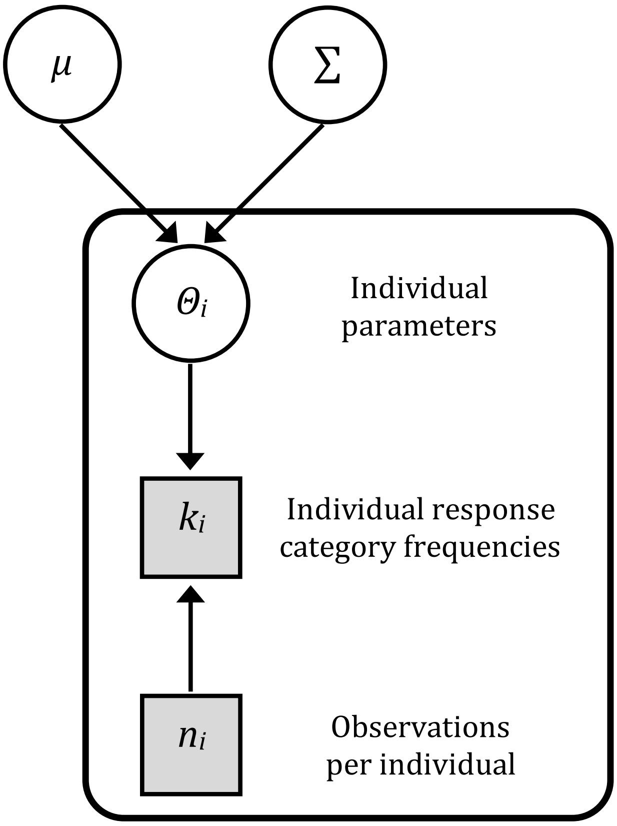

Fig. A1. Conceptual illustration of a Bayesian-hierarchical MPT model (e.g., Klauer, 2010). Observed outcomes are represented as shaded, unobserved outcomes are blank. Discrete variables are reflected in square nodes; continuous variables are reflected in circular nodes. i denotes the i-th individual; k denotes the individual categorical frequencies resulting from task responses; n denotes the number of trials per individual. Because the categorical responses depend on the number of trials and the source-monitoring parameters θ, arrows are directed towards the node k. μ and Σ represent the probit-transformed group-level parameters with mean and covariance matrix to be estimated from the data. Probit-transformed parameters assume to follow a multivariate normal distribution. Standard normal distributions for μ and a scaled inverse-Wishart distribution for Σ are used as priors.

Appendix B. Analysis of a differentiated MPT model split up by item type

Calculating the Pearson's χ^2 based model-fit statistics for hierarchical MPT models from Klauer (2010), the chosen restrictions of the MPT model adequately fitted our data as indicated by non-significant Bayesian posterior predictive p-values $> .05$ (Meng, 1994) for the mean ($T_1 = .44$) and covariance ($T_2 = .427$) structure of the data. We additionally used a more differentiated model without aggregating our data into typical and atypical statement-source combinations. Results for this analysis follow for which parameter estimates (Table B1) and their correlations (Table B2) are presented. The more differentiated model fitted the data as well (mean, $T_1 = .208$, and covariance structure, $T_2 = .13$) but comparing both models trading off model fit and complexity reflected in the Deviance Information Criterion (DCI; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002) showed worse model fit for the differentiated model ($\text{DIC}_{\text{aggregated}} = 1453$ vs. $\text{DIC}_{\text{differentiated}} = 2324$). We, thus, reported analyses based on the less differentiated model.

Table B1
Model-based estimates of source-monitoring parameter means and correlations with maximum absolute deviations for the differentiated model.

Parameter	Mean estimate	MAD
D_yy	.63 [.57, .69]	−.25 [−.45, .13]
D_yo	.62 [.54, .69]	−.23 [−.44, .04]
D_oy	.63 [.56, .69]	−.16 [−.43, .30]
D_oo = D_ny = D_no	.65 [.61, .70]	**−.31 [−.45, −.14]**
d_yy	.38 [.05, .73]	**−.31 [−.46, −.10]**
d_yo	.47 [.17, .80]	**−.34 [−.46, −.19]**
d_oo	.34 [.07, .58]	**−.32 [−.47, −.14]**
d_oy	.34 [.06, .71]	**−.31 [−.44, −.15]**
b	.43 [.37, .50]	.02 [−.18, .22]
a_y = g_y	.62 [.52, .70]	.20 [−.09, .41]
a_o = g_o	.56 [.46, .64]	.18 [−.13, .40]

Note. Mean parameter estimates of the Bayesian-hierarchical MPT model and their correlation with covariate Maximum Absolute Deviations (MAD) of the more differentiated model. The first letter of each parameter denotes the respective source-monitoring parameter: D = item memory, d = source memory, b = old-new item guessing, g = source guessing. The letter after the underscore character denotes the age of the source (either y = young or o = old) which originally presented the statement or the distractor status of a statement (n = new). The second letter after the underscore character denotes the statement's typicality (either y = typical young or o = typical old). Note that for source guessing (g) the single denotation refers to the statement's typicality. For the means of source guessing, values $> .50$ indicate a stereotype bias (i.e., guessing the source which age is consistent with the statement's typicality). As indicated by the equality sign, item-memory parameters and source-guessing parameters were equated between statement types to reach model identifiability. Model-fit indices (see Appendix B) suggest that it was legitimate to do so. 95%-Bayesian Credibility Intervals (BCI) are displayed in brackets. Credible correlations are marked in bold.

Table B2
Model-based estimates of source-monitoring parameter correlations for the differentiated model.

Parameter	*D_yy*	*D_yo*	*D_oy*	*D_oo*	*d_yy*	*d_yo*	*d_oo*	*d_oy*	*b*	*g_y*	*g_o*
D_yy	–	–	–	–	–	–	–	–	–	–	–
D_yo	.28 [−.39, .78]	–	–	–	–	–	–	–	–	–	–
D_oy	.31 [−.58, .86]	.20 [−.48, .76]	–	–	.35 [−.67, .89]	.35 [−.70, .88]	–	.34 [−.69, .87]	−.10 [−.59, .45]	−.21 [−.80, .63]	−.24 [−.85, .61]
D_oo	.50 [−.28, .88]	.31 [−.25, .75]	.34 [−.57, .84]	–	**.59 [.12, .88]**	.63 [.27, .88]	.62 [.18, .90]	**.58 [.16, .89]**	−.04 [−.47, .38]	−.32 [−.77, .23]	−.36 [−.79, .26]
d_yy	.52 [−.37, .92]	.23 [−.37, .78]	–	–	–	–	–	–	–	–	–
d_yo	.58 [−.31, .92]	.16 [−.41, .71]	–	–	**.68 [.19, .93]**	–	–	–	–	–	–
d_oo	.57 [−.26, .91]	.33 [−.27, .81]	.34 [−.66, .89]	–	**.67 [.17, .93]**	**.69 [.25, .93]**	–	**.64 [.15, .91]**	−.02 [−.49, .44]	−.40 [−.85, .32]	−.52 [−.89, .19]
d_oy	.55 [−.35, .91]	.17 [−.41, .71]	–	–	**.68 [.21, .93]**	**.81 [.53, .95]**	–	–	−.09 [−.53, .35]	−.51 [−.88, .12]	−.49 [−.90, .20]
b	−.06 [−.53, .44]	−.03 [−.51, .48]	–	–	−.13 [−.56, .34]	−.05 [−.49, .39]	–	–	–	–	–
g_y	−.36 [−.84, .45]	.07 [−.56, .64]	–	–	−.49 [−.86, .16]	−.48 [−.87, .16]	–	–	−.03 [−.47, .43]	–	–
g_o	−.39 [−.87, .49]	.02 [−.66, .63]	–	–	−.47 [−.89, .27]	−.56 [−.92, .13]	–	–	.08 [−.40, .54]	**.59 [.04, .89]**	–

Note. Parameter correlations of the more differentiated Bayesian-hierarchical MPT model. The first letter of each parameter denotes the respective source-monitoring parameter: D = item memory, d = source memory, b = old-new item guessing, g = source guessing. The letter after the underscore character denotes the age (either y = young or o = old) of source which originally presented the statement or the distractor status of a statement (n = new). The second letter after the underscore character denotes the statement's typicality (either y = typical young or o = typical old). Note that for source guessing (g), the single denotation refers to guessing the source's age for which the statement was typical. For reasons of convenience, the restricted model (see Table B1) is shown in which $D\_oo = D\_no = D\_ny$ and $a\_y = g\_y$ / $a\_o = g\_o$. 95%-Bayesian Credibility Intervals (BCI) are displayed in brackets. Credible correlations are marked in bold.

Appendix C. Parameter-recovery simulation

Table C1
Summarized results of the parameter-recovery simulation of the latent-trait 2HTSM.

	Parameter	"True"	Recovered estimates			
			Posterior mean	2.5%	97.5%	True in 95%-BCI
Mean	*D*	.65	.64	.61	.68	1
	d	.37	.38	.23	.52	1
	b	.40	.40	.35	.46	1
	g	.59	.59	.54	.63	1
Correlation	[*d*;*D*]	.59	.50	.17	.77	.86
	[*d*;*b*]	−.29	−.23	−.57	.13	.22
	[*d*;*g*]	−.52	−.44	−.73	−.08	.70
	[*D*;*b*]	−.33	−.27	−.60	.11	.27
	[*D*;*g*]	−.25	−.24	−.58	.14	.23
	[*b*;*g*]	.11	.10	−.28	.46	.08

Note. Simulation was based on 500 replications with 7 chains of 30,000 iterations (15,000 as burn-in period) each. D = item memory, d = source memory, b = old-new item guessing, g = source guessing. "True" refers to the data-generating values taken from the results of our study (best guess of true values). The simulated data were based on the number of items (90 in total) and participants (60 individuals) as in our study. Only samples that showed convergence ($\hat{R} < 1.05$; Gelman & Rubin, 1992) were retained for analyses. Group-level means μ were on the probability scale, correlations were on the latent probit scale. Posterior means and credibility intervals (2.5 and 97.5% quantiles) were estimated per replication and averaged across replication. True in 95%-BCI refers to the percentage of replications in which the 95%-Bayesian Credibility Interval (BCI) excluded 0 indicating credible estimates/correlations. Model-specific details and a commented script of the analyses can be retrieved from https://osf.io/85936/?view_only=214fa149e30748cbbd07852c43fab419.

Appendix D. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jesp.2019.103917.

References

Abney, D. H., McBride, D. M., Conte, A. M., & Vinson, D. W. (2015). Response dynamics in prospective memory. *Psychonomic Bulletin & Review, 22*(4), 1020–1028. https://doi.org/10.3758/s13423-014-0771-6.

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93*(2), 203–231. https://doi.org/10.1037/0033-2909.93.2.203.

Anderson, N. H., & Barrios, A. A. (1961). Primacy effects in personality impression formation. *Journal of Abnormal and Social Psychology, 63*(2), 346–350. https://doi.org/10.1037/h0046719.

Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin & Review, 20*(2), 326–333. https://doi.org/10.3758/s13423-012-0342-7.

Augoustinos, M., & Walker, I. (1998). The construction of stereotypes within social psychology: From social cognition to ideology. *Theory & Psychology, 8*(5), 629–652.

Bargh, J. A. (1994). The four horseman of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. Wyer, & T. Skrull (Eds.). *Handbook of social cognition: Basic processes; applications* (pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. https://doi.org/10.1007/s00572-005-0022-9.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*(4), 548–564. https://doi.org/10.1037/0033-295X.97.4.548.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin and Review, 6*(1), 57–86. https://doi.org/10.3758/BF03210812.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language, 64*(1), 1–17. https://doi.org/10.1016/j.jml.2010.09.001.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197.

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition, 28*(3), 480–500. https://doi.org/10.3758/BF03198562.

Bell, R., Buchner, A., Kroneisen, M., & Giang, T. (2012). On the flexibility of social source memory: A test of the emotional incongruity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(6), 1512–1529. https://doi.org/10.1037/a0028219.

Bordalo, P., Coffman, K., Gennaioli, N., & Shleifer, A. (2016). Stereotypes. *The Quarterly Journal of Economics, 131*(4), 1753–1794.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1), 9. https://doi.org/10.5334/joc.10.

Casper, C., Rothermund, K., & Wentura, D. (2011). The activation of specific facets of age stereotypes depends on individuating information. *Social Cognition, 29*(4), 393–414.

Cassidy, B. S., Sprout, G. T., Freeman, J. B., & Krendl, A. C. (2017). Looking the part (to me): Effects of racial prototypicality on race perception vary by prejudice. *Social Cognitive and Affective Neuroscience, 12*(4), 685–694. https://doi.org/10.1093/scan/nsw177.

Chan, W., Mccrae, R. R., De Fruyt, F., Jussim, L., Löckenhoff, C. E., De Bolle, M., ... Terracciano, A. (2012). Stereotypes of age differences in personality traits: Universal and accurate? *Journal of Personality and Social Psychology, 103*(6), 1050–1066. https://doi.org/10.1037/a0029712.

Cloutier, J., Freeman, J. B., & Ambady, N. (2014). Investigating the early stages of person perception: The asymmetry of social categorization by sex vs. age. *PLoS ONE, 9*(1), 1–7. https://doi.org/10.1371/journal.pone.0084677.

Dale, R., & Duran, N. D. (2011). The cognitive dynamics of negated sentence verification. *Cognitive Science, 35*(5), 983–996.

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition, 35*(1), 15–28. https://doi.org/10.3758/BF03195938.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5.

Dignath, D., Pfister, R., Eder, A. B., Kiesel, A., & Kunde, W. (2014). Something in the way she moves: Movement trajectories reveal dynamics of self-control. *Psychonomic Bulletin and Review, 21*(3), 809–816. https://doi.org/10.3758/s13423-013-0517-x.

Dshemuchadse, M., Scherbaum, S., & Goschke, T. (2013). How decisions emerge: Action dynamics in intertemporal decision making. *Journal of Experimental Psychology: General, 142*(1), 93–100. https://doi.org/10.1037/a0028499.

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology, 41*(4), 369–387. https://doi.org/10.1016/j.jesp.2004.08.001.

Farmer, T. A., Cargill, S. A., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: Computer mouse trajectories suggest parallel syntactic processing. *Cognitive Science, 31*(5), 889–909. https://doi.org/10.1080/03640210701530797.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Faulkenberry, T. J. (2014). Hand movements reflect competitive processing in numerical cognition. *Canadian Journal of Experimental Psychology, 68*(3), 147–151. https://doi.org/10.1037/cep0000021.

Freeman, J. B., & Ambady, N. (2009). Motions of the hand expose the partial and parallel activation of stereotypes. *Psychological Science, 20*(10), 1183–1188. https://doi.org/10.1111/j.1467-9280.2009.02422.x.

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research*

Methods, 42(1), 226–241. https://doi.org/10.3758/BRM.42.1.226.
Freeman, J. B., & Ambady, N. (2011). Hand movements reveal the time-course of shape and pigmentation processing in face categorization. *Psychonomic Bulletin and Review, 18*(4), 705–712. https://doi.org/10.3758/s13423-011-0097-6.
Freeman, J. B., & Ambady, N. (2014). The dynamic interactive model of person construal: Coordinating sensory and social processes. In J. W. Sherman, B. Gawronski, & Y. Trope (Eds.). *Dual-process theories of the social mind* (pp. 235–248). New York, NY: US: Guilford Press.
Freeman, J. B., Ambady, N., Rule, N. O., & Johnson, K. L. (2008). Will a category cue attract you? Motor output reveals dynamic competition across person construal. *Journal of Experimental Psychology: General, 137*(4), 673–690. https://doi.org/10.1037/a0013875.
Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods, 45*(1), 83–97. https://doi.org/10.3758/s13428-012-0225-x.
Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology, 2*(59), https://doi.org/10.3389/fpsyg.2011.00059.
Freeman, J. B., Pauker, K., Apfelbaum, E. P., & Ambady, N. (2010). Continuous dynamics in the real-time perception of race. *Journal of Experimental Social Psychology, 46*(1), 179–185. https://doi.org/10.1016/j.jesp.2009.10.002.
Freeman, J. B., Pauker, K., & Sanchez, D. T. (2016). A perceptual pathway to bias: Interracial exposure reduces abrupt shifts in real-time race perception that predict mixed-race bias. *Psychological Science, 27*(4), 502–517. https://doi.org/10.1177/0956797615627418.
Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–511. https://doi.org/10.1214/ss/1177011136.
Gluth, S., Ebner, N. C., & Schmiedek, F. (2010). Attitudes toward younger and older adults: The German aging semantic differential. *International Journal of Behavioral Development, 34*(2), 147–158.
Greene, R. L. (1986). Sources of recency effects in free recall. *Psychological Bulletin, 99*(2), 221–228. https://doi.org/10.1037/0033-2909.99.2.221.
Hartigan, J. A., & Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics, 13*(1), 70–84. https://doi.org/10.1214/aos/1176346577.
Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods, 50*(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7.
Heck, D. W., & Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review, 23*(5), 1440–1465. https://doi.org/10.3758/s13423-016-1025-6.
Heck, D. W., Erdfelder, E., & Kieslich, P. J. (2018). Generalized processing tree models: Jointly modeling discrete and continuous variables. *Psychometrika, 83*(4), 893–918. https://doi.org/10.1007/s11336-018-9622-0.
Hehman, E., Carpinella, C. M., Johnson, K. L., Leitner, J. B., & Freeman, J. B. (2014). Early processing of gendered facial cues predicts the electoral success of female politicians. *Social Psychological and Personality Science, 5*(7), 815–824. https://doi.org/10.1177/1948550614534701.
Hehman, E., Stolier, R. M., & Freeman, J. B. (2015). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes and Intergroup Relations, 18*(3), 384–401. https://doi.org/10.1177/1368430214538325.
Hicks, J. L., & Cockman, D. W. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language, 48*(3), 489–501.
Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika, 59*(1), 21–47. https://doi.org/10.1007/BF02294263.
Huette, S., & McMurray, B. (2010). Continuous dynamics of color categorization. *Psychonomic Bulletin and Review, 17*(3), 348–354. https://doi.org/10.3758/PBR.17.3.348.
Johnson, K. L., Freeman, J. B., & Pauker, K. (2012). Race is gendered: How covarying phenotypes and stereotypes bias sex categorization. *Journal of Personality and Social Psychology, 102*(1), 116–131. https://doi.org/10.1037/a0025335.
Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3.
Johnson, M. K., & Raye, C. L. (2000). Cognitive and brain mechanisms of false memories and beliefs. In D. L. Schacter, & E. Scarry (Eds.). *Memory, brain, and belief* (pp. 35–86). Cambridge, MA: Harvard University Press.
Kieslich, P. J., & Henninger, F. (2017). Mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49*(5), 1652–1667. https://doi.org/10.3758/s13428-017-0900-z.
Kieslich, P. J., Henninger, F., Wulff, D. U., Haslbeck, J., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: A practical guide to implementation and analysis. In M. Schulte-Mecklenbeck, A. Kühberger, & J. Johnson (Eds.). *A handbook of process tracing methods* (pp. 111–130). New York, NY: Routledge.
Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika, 75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0.
Klauer, K. C., & Ehrenberg, K. (2005). Social categorization and fit detection under cognitive load: Efficient or effortful? *European Journal of Social Psychology, 35*(4), 493–516. https://doi.org/10.1002/ejsp.266.
Koop, G. J., & Criss, A. H. (2016). The response dynamics of recognition memory: Sensitivity and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(5), 671–685. https://doi.org/10.1037/xlm0000202.
Kornadt, A. E., & Rothermund, K. (2011). Contexts of aging: Assessing evaluative age stereotypes in different life domains. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 66*(5), 547–556.
Kroneisen, M., & Bell, R. (2013). Sex, cheating, and disgust: Enhanced source memory for trait information that violates gender stereotypes. *Memory, 21*(2), 167–181. https://doi.org/10.1080/09658211.2012.713971.
Kuhlmann, B. G., & Bayen, U. J. (2016). Metacognitive aspects of source monitoring. In J. Dunlosky, & S. K. Tauber (Eds.). *The oxford handbook of metamemory* (pp. 149–168). New York, NY: Oxford University Press.
Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. *Psychology and Aging, 31*(8), 875–889. https://doi.org/10.1037/pag0000140.
Kuhlmann, B. G., Kornadt, A. E., Bayen, U. J., Meuser, K., & Wulff, L. (2017). Multidimensionality of younger and older adults' age stereotypes: The interaction of life domain and adjective dimension. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences, 72*(3), 434–440. https://doi.org/10.1093/geronb/gbv049.
Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(5), 1365–1376. https://doi.org/10.1037/a0028147.
Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.
Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and Aging, 30*(1), 180–193. https://doi.org/10.1037/a0038586.
Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology, 55*(1), 1–7. https://doi.org/10.1016/j.jmp.2010.08.013.
Lenth, R. (2019). R package version 1.3.2*emmeans: Estimated marginal means, aka least-squares means [computer software manual]*https://CRAN.R-project.org/package=emmeans.
Levy, B. R., Slade, M. D., Chung, P. H., & Gill, T. M. (2015). Resiliency over time of elders' age stereotypes after encountering stressful events. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences, 70*(6), 886–890. https://doi.org/10.1093/geronb/gbu082.
Lindsay, D. S. (2014). Memory source monitoring applied. In T. Perfect, & D. S. Lindsay (Eds.). *Sage handbook of applied memory* (pp. 59–75). London, UK: Sage. https://doi.org/10.1016/j.jelechem.2018.02.006.
Ly, A., Boehm, U., Heathcote, A., Turner, B. M., Forstmann, B., Marsman, M., & Matzke, D. (2017). A flexible and efficient hierarchical bayesian approach to the exploration of individual differences in cognitive-model-based neuroscience. In A. A. Moustafa (Ed.). *Computational models of brain and behavior* (pp. 467–479). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119159193.ch34.
Macrae, C. N., Bodenhausen, G. V., Schloerscheidt, A. M., & Milne, A. B. (1999). Tales of the unexpected: Executive function and person perception. *Journal of Personality and Social Psychology, 76*(2), 200–213. https://doi.org/10.1037/0022-3514.76.2.200.
Malejka, S., & Bröder, A. (2016). No source memory for unrecognized items when implicit feedback is avoided. *Memory & Cognition, 44*(1), 63–72. https://doi.org/10.3758/s13421-015-0549-8.
Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). Gender and orientation stereotypes bias source-monitoring attributions. *Memory, 14*(2), 148–160. https://doi.org/10.1080/09658210544000015.
Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7.
Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika, 80*(1), 205–235. https://doi.org/10.1007/s11336-013-9374-9.
McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science, 19*(1), 22–24. https://doi.org/10.1111/j.1467-9280.2008.02041.x.
Meng, X.-L. (1994). Posterior predictive *p*-values. *The Annals of Statistics, 22*(3), 1142–1160. https://doi.org/10.1214/aos/1176325622.
Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104*(2), 266–300. https://doi.org/10.1037/0033-295X.104.2.266.
Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin and Review, 19*(5), 906–913. https://doi.org/10.3758/s13423-012-0281-3.
R Core Team (2018). *R: A language and environment for statistical computing.* Vienna, Austriahttps://www.R-project.org/.
Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007.
Rudolph, U., Böhm, R., & Lummer, M. (2007). Ein Vorname sagt mehr als 1000 Worte. *Zeitschrift für Sozialpsychologie, 38*(1), 17–31. https://doi.org/10.1024/0044-3514.38.1.17.
SAS Institute (1989). *SAS/STAT user's guide*. Cary, NC.
Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metamemory expectancy illusion and schema-consistent guessing in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(3), 470–496. https://doi.org/10.1037/xlm0000602.
Sherman, J. W., & Bessenoff, G. R. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science, 10*(2), 106–110. https://doi.org/10.1111/1467-9280.00116.
Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2018). R package version 0.22-1*afex: Analysis of factorial experiments [computer software manual]*https://CRAN.R-project.

org/package = afex,.
Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology, 54*(1), 167–183. https://doi.org/10.1016/j.jmp.2009.06.007.
Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 631–651. https://doi.org/10.1037/0278-7393.28.4.631.
Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64*(4), 583–639.
Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science, 15*(5), 207–211. https://doi.org/10.1111/j.1467-8721.2006.00437.x.
Spivey, M. J., Dale, R., Knoblich, G., & Grosjean, M. (2010). Do curved reaching movements emerge from competing perceptions? A reply to van der Wel et al. (2009). *Journal of Experimental Psychology: Human Perception and Performance, 36*(1), 251–254. https://doi.org/10.1037/a0017170.
Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences, 22*(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012.
Stroebe, W., & Insko, C. A. (1989). Stereotype, prejudice, and discrimination: Changing conceptions in theory and research. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.). *Stereotyping and prejudice* (pp. 3–34). New York, NY: Springer Series in Social Psychology. https://doi.org/10.1007/978-1-4612-3582-81.
Szaszi, B., Palfi, B., Szollosi, A., Kieslich, P. J., & Aczel, B. (2018). Thinking dynamics and individual differences: Mouse-tracking analysis of the denominator neglect task. *Judgement and Decision Making, 13*(1), 23–32. https://doi.org/10.1057/9780230584556.
Tabatabaeian, M., Dale, R., & Duran, N. D. (2015). Self-serving dishonest decisions can show facilitated cognitive dynamics. *Cognitive Processing, 16*(3), 291–300. https://doi.org/10.1007/s10339-015-0660-6.
Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer. http://ggplot2.org.
Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science, 20*(11), 1428–1435. https://doi.org/10.1111/j.1467-9280.2009.02448.x.
Wulff, D. U., Haslbeck, J., Kieslich, P. J., Henninger, F., & Schulte-Mecklenbeck, M. (2019). Mouse-tracking: Detecting types in movement trajectories. In M. Schulte-Mecklenbeck, A. Kühberger, & J. Johnson (Eds.). *A handbook of process tracing methods* (pp. 131–145). New York, NY: Routledge.
Yu, Z., Wang, F., Wang, D., & Bastin, M. (2012). Beyond reaction times: Incorporating mouse-tracking measures into the implicit association test to examine its underlying process. *Social Cognition, 30*(3), 289–306. https://doi.org/10.1521/soco.2012.30.3.289.

Memory & Cognition
https://doi.org/10.3758/s13421-019-01008-1

Is knowledge reliance in source guessing a cognitive trait? Examining stability across time and domain

Liliane Wulff[1] · Beatrice G. Kuhlmann[1]

Abstract
When people need to infer the source of information in the absence of memory, they may rely on general knowledge (e.g., stereotypes) to guess the source. Prior research documented task-related determinants and individual differences of stereotype reliance in source guessing, but little is known about the underling nature of this process. In two experiments, we tested whether a cognitive trait could account for the knowledge reliance in source guessing. Participants performed two distinct study–test cycles of a classical source-monitoring paradigm in which two person sources present stereotypical information that in a later test phase had to be attributed to its origin. In Experiment 1, both tasks used item material from the same knowledge domain (age stereotypes) and were either separated by 10 minutes or 7 days. In Experiment 2, we used item material from two different knowledge domains (Task 1: age stereotypes; Task 2: gender stereotypes). Although cross-task correlations of source-guessing parameters from Bayesian-hierarchical multinomial processing tree model analyses showed only weak positive correlations, absolute source guessing remained fairly stable within individuals across time (Experiment 1) and knowledge domains (Experiment 2). Considering statistical challenges of the assessment of relative stability via correlations, we suggest based on the stricter absolute stability criterion that source guessing rather encompasses trait-like features. We discuss implications regarding the generalizability and nature of source guessing in comparison to other cognitive processes involved in source attribution, which were highly stable in both experiments.

Keywords Source guessing · Cognitive trait · Bayesian-hierarchical multinomial modeling

The ability to remember the source of information (e.g., *who* presented it) is crucial for many types of cognitive tasks in our everyday life (Johnson, Hashtroudi, & Lindsay, 1993). For instance, we have to remember who told us something or where we read the latest news. The process of attributing information to sources is highly susceptible to prior knowledge, such as stereotypes and schemas (e.g., Bayen, Nakamura, Dupuis, & Yang, 2000). For example, if we do not remember who told us about a new aerobic course in the gym, we may infer that the person source was a young adult rather than an old adult based on guessing in line with our stereotypes about aging and what we typically associate with being young and old (Kuhlmann, Bayen, Meuser, & Kornadt, 2016; Kuhlmann, Kornadt, Bayen, Meuser, & Wulff, 2017). This inherent reliance on prior knowledge, of course, does not always lead to the correct source attribution and a biased misattribution could consequently support the maintenance (and reinforcement) of stereotypes over time.

Crucially, some studies have documented individual differences in the extent to which people rely on their prior knowledge in source attributions (Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013; Spaniol & Bayen, 2002). Although influences on knowledge reliance in source guessing are well studied on the group level (contingency perception: e.g., Arnold et al., 2013; Bayen & Kuhlmann, 2011; Kuhlmann, Vaterrodt, & Bayen, 2012; Spaniol & Bayen, 2002; source memory: e.g., Kuhlmann et al., 2016), little is known about the underlying mechanisms of source guessing that may explain the origin of individual differences. That is, why do some people strongly rely on their prior knowledge and others not at all? Is knowledge reliance in source attributions an intraindividual predisposition? Thus, the purpose of the current set of experiments

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13421-019-01008-1) contains supplementary material, which is available to authorized users.

✉ Liliane Wulff
wulff@psychologie.uni-mannheim.de

[1] Department of Psychology, School of Social Sciences, University of Mannheim, D-68131 Mannheim, Germany

was to examine individual knowledge-based source guessing and whether source guessing manifests a "cognitive trait" (Kantner & Lindsay, 2012)—reflected in stability across time and (content) domain of the item material—or fluctuates across tasks, which in turn would imply that situational determinants of source guessing are primarily at play. Keeping in mind the detrimental effects of incorrect attributions of memories based on source guessing in the social environment (e.g., biased impression formation about others; Bell, Giang, Mund, & Buchner, 2013; Ehrenberg & Klauer, 2005; Sherman & Bessenoff, 1999) or in the legal context (e.g., false accusations in court; Lindsay, 1994, 2014), it is essential to understand the potential persistence of such a source-guessing bias.

Knowledge reliance in source guessing

Attributing information to sources—defined as *source monitoring*—can be based on two cognitive processes: People can either rely on memory for contextual details (e.g., spatial, temporal, episodic, perceptual or affective details) or general knowledge (Johnson et al., 1993; Mitchell & Johnson, 2000). General knowledge includes schemas that organize, link, and structure information based on previous experience (Alba & Hasher, 1983), and stereotypes, "a set of beliefs about the personal attributes of a group of people" (Stroebe & Insko, 1989, *p*. 5). The reliance on prior knowledge in source-monitoring tasks has been demonstrated in a multitude of studies (e.g., profession schemas: Bayen et al., 2000; room schemas: Küppers & Bayen, 2014; gender stereotypes: Marsh, Cook, & Hicks, 2006; social stereotypes: Ehrenberg & Klauer, 2005; Sherman & Bessenoff, 1999; age stereotypes: Kuhlmann et al., 2016).

Using Bayen et al.'s (2000) doctor–lawyer paradigm as an example, we will illustrate the standard experimental design for investigating influences of stereotypes and schemas on source monitoring. In the study phase, participants learn information presented by two sources (e.g., two persons: "Tom" & "Jim") that present an equal number of statements that are either typical for their source (i.e., consistent with the schema or stereotype associated with this source category; e.g., lawyer: "I have to be in court at nine") or statements that are typical for the other source (e.g., doctor: "It will take a couple of hours to get the results of this blood test"). At test, participants are informed about the specific category each source belongs to (e.g., profession of a doctor & a lawyer), and they have to remember the source of each statement among new distractor statements.

In previous source-monitoring studies, participants commonly made more correct source attributions when a statement was typical for its source (e.g., Bayen et al., 2000). The cognitive processes that lead to this performance benefit for typical statements are not evident from the categorical responses given in the task at first sight. Both improved memory for or biased guessing in favor of typical statement–source combinations could explain the performance benefit. Multinomial processing tree (MPT) models such as the *Two-high-threshold multinomial model of source monitoring* (2HTSM; Bayen, Murnane, & Erdfelder, 1996) disentangle different cognitive processes contributing to observable behavior (e.g., Batchelder & Riefer, 1999; Bayen et al., 1996). The 2HTSM decomposes the following processes from source-attribution behavior: item memory (i.e., the probability of recognizing a presented statement), source memory (i.e., the probability of remembering the source that presented a statement), item guessing (i.e., the probability of guessing that an unrecognized statement was old), and source guessing (i.e., the probability of guessing that a statement was presented by a specific [e.g., the typical] source).

Bayen et al. (2000) showed that schema-reliant source guessing, and not differential source memory, caused the performance benefit for typical statement–source pairs.[1] When participants did not remember which source presented a statement, they guessed the schema-consistent source. Thus, a lack of memory is often found to be compensated with preexisting knowledge at least when analyzing the data on the group-level (i.e., aggregated across participants; Arnold et al., 2013; Bayen & Kuhlmann, 2011; Bayen et al., 2000; Ehrenberg & Klauer, 2005; Kuhlmann et al., 2016; Kuhlmann et al., 2012; Küppers & Bayen, 2014; Spaniol & Bayen, 2002). As MPT models are merely measurement models that quantify certain cognitive processes, the question what determines these processes—and especially biased source guessing—remains open. Several determinants of biased source guessing have been identified already: misperception of item–source contingency (Arnold et al., 2013; Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012; Spaniol & Bayen, 2002), poor source memory (Arnold et al., 2013; Kuhlmann et al., 2016), reduced cognitive capacities at encoding (Bayen & Kuhlmann, 2011; Ehrenberg & Klauer, 2005; Kuhlmann et al., 2012), and provision of category information about sources after encoding (Hicks & Cockman, 2003; Kuhlmann et al., 2012). Even though these determinants can explain differences in source guessing between experimental conditions, they do not consider remaining interindividual variance (but see Arnold et al., 2013; Spaniol & Bayen, 2002) and whether biased guessing happens occasionally or consistently.

[1] Schemas/stereotypes can also influence source memory. If an item is very atypical for the source with which it was presented, then attention is drawn to this expectation-violating item-source combination, which may result in better source memory for the atypical, schema-incongruent or stereotype-incongruent information (i.e., inconsistency effect; e.g., Bell, Buchner, Kroneisen, & Giang, 2012; Ehrenberg & Klauer, 2005; Kroneisen & Bell, 2013; Küppers & Bayen, 2014). However, this is not consistently observed whereas schema-based/stereotype-based source guessing is consistently observed, even when this inconsistency effect occurs. In any case, this guessing bias leads to a performance advantage for typical item-source combinations.

Recent methodological advances in the estimation of MPT model parameters offer more sophisticated statistical analyses and allow for inferences about cognitive processes on an individual level. One such advance is the latent-trait model, a Bayesian-hierarchical extension of multinomial models proposed by Klauer (2010). The latent-trait model accounts for variability between individuals and estimates parameters that reflect latent cognitive processes for each individual separately. With this extension, individual parameter estimates can be compared and correlated across different applications of a task (e.g., across time and stimulus material).

Cognitive trait

One possible extension to the aforementioned group-level determinants could be that individual differences in knowledge reliance in source guessing reflect trait-like stability. Independent of the experimental condition, some people might be more inclined to rely on knowledge to fill memory gaps whereas others might be less inclined to do so. If one follows the definition of Roberts (2009), trait-like stability is “meaningfully consistent” (*p*. 139) behavior. That is, although behavior should be consistent across certain situations it must not be identical in order to be considered stable. Trait-like stability would, for instance, be reflected in correlational evidence of two variables but not necessarily in the same numeric point estimates across situations. For response tendencies such as old–new guessing bias in recognition tasks (Kantner & Lindsay, 2012, 2014), knowledge reliance in recognition heuristic (Michalkiewicz & Erdfelder, 2016), and risky decision-making (Glöckner & Pachur, 2012), trait-like stability in these cognitive processes underlying responses was reported.

In a study by Kantner and Lindsay (2012), individuals learned English nouns for a later memory test in two separate study–test cycles. After studying these items, individuals indicated whether they had learned the items before or not. The tendency to respond that a previously studied item was presented (“old”), measured as signal-detection theory (SDT) response-bias measure *c* (Macmillan, 1993), correlated highly across tests, separated by either 10 minutes or seven days, and also across stimulus material (words & digital images of paintings). The authors concluded that the old–new response bias in memory is a *cognitive trait* that should not vary deliberately within individuals and has predictive value given comparable experimental settings across tests. They explicitly distinguished it from a personality trait but acknowledged that cognitive and personality traits could be associated.

Cross-task stability of the reliance on prior knowledge has also been documented for the recognition-heuristic use in judgment tasks (i.e., choosing the recognized object and ignore knowledge about it; Goldstein & Gigerenzer, 2002). Applying a latent-trait MPT model, Michalkiewicz and Erdfelder (2016) found strong correlations of individuals’ recognition-heuristic use between two tests, separated by either one or seven days or immediately succeeding each other but with varying judgment–content domains. Thus, the general tendency to rely on the recognition heuristic (i.e., on one’s prior knowledge) in judgments seems to be a cognitive disposition that is stable across time and independent of the knowledge domain.

Even though the source-guessing bias is somewhat different from the old–new response bias (Kantner & Lindsay, 2012) and the recognition-heuristic use (Michalkiewicz & Erdfelder, 2016), these measures share not only task-related features such as the mechanism of recognition, but, especially, the recognition-heuristic use shares a common content feature with source guessing: people’s reliance on prior knowledge. Thus, the trait-like stability may also hold for knowledge-based source guessing but has not been tested yet.

Cognitive trait = personality trait?

Kantner and Lindsay (2014) examined “personality trait-like qualities” (*p*. 1273), defined as an association between a response bias and a personality trait (a detailed description of key personality traits and cognitive-processing styles discussed in this section can be retrieved from Table 1). Their results brought only weak evidence for a correlation with need for cognition (NFC; Cacioppo & Petty, 1982) and internal punishment–reward preference (behavioral inhibition system/behavioral activation system; Carver & White, 1994) scores. Whereas a cognitive trait may be independent from a personality trait, Kantner and Lindsay (2014) acknowledged that this association may nonetheless exist. Michalkiewicz, Minich, and Erdfelder (2019) assessed the relation between the recognition-heuristic use and NFC as well as faith in intuition (FII; Epstein, Pacini, Denes-Raj, & Heier, 1996). They found a negative correlation of recognition-heuristic use and NFC but no substantial correlation with FII. Participants who scored low on NFC tended to use the recognition heuristic more frequently. This relationship was unique to NFC and held even after controlling for the Big Five personality traits. So, the recognition-heuristic use possesses personality trait-like qualities at least to some extent.

If knowledge reliance in source guessing turns out to be a cognitive trait, personality traits may be related to it as well. Additionally, knowledge reliance in source guessing may be also related to stereotypical thinking in general. We often use stereotypes as a cognitive tool (Gilbert & Hixon, 1991; Macrae, Milne, & Bodenhausen, 1994) to evaluate members of social groups because individuating is mentally effortful (Fiske, 1989). This lack of motivation to process individual information deeply could also hold for knowledge-based source guessing. For instance, and in line with the aforementioned research on the recognition heuristic, stereotyping is related to NFC in some studies. That is, people who score low on NFC tend to use stereotyping as mental shortcuts more often (e.g., Carter, Hall, Carney, & Rosip, 2006;

Table 1 Definition of exemplar (personality trait) constructs related to knowledge reliance and response bias

Construct	Definition	Reference
Recognition heuristic (RH)	Simple decision strategy applied in dichotomous judgment tasks. For instance, when participants need to answer the question "Which city is more populous: Tokyo or Busan?," they should choose the recognized object (here: Tokyo) according to the RH because they immediately recognize this city (and not the other; here: Busan). They could integrate more detailed knowledge about the city (e.g., that cities with international airports such as Tokyo are often populous) and come to the same choice.	e.g., Goldstein & Gigerenzer (2002); cf. Michalkiewicz & Erdfelder (2016)
Need for cognition (NFC)	Individual's tendency to engage in effortful, deep thinking and to enjoy structuring situations in a meaningful way.	cf. Cacioppo & Petty (1982)
Faith in intuition (FII)	Individual's tendency to rely on intuitive, experiential processing of information.	cf. Epstein, Pacini, Denes-Raj, & Heier (1996)
Big Five	Taxonomy for five basal personality traits. Openness for Experience contrasts curious and exploratory trait facets with rigid and traditional ones. Conscientiousness contrasts disciplined trait facets with unambitious ones. Extraversion contrasts warm and outgoing trait facets with reserved ones. Agreeableness contrasts generous and honest trait facets with selfish and aggressive ones. Neuroticism contrasts calm and stable trait facets with sad and scared ones.	e.g., Goldberg (1993); John & Srivastava (1999); cf. McCrae & Costa (2008)
Behavioral inhibition system/behavioral activation system (BIS/BAS)	Taxonomy for two motivational systems that underlie negative and positive affect. The BIS reflects an orientation towards aversive outcomes; the BAS reflects an orientation towards pleasant outcomes. The BAS scale can be subdivided into reward responsiveness, drive, and fun seeking.	cf. Carver & White (1994)

Note. Description of central concepts mentioned in the introductory paragraph including examples where these concepts have been referred to in the literature

Perlini & Hansen, 2001) or even less often in terms of stereotype-consistent recall (Crawford & Skowronski, 1998). Transferred to source guessing, this mixed evidence in terms of the direction of relation suggests that source guessing and NFC could be related, although the direction of correlation cannot be derived. With regard to the Big Five personality traits, Flynn (2005) showed that the trait Openness to Experience was negatively correlated with explicit interracial attitudes and positively with impressions of other-race persons. Carter et al., (2006) showed that the willingness to accept stereotyping is negatively correlated with Agreeableness and positively with Extraversion and Neuroticism. Assuming that the knowledge reliance in source guessing is also used to form an impression about other persons, one could speculate that it may be negatively correlated with Openness to Experience and Agreeableness and positively correlated with Extraversion and Neuroticism (based on Carter et al., 2006; Flynn, 2005). It is, however, generally debatable whether the correlation of stereotyping with personality traits or cognitive-processing styles can be transferred to stereotyping in memory tasks, and more specifically, to the cognitive process of source guessing, and this thereof remains an exploratory research question.

Overview of the current experiments

We knew from prior research that people differ in the extent to which they make use of stereotypes/schemas in source guessing (e.g., Arnold et al., 2013; Spaniol & Bayen, 2002). In two experiments, we examined the stability of knowledge reliance in source guessing to test whether Kantner and Lindsay's (2012, 2014) findings characterizing response bias as a cognitive trait generalize above and beyond old–new item guessing and the methodological approach of signal-detection theory to multinomial processing tree modeling. Following Michalkiewicz and Erdfelder (2016), we tested two different facets of stability as a trait-like predisposition—stability across time and knowledge domain. Therefore, we applied two distinct source-monitoring tasks to each participant, using item material of the same stereotype domain but separated by a time interval of 10 minutes or seven days (Experiment 1) or different stereotype material between tasks (Experiment 2).

Other than Kantner and Lindsay (2012, 2014), we specified a Bayesian-hierarchical model version of the 2HTSM (Bayen et al., 1996) to assess parameter correlations, but our data did nonetheless also allow for a replication of their findings with both MPT and SDT measures of response bias. In addition to this relative stability, we examined the absolute stability measured as the absolute difference of source guessing between tasks. Based on Roberts's (2009) stability definition, we did not necessarily expect to observe the exact same point estimates of source guessing across tasks (i.e., absolute stability) but rather a correlation across time or knowledge domain irrespective of the overall group-level estimate (i.e., relative stability). Participants who guess stereotype based more strongly than others in one test should be more prone to do so in another test even if the mean group-level source guessing changes (e.g., due to regression to the mean or reactive effects from the first task).

Experiment 1: Stability across time

If stereotype-based source guessing is a cognitive trait, it should be stable within individuals across time (Kantner & Lindsay, 2012; Roberts, 2009). Therefore, we tested whether participants show comparable (absolute and/or relative to the group level) knowledge reliance in source guessing in two separate source-monitoring tasks, performed either 10 minutes or seven days apart. We estimated individual source-monitoring parameters and cross-task correlations/differences using Bayesian-hierarchical modeling (latent trait; Klauer, 2010).

Method

Participants and design

The design was a 2 (time interval between tasks; between subjects) × 2 (source age: old vs. young; within subjects) × 2 (age stereotypicality: statement typicality for respective source age; within subjects) mixed factorial. As there is currently no appropriate power analysis for Bayesian-hierarchical MPT models available, we computed the equivalent frequentist analysis for orientation. An a priori power analysis in G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) revealed that assuming a moderate positive correlation of source guessing across time, Pearson's $r = .30$, $n = 82$ participants per condition are needed to yield a power $(1 - \beta)$ of .80 at $\alpha = .05$ (two-tailed). We increased the number of participants per condition beyond a minimum of 100 for two reasons: first, to increase estimation precision and, second, to fulfill our counterbalancing constraints. Additionally, we conducted recovery simulations to test the precision of parameter estimation for the recruited sample size and used item number in the experiment (see Table S1 in the Supplemental Material).[2]

In total, 224 students of the Universities of Mannheim and Heidelberg participated for psychology course credit or monetary compensation.[3] We randomly assigned participants to two conditions (between subjects) in which they had to complete two age-stereotype source-monitoring tasks either separated by 10 minutes (n = 114; M = 21.37 years, SD = 2.11 years, age range: 18–26 years, 76% women, 40% psychology majors) or seven days (n = 110; M = 20.76 years, SD = 2.12 years, age range: 18–26 years, 82% women, 50% psychology majors). Exclusion criteria during participant recruitment were color-blindness, age >26 years (i.e., older than the younger participants in the survey study for norming the item material; see Material section), neurological disorders, previous participation in a similar experiment, and insufficient German proficiency (i.e., learned after the age of 6).

Material

Age-stereotypical statements for the source-monitoring tasks consisted of everyday statements selected from a previous survey study (Kuhlmann et al., 2017). In this survey, 74 older (M = 70.17 years, 60-84 years) and 69 younger (M = 22.03 years, 18-26 years) participants rated the age typicality of 368 statements reflecting the positive or negative pole of three adjective dimensions (autonomy, instrumentality, and integrity) in five life domains (family & partnership; finances; friends & acquaintances; health, fitness, & appearance; religion & spirituality). An example statement is: "I volunteer at church." (adjective dimension: instrumentality [positive]; domain: religion & spirituality). The typicality of a statement for either a "young adult" or an "old adult" (between subjects) was rated on a 5-point Likert scale (1 = *very atypical*, 2 = *atypical*, 3 = *neither typical nor atypical*, 4 = *typical*, 5 = *very typical*). Following Kuhlmann et al. (2016), we defined a statement as typically–old if the mean typicality rating for an "old adult" was ≥3 and, at the same time, <3 for a "young adult" and vice versa for typically–young statements. One hundred and thirteen typically–old and 103 typically–young statements fit our criterion. Out of these, we randomly selected 60 typically–old and typically–young statements each and divided them into two item sets of 60 statements (30 of each typicality) for the two tasks. Item sets were comparable in their mean typicality ratings (all ps ≥ .649), polarity, adjective dimension and life domain. Furthermore, each of the 60-item sets was divided into three matched subsets of 20 statements, of which two subsets served as study lists and one as distractors at test. The assignment of item lists to study and distractor lists was counterbalanced across participants.[4] Further, the source ages and names were counterbalanced across tasks. At test, one source's age was indicated to be 70 years and the other's to be 23 years because these were the ages participants thought of while rating the statement typicality for an "old adult" or a "young adult," respectively, in the

[2] We thank an anonymous reviewer for suggesting these parameter-recovery simulations.

[3] For unknown reasons, source-guessing variance was strongly restricted for the first 54 participants in the 7-day condition, preventing a meaningful correlation analysis (see Fig. S1 in the Supplemental Material). Therefore, we started data collection for this condition anew. Analyses based on the initial data set of 54 participants can be retrieved from the Supplemental Material as well as analyses based on the initial 54 (see Fig. S1) and additional 110 participants combined (see Fig. S2). Inclusion of these data did not change the conclusions (see Tables S2 and S3).

[4] The assignment of study and distractor lists was not fully counterbalanced, as for organizational reasons more participants took part than necessary to fulfill counterbalance constraints. Because the differences in the number of participants between counterbalance conditions were small and not at all systematic, we preferred to analyze all collected data.

survey (Kuhlmann et al., 2017). In both tasks, common German last names (without indicating age or gender) served as sources: either "Müller" and "Schneider" or "Fischer" and "Schmidt."

Both experiments included the BFI-2 short form (Soto & John, 2017) derived from the German self-report long form (Danner et al., 2016; Rammstedt, Danner, Soto, & John, 2018). This short form consists of 30 items assessing the Big Five personality traits under 15 specific facet traits (each two items).[5] Items were rated on a 5-point Likert scale (1 = *strongly disagree*, 5 = *strongly agree*). Need for cognition (NFC) and faith in intuition (FII) were assessed with the German translation of the Rational Experiential Inventory (REI; Epstein et al., 1996; Keller, Bohner, & Erb, 2000). The REI consists of 29 items, 14 assessing the construct of NFC, 15 assessing FII, rated on a 7-point Likert scale (1 = *completely false*, 7 = *completely true*).

Procedure

Experiments were programmed with *OpenSesame* (Mathôt, Schreij, & Theeuwes, 2012) and run on computers with a screen resolution of 1,280 × 1,024 pixels. Participants were tested in groups up to four in individual computer booths. After providing written informed consent, each participant performed two distinct study–test cycles of a source-monitoring task. In each task, participants learned information presented by two sources that they had to remember in a following test phase. To render both tasks comparable, participants were explicitly instructed before learning that both item and source memory would later be tested. Although incidental source learning results in greater source-guessing bias (Kuhlmann et al., 2012), we used intentional source-learning instructions for both tests as participants would have known that we test their source memory during the second task based on the first tasks' memory test. We further informed participants that one source person was old and the other young but did not reveal the specific ages until test. We presented the statements trial by trial in the study phase, each combined with one of two source names. Each source presented half typical information and half atypical information. For instance, the "old adult" was associated with typically–old behavior, such as going to church or visiting the doctor but, at the same time, with typically–young behavior, such as going to the gym and meeting friends for a shopping tour, equally often. The source names were printed in green or yellow (counterbalanced between subjects) on black background (font: sans; font size: 35pt); the statements were written in white below (font: sans; font size: 32pt). Each statement–source combination remained on the screen for 4 s followed by an interstimulus interval of 250 ms. We randomized the order of statement–source combinations restricted to three subsequent statements presented by the same source.

At test, we informed participants that it might be helpful to know social category information about the sources (i.e., which source was 70 vs. 23 years old) for the upcoming memory test. During the self-paced test phase, participants again saw all statements from the study phase randomly intermixed with one-third new distractor statements. Participants then decided for each statement whether it had been presented with one of the two sources or was new. We presented the question "Who said" written in white font in the top left corner of the screen, and the test statement, centered on the screen. In the center, one source name was presented to the left, the other source name was presented to the right, both in the corresponding color (green or yellow) as in the study phase with the age information below (i.e., "70 years" vs. "23 years"). We additionally presented the response option "NEW" in the bottom center in red font. Participants responded via pressing the key "D" for the left response option, "K" for the right response option, and the space bar for the new response option. Key positions on the used QWERTZ keyboard corresponded to source position on the screen and were additionally marked with colored (green, yellow, red) stickers. Finally, participants estimated their perceived ratio of age-typical statement–source combinations (contingency judgment). That is, they estimated how many of the statements from the study phase of each typicality category (i.e., young, old; order counterbalanced between tasks and within subjects) had been presented by each source. The procedure for the second source-monitoring task was equivalent to the first but used different source labels (e.g., first task: "Fischer" & "Schneider"; second task: "Müller" & "Schmidt") and age-stereotypical statements.

We randomly assigned participants to either complete the two tasks in the first session, separated by 10 minutes, or one per session, separated by seven days. In the first session, participants in the 10-minute condition worked on the first source-monitoring task, filled out the demographic questionnaire and the BFI-2 within the fixed interval of 10 minutes (measured with a stopwatch) followed by the second source-monitoring task. If participants finished the questionnaires in less than 10 minutes, they were instructed to wait quietly until time had passed (nobody needed more than 10 minutes). Participants in the 7-day condition completed the first source-monitoring task and then rated the typicality of statements for either a man or a woman (used in Experiment 2) and completed the REI. Both participant groups returned to the lab one week later on the same day and, if possible, at the same time as for the first appointment (± 8 hours) to ensure comparability in the procedures. Participants in the 10-minute condition now rated the statement's gender typicality and filled

[5] Due to experimenter error, the first 26 questionnaires lacked one item for the domain of conscientiousness (from the trait facet responsibility), which we thus dropped from analyses for all participants in Experiment 1. We used the corrected questionnaire in Experiment 2.

out the REI; participants in the 7-day condition worked on the second source-monitoring task, then filled out the demographic questionnaire and the BFI-2. Finally, all participants were debriefed and compensated.

Results and discussion

Model-based analyses

We applied the 2HTSM (Bayen et al., 1996) to estimate the following underlying cognitive processes based on the observed response frequencies collected in each source-monitoring task: The probability of item recognition is measured by parameter D, separately for statements that were originally presented by the typical source (D_T), the atypical source (D_A), or are new (D_N). When a statement is recognized, the source may also be remembered with probability d. More specifically in the present paradigm, with probability d_T the typical source is remembered, whereas with d_A the atypical source is remembered. When source memory fails (with probability $1 - d_T$ or $1 - d_A$, respectively), guessing processes are engaged. Specifically, parameter a measures the probability to guess that a recognized statement is presented by the typical source and with the complementary probability $1 - a$ that the statement is presented by the atypical source. If item memory fails, the source can also not be remembered (Bell, Mieth, & Buchner, 2017; Malejka & Bröder, 2016; see also Klauer & Kellen, 2010). With probability b, participants then guess that a statement was previously presented in the study phase (i.e., is "old"), either followed by guessing the typical (probability g) or atypical source (probability $1 - g$). With probability $1 - b$, participants guess that the statement is new.

We used Submodel 4 of the 2HTSM (cf. Bayen et al., 2000) which has been used to analyze data from similar paradigms in prior studies (Kuhlmann et al., 2016; Spaniol & Bayen, 2002), illustrated in Fig. 1. This is the most parsimonious model to start with as it explains source-monitoring data with four parameters only: D measures the probability of recognizing a statement as old, regardless of the originally presenting source, or new (i.e., $D_T = D_A = D_N = D$). Parameter d measures the probability of remembering the source of a recognized old statement (i.e., $d_T = d_A = d$). Parameter b measures the probability of guessing "old." And, crucially, g measures the probability of guessing the typical source, regardless of whether the statement was recognized or merely guessed to be old (i.e., $a = g$). Thus, the data basis for our analysis consisted of "typical" trials (aggregating attributions across typically–old statements presented by the "old" source & typically–young statements presented by the "young" source), "atypical" trials (aggregating attributions across typically–old statements presented by the "young" source & typically–young statements presented by the "old" source) and new trials (aggregating attributions across typically–young and -old distractors).[6]

To obtain individual parameter estimates and correlations, we applied a latent-trait Bayesian-hierarchical extension of multinomial modeling (Klauer, 2010). This approach accounts for variability between participants by treating parameters as random variables.[7] Individual model parameters and their correlations are estimated jointly in one model. Separate parameter estimates for each person are constrained by the population-level model that assumes a multivariate normal distribution of the probit-transformed model parameters with a mean and covariance matrix to be estimated from the data. This Bayesian-hierarchical approach is particularly advantageous here because, in addition, it allows inclusion of external covariates of model parameters (e.g., personality traits). Samples from the posterior distribution of parameters are drawn with the Monte Carlo–Marcov chain (MCMC) algorithm. We used the *R* package *TreeBUGS* for this purpose (Heck, Arnold, & Arnold, 2018). To fit a latent-trait MPT model, prior distributions for the group-level mean and covariance are needed that are updated by the incorporated data resulting in a posterior distribution of model parameters. Following Klauer (2010) and Matzke, Dolan, Batchelder, and Wagenmakers (2015), *TreeBUGS* uses weakly informative priors (see the Supplemental Material or Heck et al., 2018, for further details). The number of iterations for each model was fit to convergence. That is, we cycled through three MCMC chains with 20,000 iterations each until all parameters reached the desired convergence criterion as indicated by $\hat{R} < 1.05$ (Gelman & Rubin, 1992). In order to check for convergence stability, we ran

[6] Following a suggestion by an anonymous reviewer, we additionally fitted an alternative no-bias model attributing any typicality-based differences in source attributions to differences in source memory (i.e., d_T versus d_A estimated independently) but not source-guessing bias (i.e., $a = g = .50$; retaining the item-memory restriction $D_T = D_A = D_N$). This model neither fit the data from the 10-minute condition of Experiment 1, $T_1 < .001$, $T_2 < .001$, nor the data from Experiment 2, $T_1 = .001$, $T_2 < .001$. In the 7-day condition, this no-bias model merely fit in the mean structure, $T_1 = .09$, but not in the covariance structure $T_2 < .001$. The trade-off between model fit and complexity using the Bayesian measure DIC (Spiegelhalter, Best, Carlin, & Van der Linde, 2002) revealed that Submodel 4 (reported in the main analysis of the manuscript; equating source memory for typical and atypical statement-source combinations) is preferred across all experiments/conditions (10-minute condition of Experiment 1, $DIC_{Submodel\ 4}$: 4718 vs. $DIC_{No\text{-}bias\ Model}$: 4939; 7-day condition of Experiment 1, $DIC_{Submodel\ 4}$: 4541 vs. $DIC_{No\text{-}bias\ Model}$: 4656; Experiment 2, $DIC_{Submodel\ 4}$: 5125 vs. $DIC_{No\text{-}bias\ Model}$: 5528). This confirms that the typicality effects we observed in source attributions are best described by source-guessing biases, not source-memory differences. We thus retained Submodel 4 of the 2HTSM for the main analyses as this was the most parsimonious model, which fit the data from all our experimental conditions well and provides a valid measure of source-guessing bias.

[7] Traditionally, MPT model parameters are estimated from the group-level aggregated response frequencies using maximum likelihood estimation (e.g., Hu & Batchelder, 1994), and thus inherently assuming homogeneity of both participants and items (Batchelder & Riefer, 1999). As a result, MPT models can be misspecified and their parameter estimates biased if participants and/or items are heterogeneous (Klauer, 2010; Matzke, Dolan, Batchelder, & Wagenmakers, 2015; Smith & Batchelder, 2010). Most crucially, individual differences are ignored in this analysis approach.

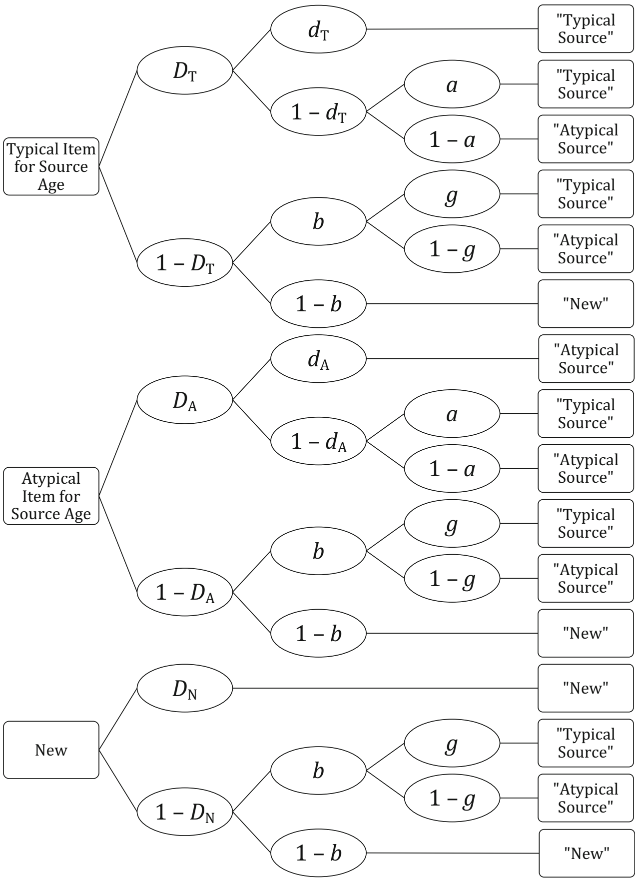

Fig. 1 Two-high-threshold multinomial model of source monitoring adapted to the current paradigm. D_T = probability of recognizing a statement that had been presented by a typical source; D_A = probability of recognizing a statement that had been presented by an atypical source; D_N = probability of knowing a statement is new; d_T = probability of correctly remembering the source of a statement that had been presented with a typical source; d_A = probability of correctly remembering the source of a statement that had been presented with an atypical source; b = probability of guessing that an unrecognized statement is old; a = probability of guessing that a recognized statement had been presented by the typical source; g = probability of guessing that an unrecognized statement had been presented by a typical source. Adapted from "Source Discrimination, Item Detection, and Multinomial Models of Source Monitoring" (Bayen et al., 1996)

one additional cycle of 20,000 iterations. The convergence criterion was maintained for this additional cycle in all our analyses; otherwise, the estimation would have started anew.

We tested whether the model adequately describes our data using the test statistics T_1 and T_2 proposed by Klauer (2010). T_1 tests the mean fit and T_2 the covariance fit by computing the distance between expected and observed mean frequencies or covariances, respectively, based on Pearson's χ^2 statistic. A satisfactory group-level model fit is indicated by a posterior predictive p value >.05. We obtained good model fit for the means T_1 in both conditions, p = .111 (10 minutes) and p = .29 (7 days) but observed a satisfactory model fit for the covariance structure only in the 7-day condition, p = .062, and not in the 10-minute condition, p = 0.[8]

We report the group-level mean estimates of the posterior distribution for the four source-monitoring parameters and corresponding 95% Bayesian credibility intervals (BCI) as presented in Table 2. In the following, we will focus on the source-guessing parameter g. We first examined whether there was a stereotype bias in source guessing on the group level. Therefore, we sampled the posterior distribution for the difference in the mean source-guessing parameter g from the chance level of .50 (i.e., $\Delta g -$ .50). A statistical meaningfully difference of posterior parameter distributions would be reflected in a BCI of the difference that excluded zero. In the 10-minute condition, source guessing differed substantially from chance-level guessing in Task 1, $\Delta(g -$ .50) = .07 [.03, .11], and Task 2, $\Delta(g -$ .50) = .06 [.02, .09], thus replicating prior findings of source-guessing biases in the domain of age stereotypes (in traditional aggregate-MPT analyses; Kuhlmann et al., 2016). In the 7-day condition, source guessing exceeded .50 only in the first task, $\Delta(g -$ .50) = .05 [.01, .08] but not in second task one week later, $\Delta(g -$ .50) = .01 [−.03, .04].

Stability

Relative stability We define relative stability as the cross-task correlations of source-monitoring parameters. These were automatically estimated in the latent-trait approach as implemented in *TreeBUGS* (Heck et al., 2018). Table 3 presents estimates of the between-task correlations of each source-monitoring parameter based on the posterior samples. For source guessing, the individual-level correlation across tasks is also depicted in Fig. 2. Numerically, these source-guessing correlations were in the positive direction, suggesting that participants whose source guessing was biased in the first test also tended to be biased in the second test. However, both correlations were rather small and their BCIs included zero, such that a negative correlation or null population correlation cannot be ruled out. Thus, there was no conclusive evidence for individual-level relative stability of the stereotype bias in source guessing. To exclude poor reliability of source guessing as an explanation for the observed weak cross-task correlations, we confirmed via split-half analyses that source guessing was estimated with satisfactory reliability within each task ($\hat{\rho}$ = .47–.70; see Appendix A for further details). In addition, we conducted simulation analyses to confirm that trait-level

[8] The misfit of the covariance structure in the 10-minute condition persisted (p < .01) for all other submodels. The covariance misfit in the 10-minute condition was caused by seven participants for whom Submodel 4 did not fit in the mean structure either. The exclusion of these seven participants led to a significant improvement of both model fit indicators (T_1 = .349, T_2 = .614), without substantially changing the overall parameter group means and correlations. We thus decided to retain Submodel 4 for all further analyses, which has been successfully applied to analyze source-monitoring data in prior studies with similar paradigms (Bayen et al., 2000; Kuhlmann et al., 2016).

Table 2 Multinomial processing tree model group-level parameter estimates of both experiments

Experiment	Task	Parameters			
		D	*d*	*b*	*g*
Experiment 1					
10 minutes	1	.73 (.40) [.70, .76]	.36 (1.46) [.24, .48]	.42 (.52) [.36, .47]	.57 (.37) [.53, .61]
	2	.67 (.40) [.63, .70]	.38 (1.83) [.23, .54]	.37 (.59) [.31, .42]	.56 (.30) [.52, .59]
7 days	1	.71 (.35) [.69, .74]	.39 (1.12) [.28, .49]	.42 (.58) [.36, .48]	.55 (.36) [.51, .58]
	2	.69 (.40) [.65, .72]	.38 (1.35) [.27, .50]	.32 (.50) [.27, .36]	.50 (.29) [.47, .54]
Experiment 2					
	1	.55 (.29) [.52, .58]	.34 (1.01) [.25, .43]	.48 (.36) [.44, .51]	.59 (.47) [.55, .62]
	2	.68 (.44) [.65, .72]	.53 (.71) [.47, .60]	.33 (.58) [.28, .38]	.51 (.45) [.47, .55]

Note. We used Submodel 4 of the Two-high-threshold multinomial processing tree model from Bayen et al., (1996) adapted to our paradigm. Parameter values were estimated using the latent-trait Bayesian-hierarchical approach (Klauer, 2010) with the *R* package *TreeBUGS* (Heck et al., 2018). See text for further details. Parentheses represent standard deviations (on the probit scale). Brackets indicate 95% Bayesian credibility intervals (BCI). D = probability of recognizing a statement as previously presented and probability of knowing that a distractor statement is new; d = probability of correctly remembering the (either typical or atypical) source of a statement; b = probability of guessing that an unrecognized statement is old; g = probability of guessing that a statement had been presented by the source with the typical age/gender for this statement when the source is not remembered

correlations of .70 between source guessing across tasks could have been recovered with satisfactory precision in the latent-trait model with the given number of items and participants. Crucially note, however, that our recovery simulations, about which more details can be retrieved from the Supplemental Material, showed that weaker correlations (i.e., ρ = .30) could not be precisely recovered; thus, there may be a true weak correlation of source-guessing bias across tasks, but it is smaller than the trait-like across-task correlation reported for old–new guessing.

Indeed, in contrast to source guessing, other memory and guessing processes showed strong relative stability across time (see Table 3). This also pertained to the item-guessing parameter b, which measures the tendency to guess old when an item is not recognized and can be interpreted as the MPT analogue to the old–new response bias in the signal-detection framework underlying Kantner and Lindsay (2012). As is evident in Table 3, we conceptually replicated their findings of cross-time relative stability in item guessing. Additionally, we generally replicated the results from Kantner and Lindsay when fitting the SDT bias measure c for item memory to the data as used by the authors (whereas adapting the bias measure c for source memory did not result in substantial correlations across tasks). A description of the SDT analyses and their results are summarized in Appendix B.

Absolute stability As the correlative evidence with regard to relative stability largely hinges on the observed interindividual variance in source guessing, a restriction of variance as present (at least to some extent) in this first experiment, may mask existing comparable source-guessing parameters on an individual level. For this reason, we further tested whether participants' source-guessing bias remained stable across tasks in absolute terms by sampling the posterior distribution of the absolute mean differences between both tasks.[9] In the 10-minute condition, the absolute mean difference in the source-guessing parameter across tasks indicated a change to a credible extent, but it was of negligible size, $|\Delta(\text{Task 1} - \text{Task 2})| = .02$ [.001, .06]. In the 7-day condition, the mean difference in the source-guessing parameter across tasks also changed credibly, albeit the change was again fairly small, $|\Delta(\text{Task 1} - \text{Task 2})| = .04$ [.003, .09]. That is, as graphically illustrated in Fig. 2, most participants' absolute reliance on stereotypes in source guessing as measured by source-guessing bias was to a fairly comparable extent in both tasks. A closer inspection of the reliance on stereotypes in source guessing revealed that for only 10.53% (12 out of 114) of participants in the 10-minute condition and 22.73% (25 out of 110) in the 7-day condition, the mean absolute difference was larger than .15 between tasks. That is, albeit some participants changed their guessing behavior to a substantial degree, the vast majority of participants did not—irrespective of the time interval between tasks. Further, the guessing bias was descriptively more pronounced in the first task for most individuals ($n_{10 \text{ minutes}} = 63$, 55.26% of individuals; $n_{7 \text{ days}}$: 74, 67.27% of individuals), indicating regression to the mean tendencies (also evident from Fig. 2). Thus, in contrast to the correlational relative stability analysis, the analysis of the more conservative absolute stability in individual source guessing across time suggested that stereotype reliance in source guessing is stable across time in the majority of participants.

Covariates of source guessing

Table 4 presents estimates of the correlations between source guessing and z-transformed covariates, and Table 5 presents

[9] We thank Adam Osth for suggesting this analysis.

Table 3 Across-task correlations between source-monitoring parameters of both experiments

Experiment	Task 1	Task 2			
		D	*d*	*b*	*g*
Experiment 1					
10 minutes	*D*	**.64 [.43, .82]**	**.37 [.12, .60]**	−.21 [−.46, .06]	−.11 [−.42, .21]
	d	**.32 [.06, .54]**	**.31 [.07, .52]**	−.16 [−.41, .10]	**−.34 [−.61, −.03]**
	b	−.15 [−.42, .14]	−.04 [−.33, .25]	**.82 [.64, .94]**	.23 [−.10, .55]
	g	−.03 [−.31, .26]	−.07 [−.35, .21]	.14 [−.16, .41]	.27 [−.06, .56]
7 days	*D*	**.45 [.19, .69]**	−.02 [−.30, .26]	−.06 [−.36, .25]	−.05 [−.37, .27]
	d	**.31 [.04, .54]**	.14 [−.12, .38]	−.24 [−.52, .04]	−.10 [−.41, .22]
	b	−.08 [−.35, .21]	−.10 [−.37, .17]	**.38 [.06, .65]**	.03 [−.32, .36]
	g	−.11 [−.38, .17]	−.03 [−.32, .26]	.17 [−.15, .47]	.20 [−.14, .52]
Experiment 2					
	D	**.70 [.50, .85]**	**.35 [.07, .60]**	−.13 [−.41, .15]	−.23 [−.49, .05]
	d	**.31 [.06, .55]**	**.29 [.03, .53]**	.06 [−.22, .34]	−.06 [−.33, .21]
	b	−.11 [−.36, .15]	.09 [−.19, .35]	**.78 [.61, .91]**	−.09 [−.34, .18]
	g	−.22 [−.43, .01]	**−.26 [−.48, −.02]**	.14 [−.11, .39]	**.31 [.07, .53]**

Note. Sampled population correlation coefficients $\hat{\rho}$ are displayed. We used Submodel 4 of the two-high-threshold multinomial processing tree model from Bayen et al. (1996) adapted to our paradigm. Parameter values were estimated using the latent-trait Bayesian-hierarchical approach (Klauer, 2010) with the *R* package *TreeBUGS* (Heck et al., 2018). See text for further details. Brackets indicate 95% Bayesian credibility intervals (BCI). *D* = probability of recognizing a statement as previously presented and probability of knowing that a distractor statement is new; *d* = probability of correctly remembering the (either typical or atypical) source of a statement; *b* = probability of guessing that an unrecognized statement is old; *g* = probability of guessing that a statement had been presented by the source with the typical age/gender for this statement when the source is not remembered. Substantial correlations (BCI excludes 0) are marked in boldface

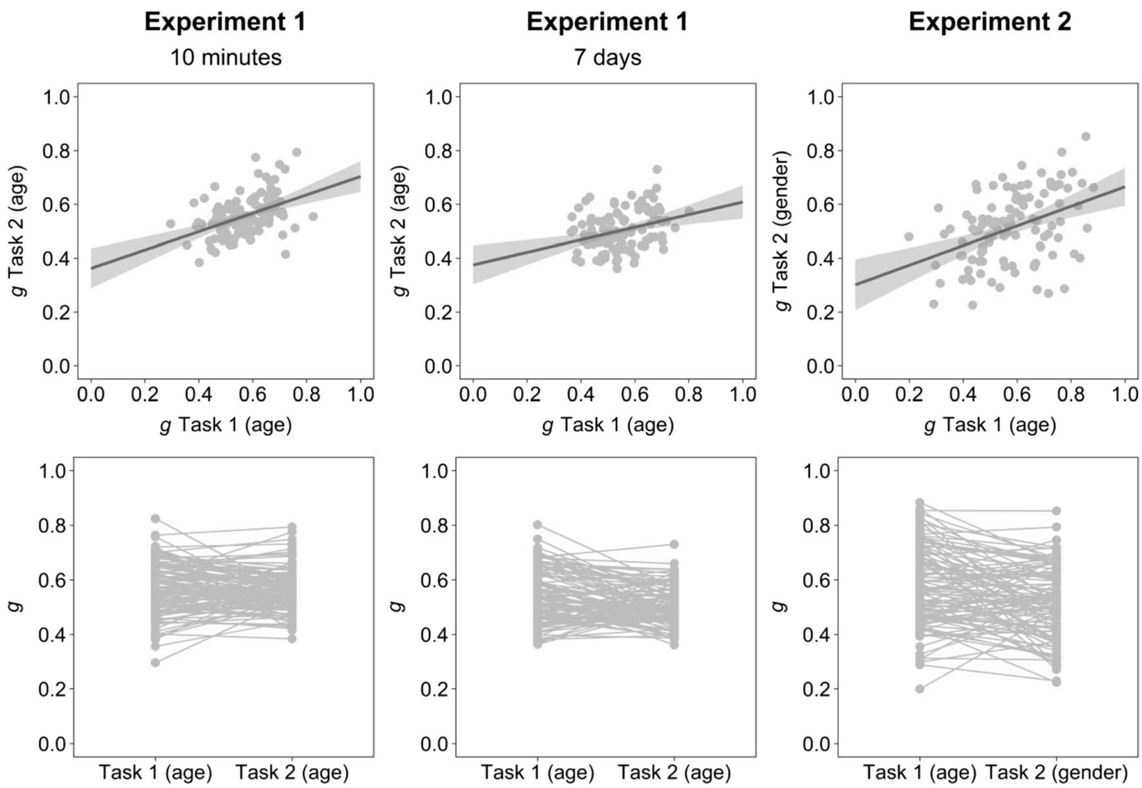

Fig. 2 Top row: Sampled population correlation coefficients $\hat{\rho}$ of source-guessing parameters (*g*) across tasks. Individual correlations are displayed in gray points as well as the linear trend line. Error bars represent standard errors (gray). Bottom row: Individual source-guessing parameter estimates connected across tasks. The respective task-specific knowledge domain is displayed in brackets. For both rows, first graph = Experiment 1, 10-minute condition; second graph = Experiment 1, 7-day condition; third graph = Experiment 2

Table 4 Correlations between source-monitoring parameters and covariates of both experiments

MPT Parameters	Task	Contingency	Cognitive-processing styles		Personality traits: Big Five				
			NFC	FII	O	C	E	A	N
Experiment 1 10 minutes									
D	1	.01 [−.09, .11]	.10 [−.01, .20]	−.09 [−.19, .01]	**.18 [.08, .28]**	−.01 [−.11, .09]	**−.13 [−.23, −.04]**	.06 [−.04, .16]	**−.13 [−.23, −.04]**
	2	.04 [−.06, .14]	**.21 [.11, .30]**	−.07 [−.17, .03]	**.26 [.16, .35]**	.07 [−.02, .17]	−.07 [−.17, .03]	**.17 [.08, .27]**	−.08 [−.18, .02]
d	1	−.02 [−.11, .07]	**.18 [.09, .27]**	−.01 [−.11, .08]	**.18 [.08, .27]**	−.07 [−.16, .03]	−.02 [−.11, .08]	.06 [−.03, .15]	−.02 [−.11, .07]
	2	.03 [−.07, .12]	**.16 [.05, .25]**	**−.10 [−.20, −.001]**	**.14 [.04, .24]**	−.01 [−.11, .09]	.04 [−.07, .14]	.07 [−.03, .16]	−.07 [−.16, .03]
b	1	.09 [−.03, .19]	−.01 [−.12, .11]	.06 [−.05, .17]	−.03 [−.14, .09]	−.06 [−.17, .05]	.03 [−.09, .14]	−.03 [−.14, .08]	−.05 [−.16, .06]
	2	.01 [−.10, .12]	−.01 [−.12, .10]	.04 [−.06, .15]	−.02 [−.13, .08]	−.03 [−.13, .07]	.01 [−.10, .12]	.01 [−.09, .12]	−.05 [−.15, .06]
g	1	**.20 [.07, .32]**	.01 [−.13, .14]	**.16 [.04, .28]**	−.01 [−.13, .12]	−.01 [−.13, .11]	**.15 [.02, .27]**	.02 [−.10, .15]	.06 [−.07, .18]
	2	.10 [−.04, .24]	−.13 [−.26, .01]	.11 [−.03, .24]	−.05 [−.18, .09]	.09 [−.05, .23]	−.08 [−.22, .06]	−.04 [−.17, .10]	.02 [−.10, .15]
7 days									
D	1	**−.15 [−.26, −.04]**	**.12 [.01, .23]**	**−.18 [−.29, −.07]**	**.12 [.01, .23]**	.03 [−.08, .14]	−.03 [−.15, .08]	0 [−.12, .11]	.08 [−.04, .19]
	2	.02 [−.09, .13]	**.12 [.02, .23]**	−.08 [−.18, .03]	.09 [−.01, .19]	.09 [−.02, .19]	.08 [−.03, .18]	.08 [−.02, .18]	.02 [−.08, .12]
d	1	−.04 [−.14, .07]	.05 [−.06, .15]	−.07 [−.17, .04]	−.03 [−.13, .07]	.06 [−.04, .16]	−.10 [−20, 0]	.04 [−.06, .15]	.04 [−.05, .14]
	2	0 [−.10, .10]	.07 [−.04, .17]	.05 [−.05, .16]	.10 [−.01, .20]	**.13 [.03, .22]**	**.17 [.07, .26]**	**.14 [.04, .24]**	−.05 [−.15, .04]
b	1	.03 [−.08, .15]	.11 [−.01, .22]	−.02 [−.14, .10]	.05 [−.06, .17]	.04 [−.08, .15]	−.01 [−.12, .12]	−.05 [−.16, .07]	−.04 [−.16, .08]
	2	.03 [−.10, .17]	−.02 [−.14, .10]	−.06 [−.18, .06]	−.01 [−.13, .11]	−.01 [−.13, .12]	−.09 [−.21, .03]	**−.13 [−.25, −.01]**	.07 [−.06, .19]
g	1	.12 [−.01, .25]	**−.15 [−.27, −.03]**	.04 [−.08, .17]	−.07 [−.19, .06]	−.01 [−.14, .12]	.09 [−.04, .21]	−.05 [−.17, .07]	.07 [−.06, .19]
	2	**.23 [.08, .37]**	**−.14 [−.28, −.01]**	.02 [−.12, .16]	−.11 [−.24, .03]	−.10 [−.24, .04]	−.05 [−.19, .10]	−.11 [−.25, .03]	.13 [−.01, .27]
Experiment 2									
D	1	.06 [−.07, .19]	**.16 [.05, .26]**	.02 [−.10, .14]	.10 [−.01, .21]	.08 [−.03, .18]	.02 [−.08, .13]	**.15 [.05, .25]**	.01 [−.10, .11]
	2	**−.16 [−.24, −.08]**	**.17 [.09, .24]**	−.02 [−.11, .06]	.07 [−.004, .15]	0 [−.08, .08]	−.01 [−.08, .07]	**.19 [.11, .26]**	.01 [−.07, .08]
d	1	**−.27 [−.38, −.16]**	**.13 [.03, .23]**	−.08 [−.19, .04]	.07 [−.03, .18]	−.09 [−.19, .01]	−.09 [−.20, .02]	−.03 [−.14, .08]	−.02 [−.12, .08]
	2	.05 [−.06, .17]	.06 [−04, .17]	**−.12 [−.22, −.02]**	−.04 [−.15, .06]	0 [−.10, .10]	**−.12 [−.22, −.02]**	**.18 [.08, .28]**	−.04 [−.13, .06]
b	1	.07 [−.04, .18]	.08 [−.03, .18]	**.22 [.11, .32]**	.07 [−.04, .17]	−.02 [−.13, .08]	.04 [−.07, .15]	−.03 [−.13, .08]	−.02 [−.12, .09]
	2	**.11 [.001, .22]**	.06 [−.04, .16]	**.26 [.16, .36]**	.06 [−.04, .15]	−.05 [−.15, .06]	.07 [−.03, .17]	−.03 [−.13, .07]	−.01 [−.11, .09]
g	1	**.51 [.44, .58]**	**−.13 [−.22, −.04]**	**.25 [.16, .33]**	0 [−.08, .09]	.02 [−.06, .11]	**.20 [.11, .28]**	−.02 [−.11, .07]	−.01 [−.09, .07]
	2	**.43 [.33, .52]**	−.10 [−.21, .01]	−.02 [−.13, .09]	0 [−.11, .11]	−.05 [−.16, .07]	−.04 [−.16, .07]	**−.13 [−.23, −.02]**	−.02 [−.13, .08]

Note. Sampled population correlation coefficients $\hat{\rho}$ are displayed. Variables were incorporated as covariates in the model (see text for further details). Brackets indicate 95% Bayesian credibility intervals (BCI). D = probability of recognizing a statement as previously presented and probability of knowing that a distractor statement is new; d = probability of correctly remembering the (either typical or atypical) source of a statement; b = probability of guessing that an unrecognized statement is old; g = probability of guessing that a statement had been presented by the source with the typical age/gender for this statement when the source is not remembered. Contingency = perceived ratio of typical statement-source combination per source-monitoring task (averaged for statement-source combinations of both typicalities). NFC = need for cognition; FII = faith in intuition. NFC and FII were rated on a 7-point Likert scale. O = openness; C = conscientiousness; E = extraversion; A = agreeableness; N = neuroticism. Big Five were rated on a 5-point Likert scale.

Table 5 Descriptive statistics for the covariates of both experiments

Experiment	Contingency		Cognitive-processing styles		Personality traits: Big Five				
			NFC	FII	O	C	E	A	N
Experiment 1 10 Minutes	Task 1	Task 2							
M	.51	.52	4.95	4.42	3.69	3.64	3.31	3.86	2.85
SD	.16	.16	.82	.84	.66	.68	.58	.58	.75
Min	.10	.13	3.07	1.80	1.83	1.80	2	2.50	1
Max	.95	.90	7	6.13	5	5	4.33	4.83	4.83
7 days									
M	.53	.51	5.14	4.22	3.76	3.66	3.23	3.88	2.81
SD	.14	.15	.72	.90	.66	.65	.65	.60	.67
Min	.15	.20	3.50	1.80	1.17	2.20	1.33	2	1.17
Max	.88	.85	6.71	6.27	5	5	4.50	5	5
Experiment 2									
M	.52	.50	5.08	4.28	3.67	3.44	3.48	3.85	2.78
SD	.16	.13	.77	.73	.62	.54	.64	.59	.75
Min	.17	.20	1.93	2	2.33	2.33	1.83	1.83	1.33
Max	.92	.83	6.79	6.20	5	4.50	4.83	5	4.67

Note. NFC = need for cognition; FII = faith in intuition; O = openness; C = conscientiousness; E = extraversion; A = agreeableness; N = neuroticism. NFC and FII were rated on a 7-point Likert scale; Big Five were rated on a 5-point Likert scale. Contingency = perceived ratio of typical statement–source combination per source-monitoring task (averaged for statement–source combinations of both typicalities). *M* = means, *SD* = standard deviation, *Min* = minimum, *Max* = maximum. All values are aggregated across items

descriptive statistics of these covariates. We found weak but substantial negative correlations between source guessing and NFC for both tasks in the 7-day condition. For all other correlations in both conditions, the 95% BCIs included zero, revealing that the influence of any Big Five domain and FII on source guessing was unreliable. Thus, we could not identify any systematic personality-trait predictors of stereotype-based source guessing.

Regarding the within-task relationship between source memory and source guessing, we found a negative trend (10 minutes; Task 1: $\hat{\rho} = -.15$ [−.44, .14]; Task 2: $\hat{\rho} = -.19$ [−.51, .14] and 7 days; Task 1: $\hat{\rho} = -.29$ [−.57, .01]; Task 2: $\hat{\rho} = -.28$ [−.59, .07]) suggesting that source guessing becomes more stereotype based with poorer source memory (in line with experimental manipulations; Kuhlmann et al., 2016). However, the BCIs of these correlations again included zero. Additionally, the contingency judgment (proportion of typical statement–source combinations in the study phase, averaged for typically–old and typically–young statement–source combinations) for each source-monitoring task was incorporated as a covariate in the model. On the mean level (see Table 5), the contingency judgment was not biased in either condition (i.e., all $ps \geq .635$ testing against .50), except for the first task in the 7-day condition, which showed a small overestimation of typical statement–source combinations, $t(109) = 2.29$, $p < .05$, $BF_{10} = 1.27$. Correlations between this contingency judgment and source guessing were numerically in the expected positive direction in all conditions. However, these correlations were only substantial (i.e., BCI excludes zero) in the first task of the 10-minute condition and the second task of the 7-day condition. Thus, we partially replicated prior findings of the probability-matching account of source-guessing bias (Arnold et al., 2013; Bayen & Kuhlmann, 2011; Spaniol & Bayen, 2002), which posits that when people do not remember the source, they match their source-guessing behavior to the perceived contingencies between items and sources.

Experiment 2

One potential limitation of our first experiment is that participants were faced with the similar (albeit different statements and names) task twice, which could elicit reactance in the second (repeated) task as participants respond to the same type of statements. They may have actively reflected on the first task during the time interval between tasks (albeit being relatively short in the 10-minute condition) and, consequently, changed their response behavior for the second task based on a strategy, or generated an internal hypothesis about the experiment. To test whether the weak relative stability in the first experiment was caused by using the same stereotype domain twice, we conducted a second experiment that specifically tested the effect of item material that reflected distinct knowledge domains: age (analogous to Experiment 1) and gender stereotypes. This also allowed us to test whether the absolute stability in source-guessing bias found in Experiment 1 generalizes across different knowledge domains.

Method

Participants and design

Participants were 108 students of the University of Mannheim (44% psychology majors; 71% women), with a mean age of 21.47 years (SD = 2.25 years, age range: 18–26 years). Our sample size considerations and exclusion criteria were identical to those of Experiment 1; we additionally excluded participants of Experiment 1. Again, participants received course credit or monetary compensation.

Material

Ninety age-stereotypical statements were selected from the same survey study (Kuhlmann et al., 2017) as in Experiment 1, using a cutoff criterion of >3.2 for one age group and ≤2.9 for the other age group. By needing age-stereotypical statements for one task only, we were able to change the item number from 60 to 90 per task compared with Experiment 1 to increase estimation precision.

Gender-stereotypical items for the second source-monitoring task consisted of everyday statements that were normed by a subset of 86 participants (M = 21.55 years, SD = 1.85 years, age range: 18–26 years) from Experiment 1.[10] Statements reflected the domains "career" and "family" for both genders (i.e., men/career, women/career, men/family, women/family). An example statement is: "I see myself as the head of our family" (men/family). Participants either rated the statements' typicality for a man or a woman (randomly assigned; between subjects) on a 5-point Likert scale (1 = *very atypical*, 5 = *very typical*).[11] We defined a statement as typically male if the mean typicality rating for a male person was >3.2 and, at the same time, ≤3 for a female person, and vice versa for typically-female statements. We randomly selected 90 statements to be equally distributed to the four gender × domain categories. Statements per category did not differ in terms of their typicality ratings from male and female participants, all $ps \geq .50$. Each of the 90-item sets was divided into three matched subsets of 30 statements, of which two subsets served as study lists and one subset as distractors at test. We used the same source labels (names) as in Experiment 1 for both tasks.

[10] All participants of Experiment 1 completed the gender-typicality rating in Experiment 1 but we prepared and started data collection for Experiment 2 before Experiment 1 was finished.

[11] The original item set consisting of 420 statements was provided by Dr. Marie Luisa Schaper (Heinrich Heine University Düsseldorf). Statements for men and women were generated based on the domains "career" and "family" for both genders (i.e., men/career, women/career, men/family, women/family) and were originally rated on a 5-point scale (1 = *strongly male*, 5 = *strongly female*). Ninety-six statements (typically-female statements: M = 3.07, SD = .33; typically-male statements: M = 3.17, SD = .28, 48 statements from each domain) with equal typicality ratings from male and female participants and 65 additional statements (typically-female statements: M = 3.23, SD = .42; typically-male statements: M = 3.01, SD = .84; 27 statements from the career domain & 38 statements from the family domain) from the original item pool were chosen for our norming study to have a sufficient large number of typically-male and typically-female statements in our source-monitoring task. To have comparable typicality ratings and cut-off criteria between source-monitoring tasks with different material, we adjusted the response scale to a 5-point Likert scale, as used in Experiment 1, ranging from 1 = *very atypical* to 5 = *very typical* for the norming in Experiment 1.

Procedure

Participants completed both source-monitoring tasks in one session, structured like the first session of Experiment 1's 10-minute condition, with the exception that participants filled out the demographic questionnaire, the BFI and the REI during the 10-minute break. We held task order constant across participants: All participants first completed the task with age-stereotypical item material, followed by the task with gender-stereotypical item material.

Results and discussion

Model-based analyses

As in Experiment 1, we analyzed the data with the 2HTSM Submodel 4 (Bayen et al., 1996) using the Bayesian-hierarchical latent-trait approach in *TreeBUGS* (Heck et al., 2018). We obtained good model fit for both the mean (T_1, p = .45) and covariance (T_2, p = .22) structure. Group-level estimates of the four parameters are presented in Table 2. On the group level, source guessing exceeded chance level (.50) for the first task, $\Delta(g - .50) = .09$ [.05, .12], but not for the second task using gender stereotypes, $\Delta(g - .50) = .01$ [−.03, .05]. Participants changed from being biased when performing an age stereotype source-monitoring task to being unbiased in the second source-monitoring task with gender-stereotypical item material 10 minutes later.

Stability

Relative stability Table 3 presents estimates of the correlations between the four source-monitoring processes across the two tasks. Although the cross-task correlation of source guessing was again rather small, it was substantial (i.e., BCI excludes zero). Thus, participants who guessed based on age stereotypes in the first task, were also more likely to guess based on gender stereotypes in the second task (see also Fig. 2). Albeit the correlation analysis suggested credible relative stability across knowledge domains, the magnitude is again rather small and descriptively comparable to Experiment 1, in which this correlation, however, did not reach statistical credibility. This seemingly inconsistent result across our experiments is to be expected, given our recovery simulations showing that with the given sample and item sizes, only large trait-like cross-task correlations of source guessing could be recovered satisfactorily, whereas there are issues with recovering smaller correlations. Thus, there is some evidence for cross-task relative stability of knowledge reliance in source guessing, but it is less stable than it would be expected for a cognitive trait. Again, within-task split-half correlations indicated good reliability of source guessing across the first task using age stereotypes, $\hat{\rho}$ = .75 [.56, .89], and the second task using gender stereotypes, $\hat{\rho}$ = .70 [.45, .88] (see Appendix A for further details), such that poor reliability cannot explain the observed weak cross-task correlation.

Other memory and guessing parameters also showed relative stability (see Table 3). Old–new item guessing was highly correlated across tasks, again conceptually replicating Kantner and Lindsay (2012). In addition, we again replicated the results from Kantner and Lindsay with the SDT bias measure *c* for item memory (whereas adapting the bias measure *c* for source memory did not result in credible correlations across tasks; see Appendix B).

Absolute stability To account for the observed group-level differences in source guessing between tasks, we calculated the difference between each individual guessing estimate and the respective task group-level mean to relativize each individual's strength of knowledge reliance within each task. We then sampled the absolute mean difference for this mean-centered guessing score between the two tasks. The difference |Δ(Task 1 – Task 2)| was .12 [.002, .36]—again indicating some changes in source guessing between tasks (see Fig. 2). However, the absolute mean difference was only larger than .15 for 36.11% of participants (39 out of 108). Thus, albeit some participants changed their guessing behavior between tasks, the majority of participants did not and instead showed fairly stable guessing tendencies. As already indicated by diverging group-level guessing biases between tasks, on an individual level, the guessing bias was descriptively more pronounced in the first task for most individuals (n = 73, 67.59% of individuals), possibly indicating regression to the mean tendencies in the second task. Further, the somewhat larger mean change in source guessing across tasks (and for a somewhat larger proportion of participants) in comparison to Experiment 1 may be due to differential strength of age versus gender stereotypes within participants (albeit matched average stereotype strength of the item material).

Covariates of source guessing

Estimated correlations between source guessing and z-transformed covariates are presented in Table 4. We found a substantial negative correlation of source guessing with NFC and a substantial positive correlation with FII in the first task. However, this pattern did not emerge for the second task. For all other correlations, the 95% BCI included zero showing that the influence of any Big Five domain on source guessing was estimated unreliably. This indicates that a (already rather weak) cognitive trait is not necessarily linked to a personality trait, as suggested by Kantner and Lindsay (2012).

Further, there was a negative correlation between source memory and source guessing for the first task (Task 1: $\hat{\rho}$ = −.48 [−.69, −.25]; Task 2: $\hat{\rho}$ = −.08 [−.34, .19]), showing that poorer source memory is, the more source guessing becomes stereotype based (confirming the trend of Experiment 1 and replicating, e.g., Arnold et al., 2013; Kuhlmann et al., 2016). Participants' contingency judgments at the end of each task showed no bias at all (see Table 5). Participants seemed to perceive the true 50:50 contingency of statement–source combinations almost perfectly, both *p*s ≥ .116, tested against .50. As presented in Table 4, the estimated correlation of source guessing with the respective contingency judgment of each task indicated a strong relationship. Beyond the positive trends observed in Experiment 1, the results of Experiment 2 replicated the probability-matching account of source-guessing bias (Bayen & Kuhlmann, 2011; Spaniol & Bayen, 2002; hierarchical MPT: Arnold et al., 2013). People appear to match their source-guessing behavior to the perceived contingencies between items and sources, regardless of the distinct knowledge domain.

General discussion

In two experiments, we tested the state versus trait-like nature of stereotype-based source guessing. The rationale behind these experiments was derived from previous research documenting intraindividual stability of response bias in item memory and of decision-making processes. Individuals performed two independent study–test cycles of a source-monitoring task. They learned information from two person sources of different ages or gender that needed to be remembered in a later (source) memory test. Information was either typical or atypical for the source's social category (age group or gender). If individuals did not remember, they could have relied on their prior knowledge (i.e., social stereotypes) to infer the source. We examined whether, and if so, to what extent, the reliance on stereotypes is an inherent cognitive trait mirrored in stable guessing behavior across time (Experiment 1) and knowledge domains (Experiment 2). We estimated memory performance and guessing tendencies of each individual with Bayesian-hierarchical MPT models. Cross-task correlations of source guessing revealed (at best) weak positive associations across 10 minutes, seven days, and knowledge domains. In contrast, item guessing (conceptually replicating Kantner & Lindsay, 2012, 2014, with the MPT analogue of their SDT-based item response bias measure) and other source-monitoring processes showed medium to large relative stability across these manipulations. Nonetheless, individual guessing tendencies were mainly stable between tasks when measured in absolute differences. Across experiments, we observed some evidence that the perceived statement–source contingency explained individuals' guessing tendencies, but none of the personality traits and cognitive-processing styles consistently did.

Methodological challenges of cross-task relative stability in source guessing

Although we obtained positive cross-task correlations of source guessing, implying its relative stability, in all experiments and conditions, we acknowledge that these correlations were rather weak and not estimated precisely (i.e., large credibility intervals), resulting in only one substantial cross-task correlation in Experiment 2 (and in Experiment 1's extended sample with n = 164; see Supplemental Material). However, we caution not to dismiss the relative stability based on these correlational analyses due to several statistical challenges involved in this correlative measure. Although our split-half reliability analyses dismiss poor reliability of source guessing as a methodological explanation, our post hoc parameter-recovery simulations (see Supplemental Material) revealed that only large correlations (i.e., ρ = .7) could be precisely recovered in the latent-trait 2HTSM Submodel 4 with the given number of observations. Weaker correlations, in particular those of the size observed in both experiments (i.e. ρ = .30), could not be recovered with adequate power, particularly because the recovery simulations were based on the smallest number of items and participants in our experiments. The credible cross-task source-guessing correlation for the extended sample indicated that even small correlations should be recovered more precisely with an increasing number of observations. We are thus careful to not overinterpret the overlap of two of three source-guessing cross-task correlation's credibility interval with zero, as evidence against relative stability of source guessing. Nonetheless, if source guessing was a highly stable cognitive trait to a comparable extent as the response bias in recognition memory (Kantner & Lindsay, 2012, 2014) or the knowledge reliance in the recognition-heuristic use (Michalkiewicz & Erdfelder, 2016), then we should have detected such large cross-task correlation based on the sample size and item numbers in both experiments. As such, we can conclude that source guessing does not show the high relative stability that one would expect for a cognitive trait.

It is notable that we detected medium to large cross-task correlations between the item response bias parameter b, confirming its trait-like status (Kantner & Lindsay, 2012, 2014) and between other source-monitoring parameters. Compared with these parameters, variance in source guessing tended to be smaller, with the largest variance observed in the source-memory parameter d. However, variance in source guessing was overall only slightly smaller than in the item response bias parameter b, which consistently showed large cross-task correlations.

Thus, a restriction of inter-individual variance in source guessing is rather implausible as an explanation of why we observed only small source-guessing correlations. In contrast, the sparse amount of information (number of observations) underlying each parameter estimate in the 2HTSM model (Bayen et al., 1996) may explain the difficulty to precisely recover estimates of cross-task correlations in source guessing, specifically, whereas the cross-task correlations of other model parameters (including the item response bias parameter b that was of additional interest to us) could be recovered precisely. Specifically, the dependence of source guessing on various preceding parameters may render the detection of small correlations in this parameter more difficult in the data at hand. In addition, in Experiment 1, in which both source-guessing correlations were not substantial, the covariance structure of the data was overall not explained well by the model.

One potential substantive explanation why relative item-guessing stability did not generalize to source guessing may be that the conceptualization of both cognitive processes differs. Item guessing refers to a conservative versus liberal response bias—individuals inherently diverge from each other with regard to their response criterion. Source guessing refers to responding stereotypically versus atypically when (source) memory is absent. Contrary to item guessing, no task-/material-unrelated criterion such as conservative versus liberal exists for source guessing without prior knowledge activation. Source guessing cannot be investigated and interpreted in a sensible manner without the activation of schema/stereotypes as a conservative or liberal response criterion (which would just mean to respond to one source more often than the other). Although the interpretation of source guessing depends on what the sources stand for, similar findings such as probability matching have been documented across paradigms with diverse schema and stereotype domains (e.g., Bayen & Kuhlmann, 2011; Kuhlmann et al., 2016; Kuhlmann et al., 2012; Spaniol & Bayen, 2002; for an overview, see Kuhlmann & Bayen, 2016). This is why one can expect our results to be generalizable also to other knowledge-based source-guessing biases (which, however, awaits to be tested in future studies).

We are confident that there is not a general problem with estimating relative correlation stability of parameters from the 2HTSM MPT model. Indeed, our supplemental correlational analyses of the SDT response-bias measure c, in item and source memory, replicated the latent-trait MPT-model-based results for item guessing (parameter b) and source guessing (parameter g), respectively. Thus, our results are independent of the employed memory measurement model (MPT vs. SDT). Regarding item response bias, this means that

Kantner and Lindsay's (2012, 2014) findings can be robustly replicated within the MPT framework, further evidencing the (relative) stability of responding either liberal or conservative with regard to old–new recognition.

The nature of stereotype-based source guessing

Given the described inherent methodological drawbacks of estimating relative stability in source guessing, we additionally considered absolute stability in our analyses. Crucially, this absolute difference captures minimal changes in guessing tendencies and can thus be defined as an even stricter measure of stability. This measure revealed rather small changes in the knowledge reliance of source guessing within individuals. In particular, these changes were of negligible size (≤.04) in Experiment 1, in which the reliance on the same stereotypes (age) was assessed in both tasks. The somewhat larger absolute change in source guessing in Experiment 2 was to be expected, as the transition of knowledge domains (age to gender) between tasks may likely still elicit a substantial change in individuals' guessing behavior. Although we matched average stereotype strength between tasks based on the surveys, the age and gender stereotypes may be of different strength within a given individual. Nonetheless, only about a third of participants showed a substantial change of .15 or larger in source guessing between tasks in Experiment 2. Thus, in absolute terms, source guessing was fairly stable across tasks in both experiments.

Both stability measures broadly result in convergent evidence that source guessing encompasses at least to some extent trait-like features. Although correlations were weaker than one would expect for a trait, the majority of participants remained highly stable in absolute terms in their source guessing, thus fulfilling the more conservative stability criterion. In any case, these stability analyses provided a more nuanced picture of the knowledge reliance in guessing tendencies. In particular, the obtained findings for (absolute) stability of source guessing may speak to the underlying process nature—namely, whether it is a rather automatic cognitive process or under conscious control (cf. Bargh, 1994). One could speculate that source guessing may rather be an automatic cognitive process not under conscious control if it is fairly stable across tasks and, more generally, across situations. Contrary to this, prior research suggested that source guessing relies on systematic rather than on heuristic processing (e.g., Bayen & Kuhlmann, 2011; Ehrenberg & Klauer, 2005; Spaniol & Bayen, 2002; for a theoretical overview, see Johnson et al., 1993; Johnson & Raye, 2000), and in the present experiments there was also some evidence that the perceived item–source contingency influenced source guessing, which seems to rather reflect systematic processing (cf. Kuhlmann et al., 2012). However, as the two presented experiments only provide indirect evidence, future studies should consider to directly manipulate and test the (automatic versus controlled) nature of source guessing.

However, our results clearly demonstrated that the trait-like nature of source guessing is not as pronounced as for other cognitive biases and thereby diverges from other cognitive processes involved in source attributions. As the process of source guessing may be of more complex nature it may not only depend on other cognitive processes such as source memory but also on the perception of contingencies already at encoding of information (e.g., Arnold et al., 2013) and the need to integrate further knowledge (e.g., about social groups). Therefore, the nature of stereotype-based source guessing may not be entirely comparable to, but qualitatively different from, other response biases (e.g., knowledge reliance in the recognition-heuristic use; Michalkiewicz & Erdfelder, 2016) or decision-making processes (e.g., such as risk seeking; Glöckner & Pachur, 2012). The potentially more complex nature of source guessing as a cognitive trait may also explain the absence of its link with personality traits, which is present for the previously named cognitive processes.

In addition to noting that the reliance on prior knowledge may often lead to correct source attributions outside the laboratory, one could further speculate that instead of mere knowledge reliance, stereotype-based source guessing reflects a rational metacognitive bias compensating for (perceived) memory differences between sources (cf. Batchelder & Batchelder, 2008). In particular, knowledge-based source guessing might be a metacognitive compensation strategy for the sometimes found memory deficit for expected sources (i.e., inconsistency effect; e.g., Bell, Buchner, Kroneisen, & Giang, 2012; Ehrenberg & Klauer, 2005; Kroneisen & Bell, 2013; Küppers & Bayen, 2014). Even though source guessing has been found to reflect rational metacognitive strategies due to item–memory differences (e.g., Kuhlmann & Touron, 2011; Meiser, Sattler, & Von Hecker, 2007), Schaper, Kuhlmann, and Bayen (2019) recently demonstrated that knowledge-based source guessing cannot be described as a rational metacognitive bias. In fact, expectancies about the memorability of expected and unexpected item–source combinations did not predict the knowledge reliance (here: schema) in source guessing.

State determinants of source guessing

The current findings are of particular importance, as age stereotypes become internalized at a young age and then become self-stereotypes at an older age (Levy, 2009; Levy, Slade, Chung, & Gill, 2015), which under threat can lead to older adults' underperformance in cognitive and physical tasks (see Lamont, Swift, & Abrams, 2015, for a meta-analytic review). The somewhat stable reliance on stereotypes and schemas in source guessing potentially maintains or even reinforces stereotypes in general, profoundly influencing how we form impressions about others (e.g., Bell et al., 2013) and how we behave (e.g., in a legal decision-making context; Lindsay, 1994, 2014). Therefore, future studies should be inspired by research questions on how to nevertheless explore state-like features to identify specific situational determinants that prompt people to rely more or less on prior knowledge, and age stereotypes in particular, in source guessing. Former research has shown that the stereotype reliance in source guessing can be debiased through effective encoding (Arnold et al., 2013; Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012; Spaniol & Bayen, 2002) and by informing people about the negative impact of stereotypes in society (Marsh et al., 2006). However, whether these interventions also help to reduce age-stereotype-based source guessing has not yet been determined. In addition, the long-term effects of such interventions on the maintenance of knowledge reliance need to be examined. Future studies could identify additional state moderators of stereotype-based source guessing related, for instance, to the uniqueness and memorability of the statements and the personal relevance of the person's sources and depicted stereotype. Other external factors that could contribute to the knowledge reliance may be the general motivation to participate or strategy use.

In sum, our experiments are a first step towards unpacking the underlying nature of the cognitive process of source guessing. Along with future studies, a better understanding of trait and state characteristics may help to determine who is most prone to stereotype-biased memory and how this bias may be effectively prevented.

Author note Liliane Wulff and Beatrice G. Kuhlmann, Department of Psychology, School of Social Sciences, University of Mannheim, Germany.

This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences and the research-training group Statistical Modeling in Psychology (GRK 2277), funded by the German Research Foundation (DFG). Parts of this research were presented at the 2018 International Meeting of the Psychonomic Society in Amsterdam, the Netherlands, the 2018 Tagung experimentell arbeitender Psychologen (TeaP) in Marburg, Germany, and the 2017 Doctoral Workshop of Experimental Psychology (A-Dok) in Ulm, Germany. Data and results for both experiments presented here are available at https://osf.io/mjcve/?view_only=058c26e0508e4d699020ffa96db276cd; the materials of the experiments are available from the corresponding author upon request. The experiments were not preregistered.

We thank Daniel Heck for his active support with regard to Bayesian data analysis. We are grateful to Marie Luisa Schaper for providing the statement material used in Experiment 2. We thank Edgar Erdfelder for helpful input on the research idea of this project and Michelle Dörnte, Helena Horn, Alina Kias, Mario Kolb, Franziska Leipold, Rebecca Maurer, Jennifer Panitz, Theresa Pfeiffer, Ronja Reinhardt, and Alin Toja for their assistance with data collection. We thank Jan Rummel for the opportunity to recruit participants in his lab and the student assistants for their help in collecting data at Heidelberg University.

Appendix A

In order to examine the stability of stereotype reliance source guessing across tasks (time), we first must ensure that we have a reliable and stable measure of source guessing within each source-monitoring task. To test this, we used the split-half method and created two different lists per test phase from each task consisting of 60 test statements each for Experiment 1. We assigned test statements to the lists based on their statement-source typicality and temporal order in which they appeared in the test phase. Specifically, the chronologically first test statement of each statement-source combination (e.g., typically–young statement originally presented by the source "Schmidt") was assigned to the first list, the second statement of each combination was assigned to the second list and so on for all statement-source combinations (3 sources ["old" source vs. "young" source vs. new] × 2 statement typicalities [old vs. young]). Thus, each list contained an equal number of statements from each statement-source combination. We chose this procedure to have comparable lists (i.e., test halves) regarding their content and testing time. We estimated parameters for the two tests lists of each task with the latent-trait approach, including estimation of the correlations of source-guessing probabilities between lists. We obtained a good model fit for the mean but, again, a less satisfactory model fit for the covariance structure (see Table 6).

For Experiment 2, data preprocessing for this split-half analysis was analogous to Experiment 1 except that we excluded the first statement of each statement-source combination due to the odd number of statements per combination in the item set in general (90 items in total; six statement-source combinations each containing 15 statements split into seven statements per half).

Appendix B

Signal detection theory based analyses of relative stability

We calculated the signal detection theory (SDT) response-bias measure c for old-new item recognition as follows (Macmillan, 1993):

$$c = -0.5(z_{\text{Hit Rate}} + z_{\text{False-Alarm Rate}})$$

Positive values indicate a bias to respond "new" and negative values indicate a bias to respond "old." Before we calculated hits and false-alarm rates, we added .5 to hits and false alarms and 1 to signal and noise trials to correct for perfect hit rates of 1 and absent false-alarm rates of 0 in the data (cf. Hautus, 1995).

We calculated the response-bias measure c for source memory analogously with our conceptualization of hits and false alarms in the source attributions following Dodson, Bawa, and Slotnick (2007). That is, we selected only those items that had been recognized as "old" and for which a source attribution was thus made. We conceptualized source memory as following two distributions: one for (recognized) items that were presented with the age-typical source and one for (recognized) items that were presented with the age-atypical source; analogous to the old-item and new-item distributions for item recognition. Positive values indicate a bias to respond the "atypical source," negative values indicate a bias to respond the "typical source." Results for the SDT analysis can be retrieved from Table 7.

Table 6 Model-fit statistics for split-half reliability analyses of both experiments

Experiment	List	Model-fit indices	
		T_1 (Mean)	T_2 (Covariance)
Experiment 1			
10 minutes	1	.305	0
	2	.345	.003
7 days	1	.521	.103
	2	.292	.005
Experiment 2			
	1	.695	.417
	2	.653	.033

Note. Values represented the respective posterior predictive p-value for the mean (T_1) and covariance structure (T_2) of the data (proposed by Klauer, 2010). Good model fit is indicated by $p > .05$

Table 7 Signal-detection theory-based analyses of response bias for item and source memory

Experiment	Cognitive process	Task	Bias c	Cross-task correlation
Experiment 1				
10 minutes	Item memory	1	.08 (.31)	$r(114) = .61$, $p < .001$, $BF_{10} = 3e^{10}$
		2	.16 (.35)	
	Source memory	1	−.13 (.35)	$r(114) = .14$, $p = .08$, $BF_{10} = .32$
		2	−.11 (.33)	
7 days	Item memory	1	.07 (.34)	$r(110) = .26$, $p = .003$, $BF_{10} = 8.97$
		2	.20 (.31)	
	Source memory	1	−.12 (.34)	$r(110) = .02$, $p = .44$, $BF_{10} = .12$
		2	−.004 (.33)	
Experiment 2				
	Item memory	1	.03 (.27)	$r(108) = .62$, $p < .001$, $BF_{10} = 1.28e^{10}$
		2	.20 (.33)	
	Source memory	1	−.21 (.42)	$r(108) = .22$, $p = .01$, $BF_{10} = 1.69$
		2	−.16 (.36)	

Note. Mean item-memory and source-memory bias values per task. Positive values of item-memory c refer to guessing that an item was "new," negative values of c refer to guessing that an item was "old." Positive values of source-memory c refer to guessing the "atypical" source, negative values of c refer to guessing the "typical" source. Standard deviations are reported in parentheses. Test statistic = Pearson product-moment correlation between bias measures of both tasks per experiment/condition; p values refer to one-tailed testing; Bayes Factors BF_{10} were computed with *JASP* (JASP Team, 2018) and are interpreted as substantial to strong evidence in favor of the alternative hypothesis H_1 for item-memory bias and as substantial evidence in favor of the absence of an effect for source-memory bias following Jeffreys (1961)

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, *93*(2), 203–231. doi:https://doi.org/10.1037/0033-2909.93.2.203

Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin & Review*, *20*(2), 326–333. doi:https://doi.org/10.3758/s13423-012-0342-7

Bargh, J. A. (1994). The four horseman of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (pp. 1–40). Hillsdale, NJ: Erlbaum. doi: https://doi.org/10.1007/s00572-005-0022-9

Batchelder, W. H., & Batchelder, E. (2008). Metacognitive guessing strategies in source monitoring. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 211–244). New York, NY: Psychology Press.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. doi:https://doi.org/10.3758/BF03210812

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language*, *64*(1), 1–17. doi:https://doi.org/10.1016/j.jml.2010.09.001

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 197–215. doi:https://doi.org/10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G V, Dupuis, S. E., & Yang, C.-L. (2000). The use of schematic knowledge about sources in source-monitoring. *Memory & Cognition*, *28*(3), 480–500. doi:https://doi.org/10.3758/BF03198562

Bell, R., Buchner, A., Kroneisen, M., & Giang, T. (2012). On the flexibility of social source memory: A test of the emotional incongruity hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1512–1529. doi: https://doi.org/10.1037/a0028219

Bell, R., Giang, T., Mund, I., & Buchner, A. (2013). Memory for reputational trait information: Is social-emotional information processing less flexible in old age? *Psychology and Aging*, *28*(4), 984–995. doi: https://doi.org/10.1037/a0034266

Bell, R., Mieth, L., & Buchner, A. (2017). Emotional memory: No source memory without old–new recognition. *Emotion*, *17*(1), 120–130. doi:https://doi.org/10.1037/emo0000211

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116–131. doi:https://doi.org/10.1037/0022-3514.42.1.116.

Carter, J. D., Hall, J. A., Carney, D. R., & Rosip, J. C. (2006). Individual differences in the acceptance of stereotyping. *Journal of Research in Personality*, *40*(6), 1103–1118. doi:https://doi.org/10.1016/j.jrp.2005.11.005

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology, 67*(2), 319–333. doi:https://doi.org/10.1037/0022-3514.67.2.319

Crawford, M. T., & Skowronski, J. J. (1998). When motivated thought leads to heightened bias: High need for cognition can enhance the impact of stereotypes on memory. *Personality and Social Psychology Bulletin*, *24*(10), 1075–1088. doi:https://doi.org/10.1177/01461672982410005

Danner, D., Rammstedt, B., Bluemke, M., Treiber, L., Berres, S., Soto, C., & John, O. (2016). Die deutsche Version des Big Five Inventory 2 (BFI-2) [The German version of the Big Five Inventory 2]. *Zusammenstellung Sozialwissenschaftlicher Items Und Skalen*. doi:https://doi.org/10.6102/zis247

Dodson, C. S., Bawa, S., & Slotnick, S. D. (2007). Aging, source memory, and misrecollections. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(1), 169–181. doi:https://doi.org/10.1037/0278-7393.33.1.169

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology*, *41*(4), 369–387. doi:https://doi.org/10.1016/j.jesp.2004.08.001

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive-experiential and analytical-rational thinking styles. *Journal of Personality and Social Psychology*, *71*(2), 390–405. doi:https://doi.org/10.1037/0022-3514.71.2.390

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi:https://doi.org/10.3758/BF03193146

Fiske, S. T. (1989). Examining the role of intent: Toward understanding its role in stereotyping and prejudice. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 253–283). New York, NY: Guilford Press.

Flynn, F. J. (2005). Having an open mind: The impact of openness to experience on interracial attitudes and impression formation. *Journal of Personality and Social Psychology*, *88*(5), 816–826. doi:https://doi.org/10.1037/0022-3514.88.5.816

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472. doi: https://doi.org/10.1214/ss/1177011136

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517. doi:https://doi.org/10.1037/0022-3514.60.4.509

Glöckner, A., & Pachur, T. (2012). Cognitive models of risky choice: Parameter stability and predictive accuracy of prospect theory. *Cognition*, *123*(1), 21–32. doi:https://doi.org/10.1016/j.cognition.2011.12.002

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. doi:https://doi.org/10.1037/0003-066X.48.1.26

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90. doi:https://doi.org/10.1093/acprof:oso/9780199744282.003.0003

Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d'*. *Behavior Research Methods, Instruments & Computers, 27*(1), 46–51. doi:https://doi.org/10.3758/BF03203619

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. doi:https://doi.org/10.3758/s13428-017-0869-7

Hicks, J. L., & Cockman, D. W. (2003). The effect of general knowledge on source memory and decision processes. *Journal of Memory and Language*, *48*(3), 489–501. doi:https://doi.org/10.1016/S0749-596X(02)00537-5

Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, *59*(1), 21–47. doi:https://doi.org/10.1007/BF02294263

JASP Team. (2018). JASP (Version 0.10.0) [Computer software]. Retrieved from https://jasp-stats.org/

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, England: Oxford University Press.

John, O. P., & Srivastava, S. (1999). The Big Five Trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York, NY: Guilford Press.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. doi:https://doi.org/10.1037/0033-2909.114.1.3

Johnson, M. K., & Raye, C. L. (2000). Cognitive and brain mechanisms of false memories and beliefs. In D. L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 35–86). Cambridge, MA: Harvard University Press.

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition*, *40*, 1163–1177. doi:https://doi.org/10.3758/s13421-012-0226-0

Kantner, J., & Lindsay, D. S. (2014). Cross-situational consistency in recognition memory response bias. *Psychonomic Bulletin & Review*, *21*, 1272–1280. doi:https://doi.org/10.3758/s13423-014-0608-3

Keller, J., Bohner, G., & Erb, H.-P. (2000). Intuitive und heuristische Urteilsbildung–verschiedene Prozesse ? Präsentation einer deutschen Fassung des "Rational-Experiential Inventory" sowie neuer Selbstberichtskalen zur Heuristiknutzung [Intuitive and heuristic judgment—different processes? Presentation of a German version of the Rational–Experiential Inventory and of new self-report scales of heuristic use]. *Zeitschrift für Sozialpsychologie*, *31*(2), 87–101. doi:https://doi.org/10.1024//0044-3514.31.2.87

Klauer, K. C. (2010). Hierarchical multinomial processing tree models—A latent-trait approach. *Psychometrika*, *75*(1), 70–98. doi: https://doi.org/10.1007/s11336-009-9141-0

Klauer, K. C., & Kellen, D. (2010). Toward a complete decision model of item and source recognition: A discrete-state approach. *Psychonomic Bulletin & Review*, *17*(4), 465–478. doi:https://doi.org/10.3758/PBR.17.4.465

Kroneisen, M., & Bell, R. (2013). Sex, cheating, and disgust: Enhanced source memory for trait information that violates gender stereotypes. *Memory*, *21*(2), 167–181. doi:https://doi.org/10.1080/09658211.2012.713971

Kuhlmann, B. G., & Bayen, U. J. (2016). Metacognitive aspects of source monitoring. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 149–168). New York, NY: Oxford University Press.

Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. *Psychology and Aging*, *31*(8), 875–889. doi:https://doi.org/10.1037/pag0000140

Kuhlmann, B. G., Kornadt, A. E., Bayen, U. J., Meuser, K., & Wulff, L. (2017). Multidimensionality of younger and older adults' age stereotypes: The interaction of life domain and adjective dimension. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *72*(3), 436–440. doi:https://doi.org/10.1093/geronb/gbv049

Kuhlmann, B. G., & Touron, D. R. (2011). Older adults' use of metacognitive knowledge in source monitoring: Spared monitoring but impaired control. *Psychology and Aging*, *26*(1), 143–149. doi:https://doi.org/10.1037/a0021055

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1365–1376. doi: https://doi.org/10.1037/a0028147

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *Quarterly Journal of Experimental Psychology*, *67*(10), 2042–2059. doi:https://doi.org/10.1080/17470218.2014.904914

Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and Aging*, *30*(1), 180–193. doi:https://doi.org/10.1037/a0038586

Levy, B. R. (2009). Stereotype embodiment: A psychosocial approach to aging. *Current Directions in Psychological Science*, *18*(6), 332–336. doi:https://doi.org/10.1111/j.1467-8721.2009.01662.x

Levy, B. R., Slade, M. D., Chung, P. H., & Gill, T. M. (2015). Resiliency over time of elders' age stereotypes after encountering stressful events. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *70*(6), 886–890. doi:https://doi.org/10.1093/geronb/gbu082

Lindsay, D. S. (1994). Memory source monitoring and eyewitness testimony. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 27–55). New York, NY: Cambridge University Press. doi: https://doi.org/10.1017/CBO9780511759192.003

Lindsay, D. S. (2014). Memory source monitoring applied. In T. Perfect and D. S. Lindsay (Eds.), *The SAGE handbook of applied memory* (pp. 59–75). London, UK: Sage. doi: https://doi.org/10.4135/9781446294703.n4

Macmillan, N. A. (1993). Signal detection theory as data analysis method and psychological decision model. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 21–57). Hillsdale, NJ: Erlbaum.

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, *66*(1), 37–47. doi:https://doi.org/10.1037/0022-3514.66.1.37

Malejka, S., & Bröder, A. (2016). No source memory for unrecognized items when implicit feedback is avoided. *Memory & Cognition*, *44*(1), 63–72. doi:https://doi.org/10.3758/s13421-015-0549-8

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). Gender and orientation stereotypes bias source-monitoring attributions. *Memory*, *14*(2), 148–160. doi:https://doi.org/10.1080/09658210544000015

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. doi:https://doi.org/10.3758/s13428-011-0168-7

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E. J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205–235. doi:https://doi.org/10.1007/s11336-013-9374-9

McCrae, R. R., & Costa, P. T., Jr. (2008). Empirical and theoretical status of the five-factor model of personality traits. In G. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (pp. 273–294). Thousand Oaks, CA: SAGE Publications.

Meiser, T., Sattler, C., & Von Hecker, U. (2007). Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *The Quarterly Journal of Experimental Psychology*, *60*(7), 1015–1040. doi:https://doi.org/10.1080/17470210600875215

Michalkiewicz, M., & Erdfelder, E. (2016). Individual differences in use of the recognition heuristic are stable across time, choice objects, domains, and presentation formats. *Memory & Cognition*, *44*(3), 454–468. doi:https://doi.org/10.3758/s13421-015-0567-6

Michalkiewicz, M., Minich, B., & Erdfelder, E. (2019). *Explaining individual differences in fast-and-frugal decision making: Need for cognition, faith in intuition, and the recognition heuristic*. Manuscript under review.

Mitchell, K., & Johnson, M. (2000). Source monitoring: Attributing mental experiences. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 179–195). New York, NY: Oxford University Press.

Perlini, A. H., & Hansen, S. D. (2001). Moderating effects of need for cognition on attractiveness stereotyping. *Social Behavior and Personality: An International Journal*, *29*(4), 313–321. doi:https://doi.org/10.2224/sbp.2001.29.4.313

Rammstedt, B., Danner, D., Soto, C. J. & John, O. P. (2018). Validation of the short and extra-short forms of the Big Five Inventory-2 (BFI-2) and their German adaptations. *European Journal of Psychological Assessment*. doi:https://doi.org/10.1027/1015-5759/a000481

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137–145. doi: https://doi.org/10.1016/j.jrp.2008.12.015

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metamemory expectancy illusion and schema-consistent guessing in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(3), 470–496. doi: https://doi.org/10.1037/xlm0000602

Sherman, J. W., & Bessenoff, G. R. (1999). Stereotypes as source-monitoring cues: On the interaction between episodic and semantic memory. *Psychological Science*, *10*(2), 106–110. doi:https://doi.org/10.1111/1467-9280.00116

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology*, *54*(1), 167–183. doi:https://doi.org/10.1016/j.jmp.2009.06.007

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81. doi:https://doi.org/10.1016/j.jrp.2017.02.004

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 631–651. doi: https://doi.org/10.1037//0278-7393.28.4.631

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. doi: https://doi.org/10.1111/1467-9868.00353

Stroebe, W., & Insko, C. A. (1989). Stereotype, prejudice, and discrimination: Changing conceptions in theory and research. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and prejudice: Changing conceptions* (pp. 3–34). New York, NY: Springer. doi: https://doi.org/10.1007/978-1-4612-3582-8_1

Is Knowledge Reliance in Source Guessing Automatic or Controlled? Evidence from Divided Attention and Aging

Liliane Wulff and Beatrice G. Kuhlmann

University of Mannheim

Author Note

Liliane Wulff and Beatrice G. Kuhlmann, Department of Psychology, School of Social Sciences, University of Mannheim, Germany.

The research reported in this article was supported by the University of Mannheim's Graduate School of Economic and Social Sciences. The research plan reported in this article was preregistered on the Open Science Framework (OSF). The research plan, data and corresponding analyses can be retrieved from `https://osf.io/2ezrn/?view_only=ce2d6ad3949c4ed39e0ca664423ebe21`. We declare that we report any deviations that occurred from our preregistered research plan. Parts of this research were presented at the 2019 *Tagung experimentell arbeitender Psychologen* (TeaP) in London, UK, and the 2019 *Doctoral Workshop of Experimental Psychology* (A-Dok) in Mannheim, Germany. We thank Katja Bitz, Alina Kias, Franziska Leipold, Jule Schilling, Paula Schmelzer, Pia Vogel, and Selina Zajdler for their assistance with the data collection. We thank Jan Rummel for the opportunity to recruit participants in his lab and the student assistants for their help in collecting data at Heidelberg University.

Correspondence concerning this article should be addressed to Liliane Wulff, M.Sc. Psychology, Department of Psychology, School of Social Sciences, University of Mannheim, D-68163 Mannheim, Germany. +49(0)621-181-3463, E-mail: wulff@psychologie.uni-mannheim.de

Abstract

People often lack source memory for the original context of information (e.g., who presented it) and thus have to guess the source. According to the probability-matching account, this guessing process is either determined by general knowledge (e.g., stereotypes) or by the item-source contingency (if perceived during encoding). In this preregistered source-monitoring experiment, we examined whether each type of source guessing is rather automatic (i.e., resource independent) or controlled (i.e., resource dependent). Individuals learned and later needed to remember age-typical or -atypical statements for two different-aged sources. To examine the resource dependence, attention at test was divided for half of the younger participants (18−26 years). In addition, older adults (61−80 years) with a natural decline in cognitive resources were tested. To examine the probability-matching account, both age groups either received specific source information (age) already at encoding or only at test (worsening item-source contingency perception at encoding). Memory and guessing processes were estimated on the group and individual level using multinomial modeling. Neither stereotype- nor contingency-based source guessing depended on cognitive resources—whether they were depleted due to the experimental manipulation or natural aging. In line with the probability-matching account, younger adults based their guessing on stereotype knowledge only when the perception of the item-source contingency was hindered. Older adults strongly guessed stereotype-based and were not able to counteract via accurate contingency representations. The results underpin the role of encoding for a successful counteraction of stereotype-based guessing and are discussed with regard to their applicability for future source-attribution attempts.

Keywords: source monitoring, source guessing, age stereotypes, cognitive aging, Bayesian-hierarchical multinomial modeling

Is Knowledge Reliance in Source Guessing Automatic or Controlled?
Evidence from Divided Attention and Aging

In our everyday life, we almost constantly process (social) information. Oftentimes, this information needs to be accessed at a later point and can then be used to draw conclusions about its trustworthiness. For instance, after interacting with a conversational partner, one must remember the person (i.e., "source") who told you something. Due to the vast amount of processed information, we often cannot remember the source. In this case, general knowledge structures (stereotypes, schemas) can be used to infer the source (Johnson, Hashtroudi, & Lindsay, 1993). For instance, one may simply rely on attitudes towards a source's age group. The separate analysis of memory and guessing processes in source monitoring via multinomial processing tree (MPT) models (Bayen, Murnane, & Erdfelder, 1996) revealed that when the source could not be remembered, people guessed that information (e.g., "I am discontent with my health") that was prototypical for different age groups (here: older adults) was presented by the source whose age was consistent (here: 70-year-old source) with the information (Kuhlmann, Bayen, Meuser, & Kornadt, 2016).

Thus, expectations about members of (social) groups that exist prior to the conversation (or the experimental task) are used to infer the source when source memory fails (cf. Kuhlmann & Bayen, 2016). This knowledge can, of course, sometimes be a valid source cue and may be even diagnostic for future social categorizations (Klauer & Ehrenberg, 2005), but it is misleading whenever a person does not behave prototypically for its social group. These "costs" of knowledge application as a lack of individuation may in this case lead to biased impression formation as these persons are attributed to be compliant and share specific attitudes with a social group that do not reflect their actual attitudes and behaviors. With regard to age stereotypes, a common and repeated misattribution of characteristics due to biased guessing may have severe consequences for affected individuals when being confronted with negative connotations about the own age group (e.g., Lamont, Swift, & Abrams, 2015). And one can only speculate whether this misattribution may foster the long-term maintenance of (age) stereotypes in society (e.g., Levy, Slade, Chung, & Gill, 2015).

General Knowledge-Based Versus Individuating Contingency-Based Source Guessing

It is, therefore, particularly important to examine the circumstances under which mostly unbiased source guessing can be achieved. Spaniol and Bayen (2002) proposed a theoretical framework to account for the application of different knowledge structures in source guessing—the *probability-matching account*. Accordingly, individuals adjust their guessing behavior to the actual contingency between attributes reflected in the item material and the (social-) group membership of the source (item-source contingency) that they perceive during encoding whenever this is possible. Thus, noticing the ratio with which each source has presented certain information either typical or atypical for its (social) group, prompts individuals to use this situation-specific knowledge in source guessing (e.g., Bayen & Kuhlmann, 2011; Kuhlmann, Vaterrodt, & Bayen, 2012). Only when the item-source contingency remains undetected, individuals rely on broader knowledge structures such as stereotypes (e.g., Bayen & Kuhlmann, 2011; Ehrenberg & Klauer, 2005; Klauer & Ehrenberg, 2005; Kuhlmann et al., 2012; Spaniol & Bayen, 2002).

Guessing based on individuating information (item-source contingencies), is thus likely when individuals are explicitly instructed to focus on the sources (in addition to the items; Kuhlmann et al., 2012), when they receive specific information about the sources (e.g., their age) already at encoding (Bayen & Kuhlmann, 2011), and when their attention is fully directed towards the item-source contingency (Bayen & Kuhlmann, 2011). Thus, prior studies have identified the encoding process as a crucial prerequisite for adequate source guessing reflecting the experimental item-source contingency.

The Nature of Knowledge Reliance in Source Guessing

But what is the nature of these knowledge-based source-guessing processes? Is it rather automatic or controlled? Following Bargh (1994), the automaticity versus controllability of a cognitive process can be disentangled in terms of its awareness, intentionality, conscious inhibition, and resource dependence. Results from two studies examining the resource dependence of knowledge-based source guessing suggest that it may be rather an automatic process, unaffected by cognitive load at test (Klauer & Ehrenberg, 2005; Marsh, Cook, & Hicks, 2006). However, Marsh et al. (2006) found that stereotype-based guessing can be consciously controlled by reflecting on the harmful impact of stereotypes. In line with this, results from another study examining the time course of source guessing point towards a controlled nature of pre-experimental knowledge applications (Spaniol & Bayen, 2002). More specifically, schema-based source guessing only emerged when a longer response time was allowed at test. The aforementioned contingency-based overwriting of stereotype-/schema-biased guessing additionally suggests its controlled nature (e.g., Bayen & Kuhlmann, 2011; Ehrenberg & Klauer, 2005; Spaniol & Bayen, 2002).

Overall, evidence thus far points towards a controlled nature of knowledge-based source guessing. However, a contrasting conclusion emerges when the scope of stereotyping is broadened. The nature of social categorization based on stereotyping can be described as rather automatic (i.e., operating outside of awareness; e.g., Devine, 1989; see Monteith, Woodcock, & Lybarger, 2013, for an overview). Automatic features in age stereotyping have also been documented (Gonsalkorale, Sherman, & Klauer, 2014; Lepore & Brown, 1997; Perdue & Gurtman, 1990, but see Chasteen, Schwarz, & Park, 2002). Furthermore, as Gilbert and Hixon (1991) already suggested, stereotype activation and application may differ in terms of their resource needs. The activation and counteraction may be resource dependent whereas the application may then be resource independent (e.g., Bodenhausen, Macrae, & Sherman, 1999; Hamilton & Sherman, 1994, but see Rivers, Sherman, Rees, Reichardt, & Klauer, 2019). Thus, if one defines knowledge-based source guessing as reflecting stereotype application, the social-cognition literature contradicts what has been predominantly reported in the source-monitoring literature regarding its nature.

Whereas the few existing studies about the nature of source guessing focused on stereotypes/schemas, even less is known about the nature of contingency-based source guessing. Notably, Bayen and Kuhlmann (2011) and Ehrenberg and Klauer (2005) manipulated attention at encoding and demonstrated that the acquisition of contingency knowledge at encoding is resource dependent. That is, biased source guessing was observed when attention was divided but not when attention was unimpeded at encoding. Whether the overwriting of stereotype influences in the moment of source guessing (at test) through contingency knowledge is also resource dependent remains unexplored, however.

Overview of Experiment and Research Questions

Whereas research on source monitoring points towards a rather controlled, resource-dependent application of pre-experimental knowledge in source guessing and has not examined the nature of peri-experimental knowledge-based guessing, evidence from the social-cognition literature suggests a more automatic, resource-independent application but resource-dependent counteraction of this knowledge reliance (e.g., Klauer & Ehrenberg, 2005, but see Rivers et al., 2019). We thus aimed to more directly examine the resource dependence of both forms of knowledge reliance in source guessing. This is often done by implementing a second task that needs to be performed while, at the same time, focusing on a primary task (as done, for instance, by Klauer & Ehrenberg, 2005). But cognitive load may also naturally emerge with increasing age due to a general decline in available cognitive resources (e.g., Park, 2000). If one then defines aging as natural divided attention, the resource-(in)dependent nature of the knowledge reliance in source guessing can be studied across age groups. Importantly, none of the aforementioned studies examined the nature of knowledge reliance based on age stereotypes in both younger and older adults.

The current study was designed to shed light on these still unanswered questions about the nature of knowledge application in source guessing by varying specific source information at encoding and by using dual tasks at retrieval (which has rarely been done so far; Klauer & Ehrenberg, 2005; Marsh et al., 2006) and aging as natural divided attention. That is, individuals of both ages performed a source-monitoring task with age-stereotypical item material and received specific information about the sources (ages) either already at encoding or only at retrieval. Additionally crossed with this manipulation, half of the younger adults performed a secondary task at the source-monitoring test. To be explicit, the primary interest was to study the nature of source guessing and not age differences. We thus adapted encoding times for older adults to compensate for age differences in source-monitoring processes and optimize accurate contingency perception; see *Procedure* section). Note that the now following hypotheses constantly refer to source guessing as measured by the *two-high-threshold multinomial model of source monitoring* (2HTSM; Bayen et al., 1996).[1]

First, we hypothesize to replicate the probability-matching account (Spaniol & Bayen, 2002) in younger adults. That is, we expect source guessing to be based on the learned item-source contingency whenever participants can easily perceive it (source age provided at encoding). Otherwise, when the source age is provided at test and contingency detection should thus be difficult, source guessing should be based on stereotype knowledge.

Second, we test the nature of stereotype-based source guessing by comparing younger adults with full and divided attention in the retrieval conditions—those conditions in which we predict stereotype-based guessing. Unaffected (or even increased) knowledge-based source guessing under divided attention would imply an automatic process. Decreased knowledge-based source guessing under divided attention, however, would imply a controlled process.

Third, we test the nature of contingency-based source guessing by comparing younger adults with full and divided attention in the encoding conditions—those conditions in which we predict contingency-based guessing. Unaffected contingency-based source guessing would imply an automatic counteraction of (stereotype-based) source guessing. If the latter corresponds to .50 for individuals with full attention and exceeds .50 for individuals with divided attention, this would imply a controlled counteraction.

We further test the nature of contingency-based guessing by examining whether older adults are still able to counteract stereotype-based source guessing with contingency knowledge. If older adults adjust source guessing to this knowledge, it would imply an automatic counteraction. If they do not adjust guessing to this knowledge, it would imply a controlled counteraction.

Methods

We preregistered the experiment's research plan on the Open Science Framework (`https://osf.io/2ezrn/?view_only=ce2d6ad3949c4ed39e0ca664423ebe21`). The experiment was programmed with the open-source software program *OpenSesame* (Mathôt, Schreij, & Theeuwes, 2012).

Design and Participants

For younger adults, the design was a 2 (dual-task condition: full vs. divided attention at test; between-subjects) × 2 (encoding condition: age information at encoding vs. at test; between-subjects) factorial. For older adults, we only manipulated the encoding condition (encoding condition: age information at encoding vs. at test; between-subjects). For reasons of convenience, we introduce a notation to abbreviate the six experimental conditions: *YA* (younger adults) and *OA* (older adults) refer to the participants' age group, *enc* (encoding) and *ret* (retrieval) represent the encoding-condition manipulation (i.e., whether the specific source information was revealed before the study or test phase), and *full* (full attention) and *div* (divided attention) correspond to the dual-task manipulation at test (i.e., whether participants performed a parallel task or not).

For our analyses based on Bayesian-hierarchical MPT modeling, an *a priori* power analysis is neither straight-forward nor appropriate for a Bayesian approach. As we were mainly interested in group-level comparisons between experimental conditions, we planned the sample size on the appropriate frequentists equivalent on an aggregate group level using *multiTree* (Moshagen, 2010). As our experiment was motivated by the nature of source guessing, we defined the comparison of younger adults with full and divided attention at test as crucial for the determination of the final sample size. Based on Kuhlmann et al. (2016) who used comparable item material and presentation times during encoding for their younger-adults sample, we set the population-level parameters for item memory (D) to .70, for source memory (d) to .60, for old-new item guessing (b) to .45, and for source guessing (g) to .60. We additionally assumed a decrease in item-memory and source-memory performance of .10 in the dual-task condition.[2] To detect a between-subjects difference of .10 in source guessing with a power of .80 and ($\alpha = .05$), a minimum of 3037 observations per condition (i.e., 34 participants $\times$ 90 items) is required. To fulfill our counterbalance constraints, we increased the sample size to 36 participants per condition (3240 observations).

Younger adults were recruited via an electronic system or ads from the University of Mannheim and Heidelberg. Older adults were recruited from our participant database, originally recruited via local newspaper ads and snowballing. Participants of both age groups were only invited for this experiment if they had not participated in previous age-stereotype based source-monitoring experiments. Further, individuals were excluded if they did not meet the age criteria (based on a survey study for the used item material, Kuhlmann et al., 2016) of 18−26 years (younger adults) or 60−84 years (older adults), if their native language (i.e., learned before the age of six) was not German, and if they were colorblind (due to different colored response options in the memory test). To avoid confounds with pathological cognitive changes, participants with a history of stroke, heart attack, severe brain injury/trauma, alcohol/drug/substance abuse, any neurological disorder (e.g., epilepsy, Alzheimer's, or Parkinson disease), current (i.e., past 6 months) diagnosis of major depression, and regular use of benzodiazepines or antidepressants were excluded. Our preregistered performance-related exclusions in the dual-task conditions based on the mean accuracy in the secondary task (< 2 SD below respective condition's mean) were not necessary.

In total, we tested 218 eligible participants, 144 younger adults (range = 18−26 years, $M = 20.90$, $SD = 2.25$) and 74 older adults (range = 61−80 years, $M = 70.88$, $SD = 5.52$).[3] Due to lower rates of university education in former times in Germany, it comes as no surprise that older adults had, on average, completed slightly fewer years of formal education ($M = 13.47$, $SD = 2.16$) than younger adults ($M = 14.06$, $SD = 1.67$), $t(215) = 2.22$, $p = .027$, $BF_{10} = 1.54$, tested two-sided.[4] The additionally assessed verbal abilities with the *SASKA* vocabulary test (Riegel, 1967) showed that older adults ($M = .87$, $SD = .12$) outperformed younger adults ($M = .65$, $SD = .15$), $t(216) = 11.15$, $p < .001$, $BF_{10} = 7.06\mathrm{e}^{19}$, tested two-sided. More details about the sample composition can be retrieved from Table 1.

Materials

The item set consisted of 90 German statements reflecting everyday behaviors or general attitudes typical for younger versus older adults. These statements were created taking the multidimensionality of age stereotypes into account (Gluth, Ebner, & Schmiedek, 2010; Kornadt & Rothermund, 2011) in a survey study by Kuhlmann, Kornadt, Bayen, Meuser, and Wulff (2017). In this study, younger and older participants initially rated 368 statements with regard to their typicality for either a young or an old target person on a 5-point Likert scale (1 = "very atypical", 2 = "atypical", 3 = "neither typical nor atypical", 4 = "typical", 5 = "very typical"). We selected those statements that were rated as "typical" ($M \geq 3$) for one age group and "atypical" ($M < 3$) for the other by both participant age groups. From this item pool, we created three lists of 30 statements each (including 15 typically-young and 15 typically-old statements) from which two lists served as to-be-learned and one as distractor list for the (source-) memory test. Assignment of lists as study or distractor list was counterbalanced between participants. (A)Typicality of the statements did not differ between lists, all ps $\geq .36$. The two common German last names "Müller" and "Schneider" (age- and gender-neutral) served as sources. Sources were described to be either 23 or 70 years old (as these were the ages participants reported thinking of when rating the typicality of statements in the survey study; Kuhlmann et al., 2017).

Procedure

Participants were tested in homogeneous age groups (up to eight persons for younger adults; up to three persons for older adults to assist with computer handling if necessary) in separate computer cubicles. After participants provided written informed consent, they were randomly assigned to the experimental conditions in the order they came to the lab. A schematic illustration of the procedure can be retrieved from Figure A1 of Appendix A.

All participants first performed the digit-symbol substitution task (*DSST;* Wechsler, 1981), a paper-pencil processing-speed test, timed by the experimenter after 90 seconds. Following this, the computerized experiment started and all participants performed a source-monitoring task with age-stereotypical item material which procedure only differed based on condition-based adjustments. Participants were initially informed that they will process everyday statements from either a young or an old person source labeled with common German last names ("Müller" vs. "Schneider"). We instructed them to memorize the statements for a later test but refrained from an explicit source-memory instruction. Participants learned 60 statements (preceded by two additional statements as a primacy and followed by two additional statements as recency buffer) of which 30 statements were typical for a person of younger age and 30 statements were typical for a person of older age (typicality ratings based on the survey study from Kuhlmann et al., 2017). Each of the two sources presented one of the aforementioned lists with half typically-young and half typically-old statements, and the third list served as a distractor (counterbalanced assignment of subsets to study and distractor lists and sources). A 250 ms blank screen followed each statement. During the retention interval, younger participants practiced the tone-monitoring task (irrespective of whether they actually had to perform it at test). A test phase followed in which all 60 statements from the initial study phase and 30 new distractor statements appeared. For each statement, participants decided whether the statement was old or new and, if it was judged to be old, which source presented it. The specific ages of the sources were written below each source label. These source attributions were made self-paced by responding via the keyboard (keys "D" and "K" for the sources [assignment counterbalanced across participants in each condition], space key for new statements). Target and distractor statements appeared in random order. In the end, participant's perception of the item-source contingency during the study phase was collected. They judged how many statements of each typicality were presented by each source (e.g., "How many typically-young statements have been presented by the young vs. old source?"). Then, all participants completed the computerized version of the *SASKA* (Riegel, 1967) in which they selected one out of five response options that represented the meaning of a target word most via key press (keys "A" to "E"). Demographic and health-related personal information were provided at the end of the experiment followed by a debriefing. Participants were compensated and dismissed.

The specific assignment of the ages to the person sources (which source was labeled to be 23 vs. 70 years old) was revealed either at encoding or at test. That is, half of all participants from both participant age groups learned the specific assignments of ages to the two sources before encoding (*YA_enc_full*, *YA_enc_div*, *OA_enc_full*; specific ages were written below each source on screen); the other half learned this assignment only before the test phase (*YA_ret_full*, *YA_ret_div*, *OA_ret_full*). Each statement was presented for four seconds for younger participants. As age-related declines in cognitive functioning may impair older adults' contingency perception at encoding (cf. load at encoding for younger adults; Bayen & Kuhlmann, 2011), we increased the presentation time to eight seconds per item-source combination to facilitate encoding and contingency perception for older participants.

Only in the younger-adults sample, participants in the divided-attention conditions (*YA_enc_div*, *YA_ret_div*) performed a secondary tone-monitoring task adapted from Boywitt, Rummel, and Meiser (2015). During source attributions at test, participants additionally listened to tones of two different frequencies and were instructed to respond via key press (key "P") whenever three consecutive tones of the same frequency were presented. The tone sequence consisted of a pseudo-randomly order of two tones of either high (440 Hz) or low frequency (330 Hz) with a length of 180 trials. We adjusted the this sequence with the following restrictions: Not more than three tones of the same frequency should be presented consecutively, the same number of target (i.e., three consecutive tones with the same frequency) and distractor trials (i.e., < three consecutive tones with the same frequency) should be presented in each frequency, and at least one target trial should be presented within the first 10 tone trials. Tones were presented for one second followed by an ISI of two seconds. In total, 60 of the 180 tones were target trials with three consecutive tones of the same frequency and 120 were single tones and therefore defined as distractors. Between study and test phase, younger participants practiced the dual task for two minutes irrespective of whether they actually had to perform it during source attributions or not. The practice dual task consisted of the tone-sequence task (as described in the main text) while working on a math task simulating proper dual-task conditions. We presented simple mathematical equations (e.g., $9 + 14 = 23?$) that participants classified as either correct (key "S") or incorrect (key "L"). Participants received performance-related feedback at the end of the practice phase. Older adults performed the math task without the dual task. We instructed participants in the dual-task conditions that both tasks were equally important and that they should respond to the tone-monitoring task on target trials immediately but otherwise respond to the source-monitoring task. We provided error feedback for two seconds whenever participants missed a target. The respective source attribution remained on the screen until a response for this task was registered. If participants needed longer than 180 tones to attribute all statements to the sources, the tone sequence was repeated until they had attributed all test statements to the sources. We calculated the overall performance for the tone-monitoring task as hit rate minus false-alarm rate (discrimination measure Pr). Hit rate corresponds to the proportion of correct responses on target-present trials (i.e., three consecutive tones of the same frequency), false-alarm rate corresponds to the proportion of incorrect responses on target-absent trials (i.e., single tones). Individuals performed fairly well in the tone-monitoring task. Their respective performance measure indicated that individuals in both dual-task conditions executed this secondary task while their primary focus was to attribute the to-be-remembered statements to their sources comparably well ($YA\_enc\_div = .74$, $SD = .13$, $YA\_ret\_div = .75$, $SD = .15$), $t(70) = .29$, $p = .775$, $BF_{10} = .25$.

Statistical Model

In this experiment, data were coded into three item types like in previous research (e.g., Ehrenberg & Klauer, 2005; Kuhlmann et al., 2016; Schaper, Kuhlmann, & Bayen, 2019; Spaniol & Bayen, 2002): statements originating from the age-consistent source (i.e., aggregate of typically-young statements presented by the young source & typically-old statements presented by the old source), statements originating from the age-*in*consistent source (i.e., aggregate of typically-old statements presented by the young source & typically-young statements presented by the old source), and new statements that had not been presented by either source (aggregate of typically-young and -old distractor statements). The 2HTSM (Bayen et al., 1996) was then used to disentangle memory and guessing processes underlying the source attributions. In order to obtain model identifiability, parameter constraints needed to be implemented. From the submodels of the 2HTSM (see Bayen et al., 1996), Submodel 5d was the most parsimonious submodel that fitted the data best.[5] This submodel contains the following parameters: Parameter D (= item memory) measures the probability of either recognizing a statement as old or as new (assumption: item memory is equal across item types, $D_{\mathrm{T}} = D_{\mathrm{A}} = D_{\mathrm{N}}$). With the complementary probability $(1 - D)$, the statement is not recognized in the first place. In this case, with probability b the item is guessed to be either old or new $(1 - b)$ because the source can also not be remembered if the statement has not been remembered before (e.g., Malejka & Bröder, 2016). Parameter d_{T} (= source memory) reflects the probability of remembering the age-typical source and parameter d_{A} reflects the probability of remembering the age-*a*typical source if the statement will be recognized before. If source memory fails ($1 - d_{\mathrm{T}}$ or $1 - d_{\mathrm{A}}$) or the statement is guessed to be old, the source will be guessed (assumption: source guessing does not depend on whether a statement was remembered before, $a = g$). A graphical illustration of the submodel can be found in Figure 1.

Multinomial modeling traditionally relies on maximum-likelihood estimation aggregating across participants and thus neglecting (potential) differences between individuals (Batchelder & Riefer, 1999; Lee, 2011; Matzke, Dolan, Batchelder, &

Wagenmakers, 2015). We thus applied a Bayesian-hierarchical extension of MPT models (*latent trait*; Klauer, 2010) to estimate group- and individual-level parameters jointly. This latent-trait approach assumes that individual parameters are constrained by an overarching population-level distribution (multivariate normal distribution) thereby accounting for individual differences and parameter correlations. Parameter values can be derived through sampling from the resulting posterior distribution based on the Markov-chain Monte Carlo (MCMC) algorithm.

Whether the 2HTSM five-parameter submodel fit our data in each condition, was tested with Bayesian posterior predictive p-values $> .05$ for both the parameter means and covariance. We followed Klauer (2010) and relied on the T_1 statistic to assess model fit for the mean (difference between observed and expected mean frequencies) and the T_2 statistic for the covariance structure of the data (summed difference between observed and expected covariance; standardized by expected SD), separately for each experimental condition.[6] To ensure parameter comparability between conditions, we fitted this submodel to all conditions of both participant age groups. Non-significant p-values for the mean and covariance structure of each data set indicated that this model well accounted for all conditions of both participant age groups (see Table 2).

Analyses were conducted using *R* (R Core Team, 2020) and *JASP* (JASP Team, 2019). The hierarchical MPT model analyses were conducted with the *TreeBUGS* package (Heck, Arnold, & Arnold, 2018), figures were created with the *ggplot2* package (Wickham, 2016). In the subsequent *Results* section, we report the model-based mean parameter values and sampled parameter differences for between-condition comparisons with corresponding 95%-Bayesian Credibility Intervals (BCI) of the posterior distribution. Differences and parameter correlations are considered meaningful if the BCI excludes 0. Condition-based group-level parameter estimates can be retrieved from Table 2. Source-guessing parameter estimates are illustrated in Figure 2.

Results

First, we aimed to replicate the probability-matching account in younger adults. Based on Kuhlmann et al. (2012), we predicted stereotype-based source guessing for the younger participants under full attention with source information provided at test but contingency-based guessing under full attention with source information provided already at encoding. Indeed, when testing each conditions' source-guessing parameter against chance level of .50, the obtained difference was not credible in the encoding condition, $\triangle g - .50 = .04\ [-.05, .13]$, indicating unbiased source guessing but there was a credible bias in the retrieval condition, $\triangle g - .50 = .11\ [.02, .19]$. However, the direct comparison of both conditions did not reveal a credible difference, $-.07\ [-.19, .06]$. This is likely due to age stereotypes being rather weak, as opposed to the stronger profession schemas used in Kuhlmann et al. (2012).

Second, we tested the resource dependence of stereotype-based source guessing at retrieval by comparing younger participants in the retrieval condition with full or divided attention. Whereas source guessing exceeded chance level for *YA_ret_full*, as reported above, it did not for *YA_ret_div*, $\triangle g - .50 = .06\ [-.02, .14]$.[7] However, source guessing tended to be biased in *YA_ret_div* and did not differ from the biased guessing in the *YA_ret_full* condition, $-.08\ [-.17, .01]$.

Third, we tested whether the counteraction of stereotype-based source guessing via contingency knowledge depended on cognitive resources by comparing younger participants in the encoding condition who had either full or divided attention. As reported earlier, source guessing for *YA_enc_full* did not exceed chance level; neither did *YA_enc_div*, $\triangle g - .50 = -.002\ [-.08, .08]$. The comparison of both conditions also revealed no credible difference, $.04\ [-.08, .16]$.

In older adults, source guessing exceeded chance level in both conditions, $\triangle g - .50 = .10\ [.01, .18]$ (encoding) and $\triangle g - .50 = .18\ [.11, .24]$ (retrieval). Although this bias descriptively reduced in the encoding condition, the between-condition comparison revealed no credible difference, $.08\ [-.19, .03]$. Thus, older adults, with declines in cognitive resources, were not able to sufficiently counteract stereotype-based source guessing based on the actual contingency in the encoding condition.

In an exploratory analysis, we implemented the subjective contingency judgment (i.e., estimated number of typically-young and typically-old statements that were presented by either source during the study phase) collected at the end of the experiment as an external covariate for the source-guessing parameters. All conditions of both participant age groups correctly perceived the item-source contingency of .50, all $BF_{10} \leq .97$, but their contingency judgments were not substantially correlated with source guessing (see Table 3).

Discussion

This study examined what underlying mechanism may describe knowledge reliance in source guessing by testing effects of a dual task at retrieval and cognitive aging. Our main finding is that in younger adults neither stereotype-based nor contingency-based guessing was strongly affected by cognitive load at retrieval suggesting their resource independence, and thus automatic nature, of both stereotype-based and contingency-based source guessing. However, the ability to counteract stereotypes based on contingency knowledge in source guessing appears to be impaired in older adults suggesting that it is not fully automatic but, as discussed later, this may rather be due to contingency detection requiring resources.

Our conclusion that stereotype-based source guessing is automatic rests on the comparison between load conditions in younger adults' retrieval conditions as well as on the observation of stereotype-based guessing in older adults. It must be noted, however, that the direct comparison against chance level revealed a significant bias only in two of these three conditions but only a descriptive bias tendency for younger adults with divided attention. In younger adults, the accurate contingency perception under optimal encoding conditions can counteract the stereotype reliance in source guessing (replicating, e.g., Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012). To our knowledge this is the first time that this is demonstrated for source guessing based on age stereotypes and thus shows that enhancing attention to younger and older adults' actual behaviors can help overcome age stereotypes. Notably, younger adults' ability to counteract stereotypes based on contingency knowledge in source guessing was independent from cognitive resources available at test—reflecting a rather automatic application of contingency knowledge during source attributions.

However, older adults were not able to guess unbiased after learning the specific source information at encoding. Although at first glance this seems contradictory to the null effect of cognitive load on contingency-based guessing in older adults, it is crucial to keep in mind that the age-related declines in cognitive resources affect older adults' performance throughout the entire task, not just during retrieval. Thus, older adults might have had difficulties perceiving the contingencies during encoding, despite the longer encoding times, as do younger adults under cognitive load at encoding (Bayen & Kuhlmann, 2011). Unfortunately, the subjective contingency judgments do not seem informative to this matter as they did not show any condition differences or relation to source guessing. As already discussed in our preregistration, these judgments are likely reactively influenced by the test and the explicit phrasing of the question (cf. Kuhlmann et al., 2012). Indeed, mean contingency estimates were near .50, even in conditions for which the true source-item contingency greatly deviated from .50 (Bayen & Kuhlmann, 2011). A less explicit, online measure of contingency perception directly after study and during the test would thus be desirable and might reveal age differences. Alternatively, older adults in the encoding condition may have accurately perceived the contingency but still stuck to their strong stereotype knowledge. Indeed, older adults' stereotype bias in the retrieval condition was the most pronounced, in line with several other findings of stronger knowledge effects in older adults (see Umanath & Marsh, 2014, for a review).

While this study contributes to the understanding of the complex nature of source guessing, the present findings should be reviewed in the light of some limitations. First, our conclusions mainly rely on null effects. These, however, may result from a generally weak age-stereotype bias in all experimental conditions which mirrors the rather weak explicit typicality ratings for this stereotype domain (compared to more pronounced expectations in other knowledge domains; e.g. Bayen, Nakamura, Dupuis, & Yang, 2000; Küppers & Bayen, 2014). Furthermore, the interpretation of the results is conditional on the scaling of source guessing. That is, the baseline stereotype-based guessing that needed to be counteracted through individuating contingency knowledge was much more pronounced in older than in younger adults. Descriptively, older adults reduced their stereotype-based guessing, though not statistically reliable, but given their strong initial bias, their counteracted guessing was still biased. Note, however, that younger adults were able to counteract even comparably strong biases in previous studies (e.g., Arnold, Bayen, Kuhlmann, & Vaterrodt, 2013; Kuhlmann et al., 2012).

Second, studying the nature of source guessing only with a 50:50% item-source contingency at encoding can be challenged. Even though we think that the use of this contingency mirrors the ambiguity in the perception of different person (groups) best, future studies should test other contingencies. Of particular interest may be a condition in which the majority of items are presented by the atypical source leading to counter-stereotype based source guessing under full attention (Bayen & Kuhlmann, 2011). Does counter-stereotype based guessing remain possible even under cognitive load? And would older adults adjust source guessing to such a strong counter-stereotype contingency?

Third, the transfer and value of our results for stereotype application in real life awaits further research. In terms of future source attributions and impression formations, the automatic stereotype application in source guessing and thus a lack of individuation may severely affect the stereotyped person group (e.g., false accusations in eyewitness testimony). The good news, however, is that at least in younger adults the contingency-based counteraction succeeds even under cognitive load (e.g., from stress; cf. Sherman, Groom, Ehrenberg, & Klauer, 2003). This underpins the necessity for unimpaired encoding of information and its contextual details as a prerequisite for unbiased guessing later on. Whether the beneficial effect for source guessing in these encoding conditions is actually traced back to contingency detection cannot be assert with reasonable certainty from this study and remains to be determined. With regards to older adults, merely revealing stereotype-relevant information at encoding may not be enough to foster contingency learning. Future research should determine, under which (if any) conditions older adults can counteract stereotype-based guessing.

Footnotes

[1]More details about the reasoning for the proposed hypotheses can be retrieved from the preregistration on OSF: https://osf.io/2ezrn/?view$_o nly = ce2d6ad3949c4ed39e0ca664423ebe21$.

[2]Although Klauer and Ehrenberg (2005) did not observe differences in the item and/or source-memory performance at test, our pretested dual-task manipulation (20 participants, 10 with full and 10 with divided attention with the specific source-age information provided at test) pointed to poorer item memory under divided attention ($Pr = hits - false\ alarms$; *YA_ret_full*: $Pr = .64$, $SD = .12$; *YA_ret_div*: $Pr = .52$, $SD = .16$).

[3]In addition to the preregistered n of 36 per condition, we were able to collect two additional older participants—one per condition.

[4]One older adult did not indicate years of formal education and was thus removed from this analysis.

[5]We also applied the least parsimonious submodel 4 ($D_{\mathrm{T}} = D_{\mathrm{A}} = D_{\mathrm{N}}$; $d_{\mathrm{T}} = d_{\mathrm{A}}$; $g = a$) to our data but obtained a substantial misfit in the mean structure for older adults who were informed about the specific source' ages only at test (*OA_ret_full*; $T_1 = .024$ and $T_2 = .223$). Therefore, we fitted all submodels that explained the data with five parameters and submodel 5d revealed the best model fit for the respective *OA_ret_full* condition (as indicted by the non-significant p-values of $T_1 = .102$ and $T_2 = .139$).

[6]Even though other submodels indicated descriptively larger posterior predictive p-values for the covariance structure, none of the other submodels exceeded the one for the means (which was of main interest to test the proposed hypotheses).

[7]Descriptively, source guessing exceeded chance level for *YA_ret_div* and substantially so when we excluded one participant with the most pronounced counter-stereotype guessing parameter of. 30 in an exploratory analysis, $\triangle g - .50 = .08$ $[.001, .15]$.

References

Arnold, N. R., Bayen, U. J., Kuhlmann, B. G., & Vaterrodt, B. (2013). Hierarchical modeling of contingency-based source monitoring: A test of the probability-matching account. *Psychonomic Bulletin & Review*, *20*(2), 326–333. https://doi.org/10.3758/s13423-012-0342-7

Bargh, J. A. (1994). The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition. In R. Wyer & T. Skrull (Eds.), *Handbook of social cognition: Basic processes; applications* (pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. https://doi.org/10.1007/s00572-005-0022-9

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. https://doi.org/10.3758/BF03210812

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language*, *64*(1), 1–17. https://doi.org/10.1016/j.jml.2010.09.001

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*(3), 480–500. https://doi.org/10.3758/BF03198562

Bodenhausen, G. V., Macrae, C. N., & Sherman, J. W. (1999). On the dialectics of discrimination: Dual processes in social stereotyping. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 271–290). New York, NY: Guilford Press.

Boywitt, C. D., Rummel, J., & Meiser, T. (2015). Commission errors of active intentions: The roles of aging, cognitive load, and practice. *Aging,* *Neuropsychology, and Cognition*, *22*(5), 560–576. https://doi.org/10.1080/13825585.2014.1002446

Chasteen, A. L., Schwarz, N., & Park, D. C. (2002). The activation of aging stereotypes in younger and older adults. *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *57*(6), 540–547. https://doi.org/10.1093/geronb/57.6.P540

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology*, *41*(4), 369–387. https://doi.org/10.1016/j.jesp.2004.08.001

Gilbert, D. T., & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517. https://doi.org/https://doi.org/10.1037/0022-3514.60.4.509

Gluth, S., Ebner, N. C., & Schmiedek, F. (2010). Attitudes toward younger and older adults: The german aging semantic differential. *International Journal of Behavioral Development*, *34*(2), 147–158. https://doi.org/10.1177/0165025409350947

Gonsalkorale, K., Sherman, J. W., & Klauer, K. C. (2014). Measures of implicit attitudes may conceal differences in implicit associations: The case of antiaging bias. *Social Psychological and Personality Science*, *5*(3), 271–278. https://doi.org/10.1177/1948550613499239

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. Wyer & T. Skrull (Eds.), *Handbook of social cognition: Basic processes; applications* (pp. 1–68). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7

JASP Team. (2019). JASP (Version 0.11.1)[Computer software]. Retrieved from https://jasp-stats.org/

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Klauer, K. C., & Ehrenberg, K. (2005). Social categorization and fit detection under cognitive load: Efficient or effortful? *European Journal of Social Psychology*, *35*(4), 493–516. https://doi.org/10.1002/ejsp.266

Kornadt, A. E., & Rothermund, K. (2011). Contexts of aging: Assessing evaluative age stereotypes in different life domains. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *66*(5), 547–556. https://doi.org/10.1093/geronb/gbr036

Kuhlmann, B. G., & Bayen, U. J. (2016). Metacognitive aspects of source monitoring. In J. Dunlosky & S. K. Tauber (Eds.), *The oxford handbook of metamemory* (pp. 149–168). New York, NY: Oxford University Press.

Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. *Psychology and Aging*, *31*(8), 875–889. https://doi.org/10.1037/pag0000140

Kuhlmann, B. G., Kornadt, A. E., Bayen, U. J., Meuser, K., & Wulff, L. (2017). Multidimensionality of younger and older adults' age stereotypes: The interaction of life domain and adjective dimension. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *72*(3), 434–440. https://doi.org/10.1093/geronb/gbv049

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1365–1376. https://doi.org/10.1037/a0028147

Küppers, V., & Bayen, U. J. (2014). Inconsistency effects in source memory and compensatory schema-consistent guessing. *Quarterly Journal of Experimental Psychology*, *67*(10), 2042–2059. https://doi.org/10.1080/17470218.2014.904914

Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage. *Psychology and Aging*, *30*(1), 180–193. https://doi.org/10.1037/a0038586

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7. https://doi.org/10.1016/j.jmp.2010.08.013

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian data analysis for cognitive science: A practical course.* Cambridge University Press.

Lepore, L., & Brown, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology*, *72*(2), 275–287. https://doi.org/10.1037/0022-3514.72.2.275

Levy, B. R., Slade, M. D., Chung, P. H., & Gill, T. M. (2015). Resiliency over time of elders' age stereotypes after encountering stressful events. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences*, *70*(6), 886–890. https://doi.org/10.1093/geronb/gbu082

Malejka, S., & Bröder, A. (2016). No source memory for unrecognized items when implicit feedback is avoided. *Memory & Cognition*, *44*(1), 63–72. https://doi.org/10.3758/s13421-015-0549-8

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). Gender and orientation stereotypes bias source-monitoring attributions. *Memory*, *14*(2), 148–160. https://doi.org/10.1080/09658210544000015

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Matzke, D., Dolan, C. V., Batchelder, W. H., & Wagenmakers, E.-J. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205–235. https://doi.org/10.1007/s11336-013-9374-9

Monteith, M. J., Woodcock, A., & Lybarger, J. (2013). Automaticity and control in stereotyping and prejudice: The revolutionary role of social cognition across three decades. In D. E. Carlston (Ed.), *Oxford handbook of social cognition* (pp. 74–94). New York, NY: Oxford University Press.

Moshagen, M. (2010). multiTree : A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

Park, D. C. (2000). The basic mechanisms accounting for age-related decline in cognitive function. In D. C. Park & N. Schwarz (Eds.), *Cognitive aging: A primer* (pp. 3–19). Taylor & Francis. https://doi.org/10.4324/9780203727027

Perdue, C. W., & Gurtman, M. B. (1990). Evidence for the automaticity of ageism. *Journal of Experimental Social Psychology*, *26*(3), 199–216. https://doi.org/10.1016/0022-1031(90)90035-K

R Core Team. (2020). R: A language and environment for statistical computing [Computer software]. Retrieved from `https://www.R-project.org/`

Riegel, K. F. (1967). *Der sprachliche leistungstest saska: Synonym-, antonym-, selektions-, klassifikations-, analogietest*. Verlag für Psychologie Hogrefe.

Rivers, A. M., Sherman, J. W., Rees, H. R., Reichardt, R., & Klauer, K. C. (2019). On the roles of stereotype activation and application in diminishing implicit bias. *Personality and Social Psychology Bulletin*, *46*(3), 349–364. https://doi.org/10.1177/0146167219853842

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metamemory expectancy illusion and schema-consistent guessing in source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(3), 470–496. https://doi.org/10.1037/xlm0000602

Sherman, J. W., Groom, C. J., Ehrenberg, K., & Klauer, K. C. (2003). Bearing false witness under pressure: Implicit and explicit components of stereotype-driven memory distortions. *Social Cognition*, *21*(3), 213–246. https://doi.org/10.1521/soco.21.3.213.25340

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 631–651. https://doi.org/10.1037/0278-7393.28.4.631

Umanath, S., & Marsh, E. J. (2014). Understanding how prior knowledge influences memory in older adults. *Perspectives on Psychological Science*, *9*(4), 408–426. https://doi.org/10.1177/1745691614535933

Wechsler, D. (1981). *Wechsler adult intelligence scale - revised*. Psychological Corporation, New York.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer. Retrieved from `http://ggplot2.org`

Table 1

Demographic Sample Information

Participant group	Age	Sex		Formal education	*DSST*		Vocabulary test
		m	f		Number of correctly transcribed items	Free recall	
YA_enc_full	20.97 (2.08)	7	29	14.15 (1.81)	66.94 (10.59)	8.31 (1.67)	65.48 (.13)
YA_ret_full	21.44 (2.42)	6	30	14.31 (1.60)	66.78 (10.82)	8.09 (1.82)	63.49 (.16)
YA_enc_div	20.28 (2.27)	7	29	13.54 (1.34)	63.47 (9.71)	8.25 (1.42)	63.36 (.13)
YA_ret_div	20.83 (2.08)	12	24	14.26 (1.85)	66.37 (9.13)	7.94 (1.49)	67.99 (.16)
OA_enc_full	70.24 (5.82)	13	24	14.80 (2.95)	45.42 (10.97)	6.61 (1.93)	85.71 (.13)
OA_ret_full	71.54 (5.21)	16	21	15.97 (4.02)	45.89 (10.55)	6.03 (3.05)	88.67 (.12)

Note. Demographic sample description per experimental condition: *YA* = younger adults, *OA* = older adults; *enc* = source information (age) was revealed before encoding, *ret* = source information (age) was revealed before retrieval; *full* = full attention at test (no dual task), *div* = divided attention at test (dual task). Standard deviations are displayed in parentheses. Age and formal education is measured in years. *DSST* = Digit-symbol substitution task (Wechsler, 1981). The performance in the vocabulary test refers to the verbal intelligence test *SASKA* (Riegel, 1967).

Table 2

Model-Fit Statistics and Group-Level Parameter Estimates of Source-Monitoring Processes

Participant group	Model Fit		Source-Monitoring Parameter				
	T_1	T_2	D	d_{T}	d_{A}	b	g
YA_enc_full	.597	.716	.59 [.53, .65]	.92 [.81, .99]	.93 [.84, .1]	.43 [.37, .49]	.54 [.45, .63]
YA_ret_full	.542	.457	.54 [.47, .61]	.31 [.02, .74]	.54 [.19, .87]	.50 [.44, .57]	.61 [.52, .69]
YA_enc_div	.621	.682	.58 [.50, .64]	.93 [.77, 1]	.93 [.77, 1]	.49 [.42, .55]	.50 [.42, .58]
YA_ret_div	.591	.343	.45 [.37, .54]	.63 [.14, .96]	.49 [.15, .85]	.58 [.51, .65]	.56 [.48, .64]
OA_enc_full	.485	.348	.61 [.56, .66]	.91 [.70, 1]	.94 [.80, 1]	.56 [.48, .65]	.60 [.51, .68]
OA_ret_full	.106	.11	.51 [.44, .58]	.11 [.004, .33]	.36 [.04, .82]	.65 [.55, .73]	.68 [.61, .74]

Note. Assessed model fit and estimated memory and guessing processes per experimental condition: *YA* = younger adults, *OA* = older adults; *enc* = source information (age) was revealed before encoding, *ret* = source information (age) was revealed before retrieval; *full* = full attention at test (no dual task), *div* = divided attention at test (dual task). Model fit was assessed with posterior predictive p-values for the mean (T_1) and covariance (T_2) structure of the respective data set per condition when fitting submodel 5d of the 2HTSM (Bayen et al., 1996). p-values > .05 indicate adequate model fit. D = probability of recognizing a statement that had either been presented by the typical or atypical source or was known to be new; d_{T} = probability of correctly remembering the source of a statement that had been presented by the typical source; d_{A} = probability of correctly remembering the source of a statement that had been presented by the atypical source; b = probability of guessing that an unrecognized statement is old; g = probability of guessing that a recognized or unrecognized statement had been presented by the typical source. The 95%-BCI is displayed in brackets.

Table 3

Contingency Judgment

Participant group	Perceived Contingency	
	Mean Judgment	Correlation with Source Guessing
YA_enc_full	.50 (.06), BF_{10} = .20	-.06 [-.33, .21]
YA_ret_full	.50 (.06), BF_{10} = .18	-.14 [-.34, .07]
YA_enc_div	.49 (.04), BF_{10} = .97	-.01 [-.30, .29]
YA_ret_div	.49 (.07), BF_{10} = .25	.14 [-.08, .37]
OA_enc_full	.52 (.07), BF_{10} = .66	.07 [-.15, .31]
OA_ret_full	.51 (.09), BF_{10} = .23	.09 [-.08, .26]

Note. Contingency judgment per experimental condition: *YA* = younger adults, *OA* = older adults; *enc* = source information (age) was revealed before encoding, *ret* = source information (age) was revealed before retrieval; *full* = full attention at test (no dual task), *div* = divided attention at test (dual task). Contingency judgements were averaged for typically-young and typically-old statements and their difference to chance level of .50 was tested. Bayes factors refer to the test against .50. Following the Bayes-factor classification from Lee & Wagenmakers (2013), we interpreted these test results as anecdotal to moderate evidence for a lack of a difference to .50 in either condition.

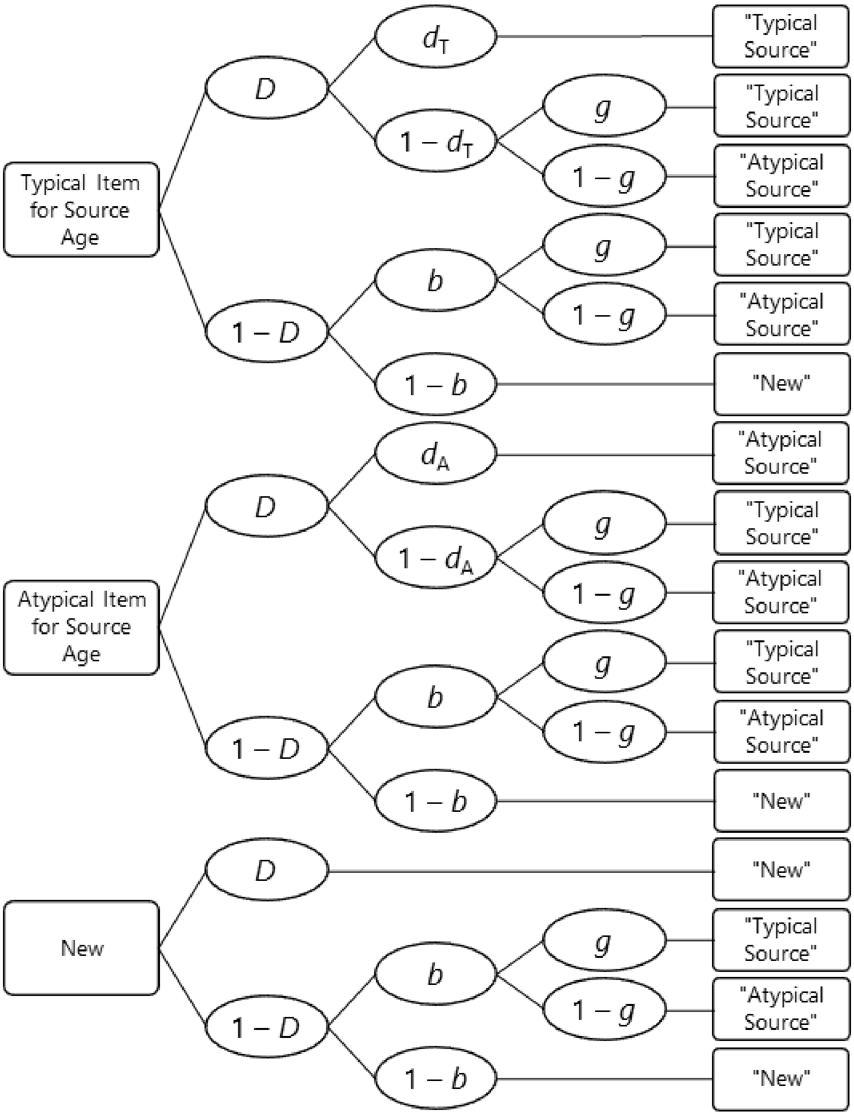

Figure 1. Submodel 5d of the 2HTSM adapted from Bayen et al. (1996). The three trees represent the respective item types, source-attribution responses are displayed in quotation marks. D = probability of recognizing a statement that had either been presented by the typical or atypical source or was known to be new; d_T = probability of correctly remembering the source of a statement that had been presented by the typical source; d_A = probability of correctly remembering the source of a statement that had been presented by the atypical source; b = probability of guessing that an unrecognized statement is old; g = probability of guessing that a recognized or unrecognized statement had been presented by the typical source.

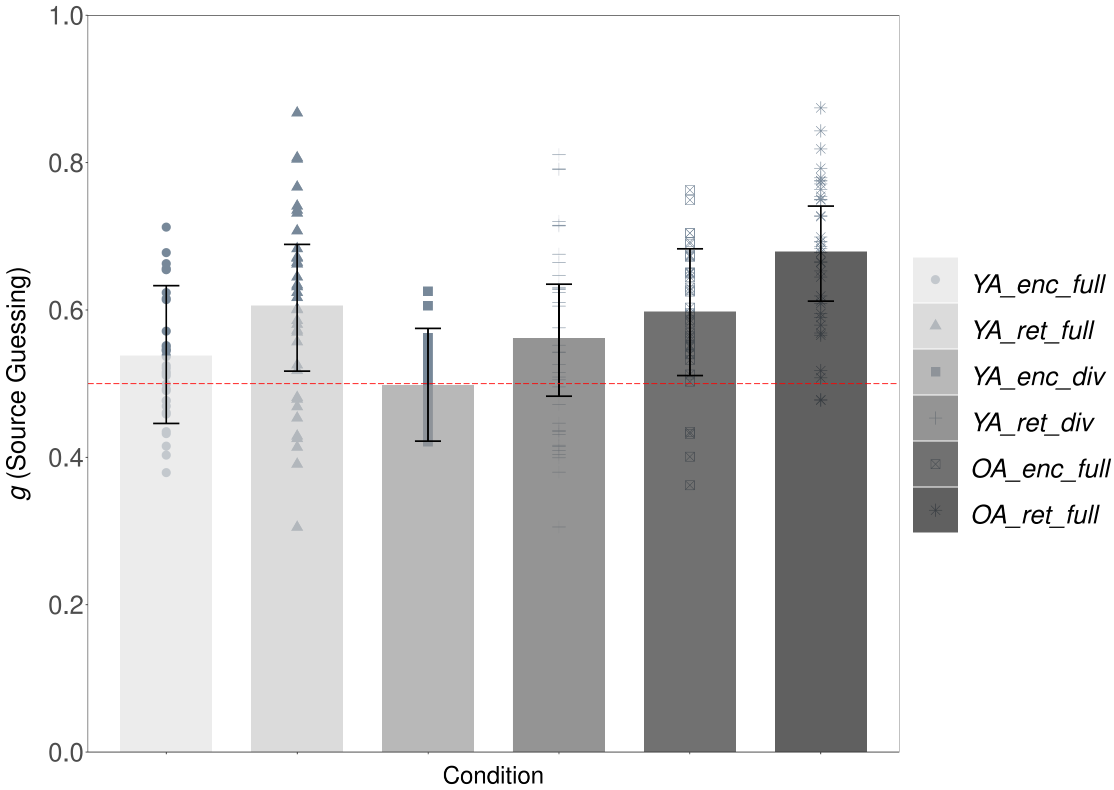

Figure 2. Mean source guessing per experimental condition: *YA* = younger adults, *OA* = older adults; *enc* = source information (age) was revealed before encoding, *ret* = source information (age) was revealed before retrieval; *full* = full attention at test (no dual task), *div* = divided attention at test (dual task). The dashed (red) line represents chance-level guessing of .50 (also representing the experiment's item-source contingency), values above .50 indicate stereotype-based source guessing, values below .50 indicate counter-stereotype-based source guessing. Group-level parameter means are displayed as bars, individual parameter estimates are displayed as dots. The error bars represent the 95%-BCI.

Appendix

Schematic Illustration of Experimental Procedure

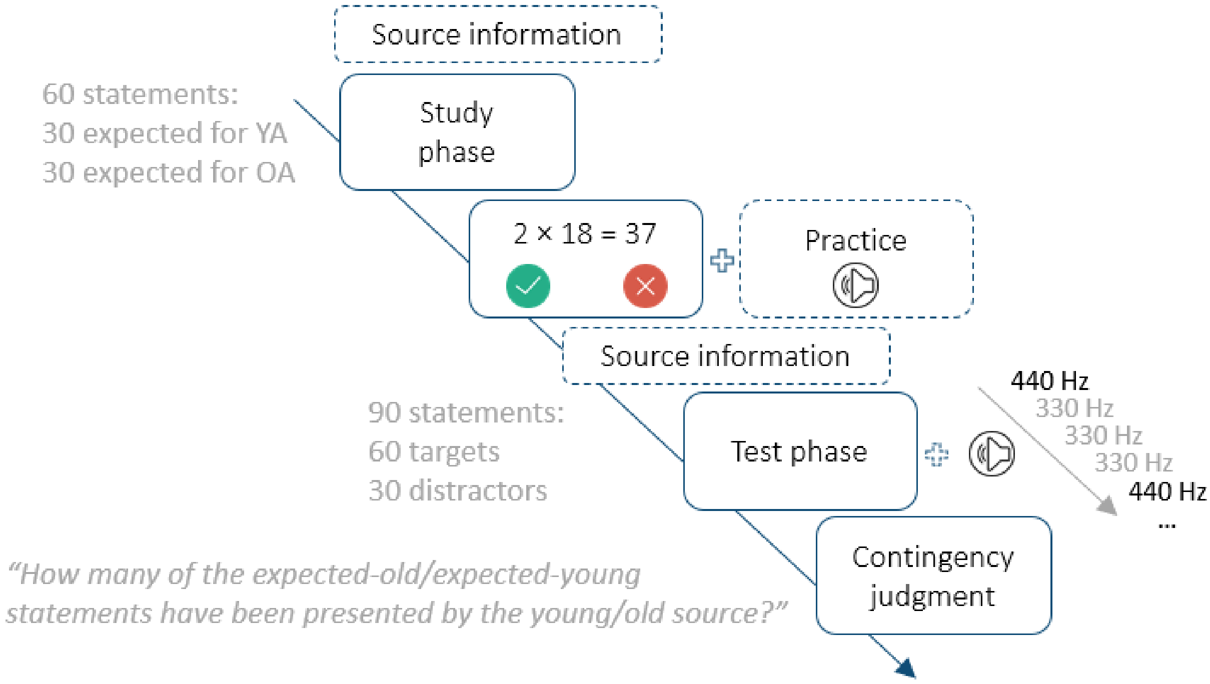

Figure A1. Conceptual illustration of the procedure. Dashed lines represent condition-specific adjustments of the general procedures (see *Procedure* section). Icons made by Smashicons from `https://www.flaticon.com/`.

Guess What?! Different Source-Guessing Mechanisms for Old Versus New Information

Liliane Wulff

University of Mannheim

Raoul Bell

Heinrich-Heine-University Düsseldorf

Laura Mieth

Heinrich-Heine-University Düsseldorf

Beatrice G. Kuhlmann

University of Mannheim

Author Note

Liliane Wulff, Department of Psychology, School of Social Sciences, University of Mannheim, Germany, Raoul Bell, Department of Experimental Psychology, Heinrich-Heine-University Düsseldorf, Germany, Laura Mieth, Department of Experimental Psychology, Heinrich-Heine-University Düsseldorf, Germany, and Beatrice G. Kuhlmann, Department of Psychology, School of Social Sciences, University of Mannheim, Germany. The research reported in this article was supported by the University of Mannheim's Graduate School of Economic and Social Sciences. The data and corresponding analyses can be retrieved from the Open Science Framework (OSF): `https://osf.io/fx9q4/?view_only=ad76a5c02496477a914af9ece141e2aa`. We thank Katja Bitz, Alina Kias, Franziska Leipold, Michelle Dörnte, Jule Schilling, Paula Schmelzer, and Selina Zajdler for their assistance with the data collection.

Correspondence concerning this article should be addressed to Liliane Wulff, M.Sc. Psychology, Department of Psychology, School of Social Sciences, University of Mannheim, D-68163 Mannheim, Germany. +49(0)621-181-3463. E-mail: wulff@psychologie.uni-mannheim.de

Abstract

The *probability-matching account* (Spaniol & Bayen, 2002) states that source guessing is matched to experimental contingencies of item types and sources. However, recently Bell, Mieth, and Buchner (2020) provided first evidence that this account may only apply to source attributions for previously-encountered contextual information. For source attributions about novel information, which participants detected as new, source guessing did not match the contingency knowledge perceived during encoding. In this case, source guessing continued to follow generic knowledge as the default thus neglecting past on-line contingency knowledge as a cue for prospective source attributions. The focus of the present study was to test whether this schema reliance for detected-new items could be explained by the lack of constant, individuating descriptions about the person sources in Bell et al. (2020). We accounted for this by presenting two specific persons with profession labels during encoding. Replicating Bell et al. (2020), source guessing for detected-old items was based on the item-source contingency and source guessing for detected-new items was based on schematic expectations about the professions—notwithstanding the constant use of specific source exemplars across previously-processed and novel contextual information. The present study thus adds evidence for persistent schema-based source guessing for detected-new items, even when novel information is attributed to known sources.

Keywords: source monitoring, schemas, source guessing, probability-matching account, Bayesian-hierarchical multinomial modeling

Guess What?! Different Source-Guessing Mechanisms for Old Versus New Information

How do we determine the origin of information? For instance, when to decide who told you about a recently-released movie in the theatre or where you read the latest news, which cues can we make use of to infer the source? Of course, it would be best if we can retrieve contextual details about the past situation in which we initially processed the information itself. But, for example due to the similarity of different sources that we usually process, we sometimes may fail to retrieve the distinct source (e.g., Bayen, Murnane, & Erdfelder, 1996). In this case, a promising cue for source attributions lacking source memory may be simply guessing, for instance, based on our experiences from the past. That is, when looking back at the example from above, we may simply guess that our best friend told us about the movie as they are film fans and, therefore, most likely have told us about it.

Johnson, Hashtroudi, and Lindsay (1993) conceptualized this dissociation between memory- and judgment-based source attributions and proposed an overarching theoretical framework according to which source attributions can be influenced by contextual memory processes for episodic, perceptual, or affective details about the origin of information. When source memory is absent, source attributions can be based on judgment processes based on pre-experimental knowledge (e.g., schemas, stereotypes) and metacognitive beliefs. A large body of evidence highlighted that, indeed, pre-experimental knowledge is used to infer the source when no explicit memory for the source is available (see Kuhlmann & Bayen, 2016, for an overview). One of the first knowledge domains in which this finding has been observed is the schema about different social groups, examined specifically for professional groups. In a source-monitoring experiment relying on the doctor-lawyer paradigm (e.g., Bayen, Nakamura, Dupuis, & Yang, 2000, Experiment 2), two person sources, a doctor (i.e., physician) and a lawyer, are shown with everyday statements reflecting activities and attitudes that are either expected for a doctor or a lawyer. Both sources present the same amount of statements from each item type (expected-doctor vs. expected-lawyer) such that there is a null contingency between the item type and the sources' professional groups. In a later memory test, individuals are asked to attribute the previously-learned and new statements, that have not been encountered before, to either source or classify them as new. Using multinomial modeling (Bayen et al., 1996), memory and guessing processes underlying individuals' source attributions could be disentangled revealing that source guessing was predominantly biased towards the schematically-expected source (i.e., neglecting the null contingency reflected in the experiment's item material) if veridical source information could not be retrieved from memory.

Importantly, source guessing is not always based on pre-experimental knowledge structures. Source guessing can be also influenced by peri-experimental knowledge that is acquired during the course of encoding of the item-source combinations. That is, source guessing can be adjusted based on the contingencies reflected in the item material (e.g., Bayen & Kuhlmann, 2011; Bell, Mieth, & Buchner, 2015; Ehrenberg & Klauer, 2005; Kuhlmann, Vaterrodt, & Bayen, 2012). Contingency-based source guessing is typically examined using sources about which participants have no prior knowledge (e.g., two unfamiliar male persons; Ehrenberg & Klauer, 2005). During the encoding of item-source pairings, the ratio of certain item types (e.g., positive or negative behaviors reflected in the descriptions of the two men) presented by either source is typically varied. More specifically, one source is associated with more items of a particular type (e.g., 75% negative behavior, 25% positive behavior), whereas this ratio is reversed for the other source (25% positive behaviors, 75% negative behaviors). In the test phase, participants can then rely on this previously learned item-source contingency in source attributions when source memory is absent.

The question under which conditions individuals make use of these different kinds of knowledge structures (e.g., pre- vs. peri-experimental) in source guessing has attracted the interest of researchers in the source-monitoring literature and ultimately led to the development of a theoretical account that formalized the use of different cues for source guessing and which is described in the next section.

The Probability-Matching Account and Its Extension to New Items

According to the *probability-matching account* of source guessing by Spaniol and Bayen (2002), individuals' source guessing should be based on probability matching in the first place (i.e., guessing in line with the peri-experimentally perceived proportion of outcome frequencies, for instance, different item types; cf. Schulze, James, Koehler, & Newell, 2019; Spaniol & Bayen, 2002). When speaking of probability matching in the context of source attributions, it refers to source guessing following the perceived distribution of item types (e.g., expected vs. unexpected for a certain source characteristic) to each source (cf. Spaniol & Bayen, 2002). Source guessing based on pre-experimental general world knowledge (such as schemas and stereotypes) only sets in if the situational conditions at encoding (e.g., divided attention; fast presentation time) hinder an accurate perception of the item-source contingency (e.g., Bayen & Kuhlmann, 2011; Kuhlmann et al., 2012). While, until recently, the interpretation of probability matching was limited to detected-old items, Bell et al. (2020) have since argued that individuals may, nonetheless, infer the source even for information of which they know it is new. The content itself may serve as a cue for the most likely source even though the information was previously not encountered and associated with a source.

For this reason, Bell et al. (2020) recently posed the question whether the probability-matching account could be extended to situations where source attributions are made for novel information for which individuals may have an intuition about what kind of source such an information may originate from (e.g., based on plausibility considerations)? In a series of studies, the authors demonstrated a clear distinction for source guessing between detected-old and detected-new items. More specifically (illustration from Experiment 1), participants performed a source-monitoring task in which multiple faces (each only once thus serving as the *items*) were shown with a profession label (farmer or lawyer) and behavior descriptions that were either expected for a farmer, a lawyer, or neutral (irrelevant for the distinction between the professions; serving as the *sources*). Each type of behavior description was paired equally often with each profession label. Thus, during the encoding phase, participants should have learned that there was a null contingency between a face's professional label and the type of behavior description provided. In the test phase, participants were first asked to decide whether a face (presented with the profession labels of farmer or lawyer) among new, distractor faces (labeled either farmer or lawyer) was old or new. After participants provided their old-new decision, they were asked to attribute each face to either of the three behavior descriptions (as the sources). That is, they attributed the faces (with profession labels) to behaviors either expected for a farmer, a lawyer, or neutral behavior. Importantly, participants attributed the faces to the behaviors irrespective of whether they initially classified the face as old or new. By this, the authors showed that the recognition status of an item (i.e., face) was incorporated into source guessing (see also Meiser, Sattler, & von Hecker, 2007): Individuals based their source guessing on the peri-experimental item-source contingency perceived during encoding when they had to infer the expectancy of the behavior description for items that they had previously detected to be old (i.e., unbiased guessing in line with the null contingency reflected in the stimulus material). Individuals, however, based their source guessing on pre-experimental schematic beliefs when they had to infer the expectancy of the behavior description for faces that they had detected to be new (i.e., bias towards the profession-expected behavior).

Research Question and Overview of Experiment

In summary, guessing for detected-old items is based on probability matching; guessing for detected-new items is based on schematic expectations about sources (Bell et al., 2020). Hence, the empirical evidence recently provided by Bell et al. (2020) implies that the probability-matching account only holds for previously-encountered, and not novel, information.

However, in the experimental paradigm used by Bell et al. (2020), the only cue that individuals could make use of when inferring the profession expectancy for detected-new faces was the accompanying label indicating the profession of the source. Except for the profession label, no further information connected the new faces to the studied faces. The rather unspecific face exemplars and their labels may activate the mental concept of the respective professional category but did not elicit an individual representation of this person nor trigger transfer of the previously-learned information about other individual faces. In our everyday life, however, we often interact with a finite number of specific persons in our social network (Dunbar, 2010, e.g., at the workplace or among our peer group) that we uniquely identify as those and who constitute of multifaceted attributes and behaviors. Therefore, one potential explanation for the schema reliance (instead of contingency reliance) for detected-new items in source guessing could be the lack of specific and constant individuating source information during encoding which is usually given in recurring social interactions in real life. This missing cue may prompt individuals to rely on schematic knowledge as default when performing a source-attribution attempt as they cannot draw on a learning history of the distribution of item types to the sources during the encoding process. Alternatively, a replication of the distinction between source guessing for detected-old and detected-new information with constant, known (from the encoding phase) sources would provide strong evidence that source guessing for detected-new items indeed is governed by different mechanisms than the well-studied source guessing for detected-old items.

The goal of the present study was to extend the findings on the probability-matching account for detected-old and detected-new items (Bell et al., 2020) using constant and distinct exemplars of two sources (i.e., persons) who present various information (i.e., more akin to the original doctor-lawyer paradigm; Bayen et al., 2000). More specifically, we aim to answer the question, whether people are capable to generalize their source guessing based on the item-source contingency perceived during encoding to an unrelated new decision context for the same, previously-encountered, sources uniquely identified by their profession and name? We, first, hypothesized that participants' source guessing for detected-old items should be determined by probability matching according to the proposed framework by Spaniol and Bayen (2002). That is, given that participants correctly perceive the true null contingency between item types and source professions at encoding, source guessing should be based on exactly this probability matching resulting in equal guessing of either source. Second, we hypothesized that participants apply the previously-experienced item-source contingency with already-familiar and specific source exemplars to detected-new items in their source guessing. This prediction is in contrast to Bell et al. (2020). As explained earlier, we predicted that we considered contingency-based source guessing for detected-new items as possible here due to our constant use of two specific source exemplars only across source attributions for detected-old and -new items.

Method

Design and Participants

The design was a 2 (statement profession expectancy [items]: statements expected vs. unexpected for the respective source's profession; within-subjects) × 2 (professional group [sources]: doctor vs. lawyer; within-subjects) factorial. The sources were presented by two male faces, each consistently accompanied by the same name and profession label. Furthermore, we counterbalanced source labels, corresponding images, and the order of the sources presented on the test screen between-subjects.

Sixty undergraduates (age range = 18−30 years; $M = 21.90$, $SD = 3$) were recruited via an electronic system or ads from the University of Mannheim (refer to the *Statistical Model* section for the sample-size reasoning). We excluded individuals whose native language (i.e., learned before the age of six) was not German.

Materials

The item set consisted of 108 German statements reflecting behaviors and attitudes to be expected for the profession of a doctor (= physician) or a lawyer (adapted from a German norming study by Kuhlmann et al., 2012, based on Bayen et al., 2000). The statements were initially rated with regard to their expectancy for both professional groups on a 5-point Likert scale (1 = "very unexpected", 2 = "unexpected", 3 = "neither expected nor unexpected", 4 = "expected", 5 = "very expected”) by 60 German undergraduates. We selected statements that were rated as expected ($M \geq 3.5$) for either a doctor (e.g., “Your blood pressure is too high.”) or a lawyer (e.g., “I have to be in court at 9 am.”) while at the same time rated as unexpected ($M < 2.5$) for the other profession such that only unambiguous statements remained from which we randomly selected 54 statements per profession.[1] Black-and-white images showing two similarly-looking adult men (taken from Bayen et al., 1996) labeled with the two German first names "Ralf" or "Uwe" (as in Kuhlmann et al., 2012) served as sources. The experiment was programmed with the open-source software program *OpenSesame* (Mathôt, Schreij, & Theeuwes, 2012).

Procedure

Participants were tested in groups up to four persons in separate computer cubicles. After they provided written informed consent, the computerized experiment started and participants performed a source-monitoring task with verbal item material reflecting profession-related attitudes and behaviors.

We informed participants about that we would present them with everyday statements that were said by either a doctor or a lawyer ("Ralf" or "Uwe"; assignment of names to sources counterbalanced between-subjects) accompanied by an image of the respective person or statements that were not assigned to either person and thus presented without a profession label, name, and an image. Thus, statements could originate from three sources, two persons presented with different professions/names/images and one person that we did not describe in any further detail. We provided an explicit source-memory instruction already before encoding. In the encoding phase, participants learned 54 statements (preceded and followed by three buffer statements of equal expectancy for both professional groups, respectively). Twenty-seven statements were expected to be said by a doctor and 27 statements were expected to be said by a lawyer. Each source was associated with 18 statements from which nine were expected for a doctor and nine were expected for a lawyer such that there was a null contingency between item types and sources. Statements were randomly selected from the item pool for each participant. Each statement—and, if applicable the face, name, and profession label—was displayed for four seconds until the next trial followed.

During the retention interval, participants performed a working-memory task in which they were instructed to monitor and remember digits in a subsequent free recall (serial-reproduction task). Participants were informed that they should remember eight subsequently-presented digits that they needed to reproduce within 10 seconds immediately afterwards (enter via keyboard). Digits varied from one to nine and were presented in random order. Each digit lasted on the screen for one second with an ISI of 250 ms. Participants received performance and detailed error feedback after each trial. Performance feedback was either “Your response was correct.” or “Your response was incorrect”. Error feedback would emerge if participants either entered too few digits than required (“Incorrect! You entered too few digits.”), too many digits than required (“Incorrect! You entered too many digits.”), or did not enter any digit (“You did not enter any digit. Please respond faster during the next trial.”). The distractor task lasted 10 minutes in total.

Next, we instructed participants for the upcoming self-paced (source-) memory test. They should indicate whether the statements were shown before ("old") or not ("new"). In addition, participants should indicate whether the statement (content) was associated with either of the three sources from the study phase: the doctor, the lawyer, or the third, unknown source (to the participants introduced as an unspecific person without a name and an image; referred to as "irrelevant" throughout the now following description, as this source did not contain any profession-related information). To be explicit, a source attribution needed to be made for all statements irrespective of whether they were judged to be old or new to continue with the next test trial after a 250 ms ISI. If participants could not retrieve the statement and/or source, they were instructed to simply guess for the old-new and/or source decision. The old-new decision was placed on the upper part of the screen and the three source response options were placed underneath (with the corresponding label name and image adjacent to them for the doctor and lawyer). In total, source attributions were made for 54 target and 54 distractor statements in random order. Participants made their responses via mouse clicks just as to proceed to the next trial.

In the end, we asked for participants' judgment of the item-source contingency perceived during encoding. That is, they estimated the number of statements of each profession expectancy associated with each of the three sources (e.g., "How many expected-doctor statements have been presented by the doctor, lawyer, and neither of them?"). Following this, personal information (age, gender, and subject of study) were collected before we debriefed, compensated (with course credit or monetary compensation), and dismissed participants.

Statistical Model

We applied a multinomial model to our data in order to disentangle memory and guessing processes underlying the individuals' categorical response data. Specifically, we employed the three-sources model variant of the *two-high threshold multinomial model of source monitoring* (Bayen et al., 1996; Keefe et al., 2002) from Bell et al. (2020, see also Bell, Mieth, & Buchner, 2017). This model accounts for source attributions judged to be old and, with the extension from Bell et al. (2020), also those judged to be new and is thereby well suited to test the proposed hypothesis as it estimates different source-guessing parameters based on the recognition status of the item.

We planned our sample size based on the experiments by Bell et al. (2020) but increased it to $N = 60$ participants to fulfill our counterbalance constraints. As our experimental paradigm and, associated therewith, the interpretation of parameters differed from Bell et al. (2020), parameter values from their study were not diagnostic with regard to an *a priori* power analysis. To nonetheless estimate the power for the difference test of source guessing for detected-old and detected-new items, we ran a *post-hoc* power analysis for an appropriate equivalent analysis on an aggregate group level using *multiTree* (Moshagen, 2010). Source guessing was set to .50 (indicating contingency-based guessing) for detected-old items (a_{E}) and to .70 (indicating schema-based guessing; parameter value identified as a benchmark for a schema bias in the literature on source monitoring using the doctor-lawyer paradigm e.g. Bayen & Kuhlmann, 2011; Bayen et al., 2000; Kuhlmann et al., 2012) for detected-new items (e_{E}). The analysis revealed that we had sufficient power ($1 - \beta > .99$) to detect substantial differences of .20 between source guessing for detected-old and detected-new items with $N = 60$ and $\alpha = .05$.

Exemplar model trees are illustrated in Figure 1. The following parameters representing underlying cognitive processes in source attributions are modeled with the MPT model for detected old and new items (adapted from Bell et al., 2020): Parameter D reflects the probability of recognizing an item that was either presented with the expected (D_{E}), unexpected (D_{U}), or irrelevant source (D_{I}) or was detected to be new (D_{N}). The respective complementary probability ($1 - D$) mirrors that an item was not recognized or detected as new. The old-new status of an item is then guessed to be either old with probability b or new with its complementary probability ($1 - b$). If an item was recognized as old, parameter d reflects the probability of remembering the respective source (d_{E}, d_{U}, or d_{I}). The respective complementary probability ($1 - d$) mirrors that the source was not remembered. In this case or if an item was guessed to be old, the source must be guessed. Two parameters allow source guessing to potentially vary between the recognition status of items. That is, parameter a_{I} reflects the probability of guessing the irrelevant source if the item was recognized before (but source memory failed). If the irrelevant source was not guessed, the expected source was guessed with probability a_{E} or the unexpected source was guessed with its complementary probability $1 - a_{\mathrm{E}}$. If the item was not recognized in the first place, parameter $g_{\mathrm{O\,I}}$ measures the probability of guessing the irrelevant source given that the item was guessed to be old before. If the irrelevant source was not guessed, the expected source is guessed with probability $g_{\mathrm{O\,E}}$ or the unexpected source with its complementary probability $1 - g_{\mathrm{O\,E}}$. If the item was not recognized and the status of the item was guessed to be new, parameter $g_{\mathrm{N\,I}}$ reflects the probability to guess the irrelevant source. If the irrelevant source was not guessed, the expected source is guessed with probability

g_{NE} or the unexpected source with its complementary probability $1 - g_{\mathrm{NE}}$. Furthermore, the model includes an additional parameter e that reflects source guessing for detected-new items. If a distractor item was detected as new in the first place, parameter e_{I} measures the probability of guessing the irrelevant source. If the irrelevant source was not guessed, the expected source was guessed with probability e_{E} and the unexpected source was guessed with its complementary probability $1 - e_{\mathrm{E}}$.

Thus, the multinomial model estimated its parameters based on test responses to four types of item-source combinations: statements either originating from the "expected" source (i.e., expected-doctor statements associated with the doctor and expected-lawyer statements associated with the lawyer), from the "unexpected" source (i.e., expected-doctor statements associated with the lawyer and expected-lawyer statements associated with the doctor), from the irrelevant source (i.e., expected-doctor and expected-lawyer statements presented without any label or image), or new (distractor) statements never presented during encoding (i.e., expected-doctor and expected-lawyer statements without any source).

Multinomial modeling traditionally relies on maximum-likelihood estimation of data (i.e., test response frequencies) aggregated across participants and thus neglect (potential) differences between individuals (Batchelder & Riefer, 1999; Lee, 2011; Matzke, 2015). We implemented a Bayesian-hierarchical approach of multinomial modeling, specifically the *latent trait* approach (Klauer, 2010), to consider individual differences and potential covariates which may explain variability in memory and guessing parameters. This approach enables researchers to model group-level and individual-level parameters jointly. The implemented latent-trait approach assumes that individual parameters are constrained by an overarching population-level distribution (multivariate normal distribution) thereby accounting for individual differences and parameter correlations. Parameter values can be derived through sampling from the resulting posterior distribution based on the Markov-chain Monte Carlo (MCMC) algorithm. The parameters of most interest to test our hypothesis were a_{E} and e_{E}, as they reflect the probability to guess the expected, schema-based, source for detected-old and detected-new items.

Results

All analyses were conducted using *R* (R Core Team, 2020) and *JASP* (JASP Team, 2019). The hierarchical MPT model analyses were conducted with the *TreeBUGS* package (Heck, Arnold, & Arnold, 2018), figures were created with the *ggplot2* package (Wickham, 2016).

As we did not expect item-memory differences between item types *a priori*, the baseline MPT model was restricted to just one item-memory parameter D (= D_{E} = D_{U} = D_{I} = D_{N}) assuming equality in the memorability between item types. We assessed model fit with Bayesian posterior predictive p-values for the mean (T_1 statistic; difference between observed and expected mean frequencies) and covariance structure (T_2 statistic; summed difference between observed and expected covariance; standardized by expected SD) of the data as proposed by Klauer (2010) and implemented in Heck et al. (2018). Both model-fit indices were estimated non-significant ($T_1 = .446$, $T_2 = .401$). Thus, the model and imposed restrictions on the item-memory parameters explained the data well. Sampled mean parameter estimates, parameter differences (also against .50) in source guessing, and parameter correlations, as well as parameter correlations with the contingency judgments included as external covariate as implemented in *TreeBUGS* (Heck et al., 2018) are reported with respective 95%-Bayesian Credibility Intervals (BCI) of the posterior distribution. If the BCI excluded 0, parameter differences and correlations were considered statistically meaningful. An overview of mean parameter estimates (and correlations with contingency judgments) is given in Table 1. Estimates of source-guessing parameters are graphically illustrated in Figure 2.

Source Guessing for Detected-Old and Detected-New Items

To test the probability-matching account for items remembered as old or detected as new, we, first, compared the respective guessing parameters against chance level of .50. We did so because the null contingency between item types presented by each source at encoding should lead to an indifference in the reliance on the profession schema in source guessing among expected and unexpected item-source combinations. If, in contrast, source guessing exceeds .50, items expected for either professional group were guessed to stem from the expected source more often than from the unexpected source, indicating reliance on the profession schema. Source guessing for expected items recognized as old was equal to .50 (and thus perfectly matched to the factual contingency) as indicated by a non-credible difference, $\triangle a_{\mathrm{E}} - .50 = -0.002$ $[-.19, .17]$.

In contrast, source guessing for expected items detected to be new largely exceeded the actual .50 contingency as indicated by a credible positive difference, $\triangle e_{\mathrm{E}} - .50 = .40$ $[.06, .50]$. Thus, participants' source guessing was reset to a pronounced reliance on schematic knowledge as default when source attributions for new items needed to be made. The divergence between source guessing for items detected as old and those detected as new was also mirrored in the direct comparison between both parameters. The sampled difference revealed credibility, $\triangle a_{\mathrm{E}} - e_{\mathrm{E}} = -.40$ $[-.63, -.06]$.

Source Guessing for Guessed-Old and Guessed-New Items

By means of source guessing for items for which the old-new status needed to be guessed in the first place (parameter g; for old and new items, respectively), contingency knowledge perceived during encoding could not be taken into account. In a state of complete uncertainty with regard to the item status and its origin, participants based their source guessing predominantly on schematic knowledge for guessed-old items, $\triangle g_{\mathrm{OE}} - .50 = .18$ $[.09, .27]$, and did so even more if the item status was guessed to be new, $\triangle g_{\mathrm{NE}} - .50 = .42$ $[.29, .49]$, as indicated by a credible difference of both parameters, $\triangle g_{\mathrm{OE}} - g_{\mathrm{NE}} = -.24$ $[-.36, -.09]$.

Contingency Judgment

The explicit post-judgments of perceived contingencies during encoding revealed that individuals estimated the contingency (percentage of expected-doctor and expected-lawyer statements presented by each of the three sources), although descriptively close to the actual contingency ($M = .41$, $SD = .10$), biased in favor of expected source, $t(59) = 5.79$, $p < .001$, $BF_{10} = 50100.33$, tested two-sided against .33. Individuals thus stated that more expected-doctor items were presented by the doctor and more expected-lawyer items were presented by the lawyer than each source actually presented. The correlation of the contingency judgment and source guessing for detected-old items reached statistical credibility as indicated by .40 $[.22, .54]$. Further, we observed a positive correlation coefficient of comparable size for detected-new items. Here, again, the contingency judgments were linked to source guessing to a credible extent (.30 [.03, .48]). That is, at least to some extent as indicated by medium-sized correlation coefficients, individuals who estimated the item-source contingency in a rather schema-biased manner, tended to show a more pronounced schema bias in source guessing for both detected-old and detected-new items. Note, however, that only the correlation of source guessing and the contingency judgments for detected-old items remained credible [.08, .62] when controlling for the sampling error of the estimated population correlation according to the sample size. This was not true for detected-new items $[-.08, .56]$.

Discussion

Is peoples' source guessing guided by probability matching generalized to detected-new items when processing information from two distinct person sources? The present research aimed to answer this question by implementing a source-monitoring task based on the doctor-lawyer paradigm and by modeling the respective cognitive processes at play during source attributions for old and new items. In line with Bell et al. (2020), the present results replicate the recent evidence on probability matching for detected-old items but did not hold for detected-new items using a modified source-monitoring task of the doctor-lawyer paradigm with two constant, distinct source exemplars. That is, instead of applying the previously-encountered item-source contingencies as done for detected-old items to a new decision context without source information available, participants rather relied on pre-experimental generic schematic knowledge about the professions of a doctor and a lawyer as guessing heuristic for new items.

In source attributions for detected-old and guessed-old items, participants considered the recognition status of the specific item in their source decision based on guessing. When participants detected an item as old, they used the previously-encountered item-source contingencies during encoding as a cue for attributing the information to its origin. Particularly, on average, participants' source guessing was based on the factual contingency presented during encoding for those items that were detected to be old. This findings adds evidence in favor of the probability-matching account (developed for within the scope of information detected and believed to be old; Bayen & Kuhlmann, 2011; Bell et al., 2020; Kuhlmann et al., 2012; Spaniol & Bayen, 2002). When participants, however, guessed an item to be old, they refrained from using the previously-encountered item-source contingencies and, instead, based their source guessing on pre-experimental generic schematic knowledge about the doctor and the lawyer (in line with Bell et al., 2020).

How can the dissociation in the use of knowledge cues, which was peri-experimental knowledge for detected-old items and pre-experimental knowledge for guessed-old items, in source guessing be explained? Based on the empirical evidence reported here, one can only speculate about the potential candidates that drive this dissociation between detected and non-detected items. Participants' metacognitive reasoning could be one such candidate as judgment processes in source monitoring have been shown to be influenced by the subjective beliefs about source inference (see e.g., Batchelder & Batchelder, 2008; Kuhlmann & Touron, 2011; Meiser et al., 2007). Thus, participants' source inference for old items could be guided by their metacognitive beliefs about the expected difference in the recognition status of an item. That is, for source attributions based on a state of uncertainty for guessed-old items, contingency knowledge acquired during the course of encoding is not applied to source guessing. Based on the null contingency reflected in the stimulus material (i.e., balanced ratio of expected-doctor and expected-lawyer items presented by either source) unbiased source guessing for both types of old items seems likely but participants' source guessing indeed varied as a function of their belief about the recognition status of old items which was valued as an additional cue for the judgment processes during source attributions.

In source attributions for detected-new and guessed-new items, participants did not consider the recognition status of the specific item in their source decision based on guessing. Irrespective of whether an item was detected as or guessed to be new, they relied on their generic, pre-experimental schema knowledge about the professions. Previously-learned associations between the items and distinct sources were not transferred to a new decision context for the very same person sources. That is, the contingency knowledge acquired during encoding of information from distinct sources could not be generalized to source guessing for novel information (presented without any source information) of these specific persons.

As Bell et al. (2020) already discussed, the schema bias in source guessing for detected-new items may serve the purpose of the most logical decision criterion as the reliance on pre-experimental knowledge can lead to correct source attributions in many instances—especially compared to contingency representations that may fluctuate considerably across contexts. In contrast to the present study, participants in the study by Bell et al. (2020) processed information that was associated with multiple persons, which is why a schema reliance in source guessing thereby seemed (more) logical. But the validity of pre-existing knowledge such as schemas and stereotypes can, of course, be questioned particularly whenever these attitudes and associated behaviors do not precisely reflect features of an individual's personality. This holds particularly true in the present experiment as participants learned unique information about two constant sources, which renders the reliance on the profession schema unreasonable here as the sources did not behave schema-conform. Even more so, it can be described as nothing less than bad news when it comes to the practical implications of the present finding for detected-new items. Namely, this study clearly demonstrated that even though inferences for novel information could be drawn for the exact same and distinct source exemplars from whom individuals have learned expected *and* unexpected behavioral descriptions before, they draw on non-individualized, generic knowledge about the professions. More precisely, irrespective of past behaviors that were associated with the sources, and therefore also encompassed unexpected behaviors for the sources' professions, sources were nonetheless associated with their respective prototypical profession behavior. The acquired knowledge about specific person sources from a past learning episode was not transferred to novel information from these already-known sources. This finding is particularly remarkable against the background of the potential persistence of attitudes towards certain person groups inaccurately conforming schemas and stereotypes in society. However, we can only speculate about how source guessing for detected-new items would evolve given, for instance, more detailed background knowledge about the sources in an experimental setting or increased degrees of familiarity with the sources in real life. An interesting and fruitful avenue for prospective research foci could be to further examine the persistence of pre-experimental knowledge reliance under decisional uncertainty for new information, its disproportionate application, and interventions to overcome such biases.

In order to conclusively claim that source guessing is generally biased by pre-experimental knowledge for detected-new items with reasonable confidence, we propose to, first, disqualify other limitations as potential explanations. For instance, the null contingency may have been difficult to detect for participants, as evident in the, on average, biased contingency judgments and marked individual difference in source-guessing parameter for detected-old items (a_{E})—despite unbiased probability matching on the group level. One potential explanation for schema-biased guessing for detected-new items could be that the contingency detection renders more difficult with the increase in complexity due to including the third rather abstract source. We would, therefore, suggest to replicate the study with more pronounced, unbalanced, item-source contingencies (e.g., as done in Bayen & Kuhlmann, 2011) and place a greater emphasis on the unaffected integration of item (types) and sources during intentional learning (e.g., due to explicitly informing about the contingency and instructing participants to pay attention to it). As a consequence, a generalization and application of this mental representation to source attributions for detected-new items may become more likely. Nonetheless, the preconditions have indeed been ideal detect the accurate contingency (based on the implemented source-memory instruction, Kuhlmann et al., 2012, and full attention during encoding, Bayen & Kuhlmann, 2011). In the same vein, effective encoding provides a basis for probability matching, at least for sources who were associated with a specific context, and turned out to be one promising candidate to counteract schematic and stereotypical influences on source guessing in the past (Kuhlmann, Bayen, Meuser, & Kornadt, 2016; Wulff & Kuhlmann, 2020b). The particular conditions under which source guessing for novel information most likely also benefits from accurate contingency perception is still to be examined, as well as the challenging issue of how to address individual differences in prospective interventions. The recent statistical advances that made the measurement of individual difference possible and found their legitimate application in an increasing number of studies (e.g., Arnold, Bayen, & Smith, 2015; Lee, Bock, Cushman, & Shankle, 2020; Schaper, Kuhlmann, & Bayen, 2019; Wulff & Kuhlmann, 2020a) should be used to examine the interplay and dependencies of source guessing for old and new items (also on an individual level) in prospective research.

It should be noted, however, that individuals dissociated in their source guessing, depending on the recognition status of the item, between items presented by the expected/unexpected sources and items presented without specific source information. That is, participants were most inclined to guess the third, unspecific source, for detected-new items but were more hesitant to do so for detected-old items. The recognition status of an item thus affected source guessing also for the irrelevant source for which no specific source characteristic was provided. The same pattern emerged for guessed-old and -new items with an increased probability to attribute guessed-new items to the non-specified source reflecting their general decisional uncertainty. When participants were in a state of lacking memory, they were, nonetheless, forced to make a source attribution in the present paradigm. In such cases, the recognition status of the item seemed to be used as a cue to make inferences about the origin of information such that new items for which no adequate source attributions could be made were more often associated with the least informative source for the decision criterion based on the profession. This observation, in addition to the recognition-status specificity of guessing of schema-expected/-unexpected exemplars, further underpins that individuals potentially rely on various cues when judging the source (cf. Bell et al., 2020).

In conclusion, the scope of application of the probability-matching account appears to be restricted to items that were detected as or believed to be old. Albeit participants learned information about distinct source exemplars of a doctor and a lawyer (unlike in Bell et al., 2020) and used the contingency knowledge reflected in the item material to adjust their source guessing for previously-encountered information detected as such, they nonetheless refrained from generalizing this knowledge application to novel stimuli and based their source guessing on a schema acquired prior to the experiment. This divergence in the use of knowledge cues in source guessing between detected-old and detected-new items goes beyond the explanatory power of the probability-matching account.

Footnotes

[1]Marginal differences between the expectancy of professions were inherent to the ratings and were kept at a minimum in the item composition. Due to the construction of the original item set of parallel statements (e.g., "I have to be in court at 9 am." vs. "I have to be in court at 12 pm."), item selection was limited to either of both parallel statements to reduce the likelihood of confusion. We did our very best to, nonetheless, create a sample of statements expected for each professional group that did not differ between professions. The incongruent ratings did not differ between professions, $p \geq .623$, but we were, unfortunately, not able to achieve this for the congruent ratings, $p < .001$. Descriptively, the mean ratings between professions were, nonetheless, in a comparable range of the scale: $M_{\text{doctor}} = 4.64$ and $M_{\text{lawyer}} = 4.37$.

References

Arnold, N. R., Bayen, U. J., & Smith, R. E. (2015). Hierarchical multinomial modeling approaches: An application to prospective memory and working memory. *Experimental Psychology*, *62*(3), 143–152. https://doi.org/10.1027/1618-3169/a000287

Batchelder, W. H., & Batchelder, E. (2008). Metacognitive guessing strategies in source monitoring. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 211–244). New York, NY: Psychology Press. https://doi.org/10.4324/9780203805503.ch12

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, *6*(1), 57–86. https://doi.org/10.3758/BF03210812

Bayen, U. J., & Kuhlmann, B. G. (2011). Influences of source-item contingency and schematic knowledge on source monitoring: Tests of the probability-matching account. *Journal of Memory and Language*, *64*(1), 1–17. https://doi.org/10.1016/j.jml.2010.09.001

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 197–215. https://doi.org/10.1037/0278-7393.22.1.197

Bayen, U. J., Nakamura, G. V., Dupuis, S. E., & Yang, C. L. (2000). The use of schematic knowledge about sources in source monitoring. *Memory & Cognition*, *28*(3), 480–500. https://doi.org/10.3758/BF03198562

Bell, R., Mieth, L., & Buchner, A. (2015). Appearance-based first impressions and person memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 456–472. https://doi.org/10.1037/xlm0000034

Bell, R., Mieth, L., & Buchner, A. (2017). Emotional memory: No source memory without old–new recognition. *Emotion*, *17*(1), 120–130. https://doi.org/10.1037/emo0000211

Bell, R., Mieth, L., & Buchner, A. (2020). Source attributions for detected new items: Persistent evidence for schematic guessing. *Quarterly Journal of Experimental Psychology*. https://doi.org/10.1177/1747021820911004

Dunbar, R. (2010). *How many friends does one person need?: Dunbar's number and other evolutionary quirks*. Cambridge, MA: Harvard University Press.

Ehrenberg, K., & Klauer, K. C. (2005). Flexible use of source information: Processing components of the inconsistency effect in person memory. *Journal of Experimental Social Psychology*, *41*(4), 369–387. https://doi.org/10.1016/j.jesp.2004.08.001

Heck, D. W., Arnold, N. R., & Arnold, D. (2018). TreeBUGS: An R package for hierarchical multinomial-processing-tree modeling. *Behavior Research Methods*, *50*(1), 264–284. https://doi.org/10.3758/s13428-017-0869-7

JASP Team. (2019). JASP (Version 0.11.1)[Computer software]. Retrieved from `https://jasp-stats.org/`

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3–28. https://doi.org/10.1037/0033-2909.114.1.3

Keefe, R. S. E., Arnold, M. C., Bayen, U. J., McEvoy, J. P., Wilson, W. H., Larkin, W., ... Diforio, D. (2002). Source-monitoring deficits for self-generated stimuli in schizophrenia: Multinomial modeling of data from three sources. *Schizophrenia Research*, *57*(1), 51–67. https://doi.org/10.1016/S0920-9964(01)00306-1

Klauer, K. C. (2010). Hierarchical multinomial processing tree models: A latent-trait approach. *Psychometrika*, *75*(1), 70–98. https://doi.org/10.1007/s11336-009-9141-0

Kuhlmann, B. G., & Bayen, U. J. (2016). Metacognitive aspects of source monitoring. In J. Dunlosky & S. K. Tauber (Eds.), *The oxford handbook of metamemory* (pp. 149–168). New York, NY: Oxford University Press.

Kuhlmann, B. G., Bayen, U. J., Meuser, K., & Kornadt, A. E. (2016). The impact of age stereotypes on source monitoring in younger and older adults. *Psychology and Aging*, *31*(8), 875–889. https://doi.org/10.1037/pag0000140

Kuhlmann, B. G., & Touron, D. R. (2011). Older adults' use of metacognitive knowledge in source monitoring: Spared monitoring but impaired control. *Psychology and Aging*, *26*(1), 143.

Kuhlmann, B. G., Vaterrodt, B., & Bayen, U. J. (2012). Schema bias in source monitoring varies with encoding conditions: Support for a probability-matching account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1365–1376. https://doi.org/10.1037/a0028147

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. *Journal of Mathematical Psychology*, *55*(1), 1–7. https://doi.org/10.1016/j.jmp.2010.08.013

Lee, M. D., Bock, J. R., Cushman, I., & Shankle, W. R. (2020). An application of multinomial processing tree models and Bayesian methods to understanding memory impairment. *Journal of Mathematical Psychology*, *95*, 102328.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314–324. https://doi.org/10.3758/s13428-011-0168-7

Matzke, D. (2015). Bayesian estimation of multinomial processing tree models with heterogeneity in participants and items. *Psychometrika*, *80*(1), 205–235. https://doi.org/10.1007/s11336-013-9374-9

Meiser, T., Sattler, C., & von Hecker, U. (2007). Metacognitive inferences in source memory judgements: The role of perceived differences in item recognition. *The Quarterly Journal of Experimental Psychology*, *60*(7), 1015–1040. https://doi.org/10.1080/17470210600875215

Moshagen, M. (2010). multiTree : A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54. https://doi.org/10.3758/BRM.42.1.42

R Core Team. (2020). R: A language and environment for statistical computing [Computer software]. Retrieved from `https://www.R-project.org/`

Schaper, M. L., Kuhlmann, B. G., & Bayen, U. J. (2019). Metamemory expectancy illusion and schema-consistent guessing in source monitoring. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *45*(3), 470–496. https://doi.org/10.1037/xlm0000602

Schulze, C., James, G., Koehler, D. J., & Newell, B. R. (2019). Probability matching does not decrease under cognitive load: A preregistered failure to replicate. *Memory & Cognition*, *47*(3), 511–518. https://doi.org/10.3758/s13421-018-0888-3

Spaniol, J., & Bayen, U. J. (2002). When is schematic knowledge used in source monitoring? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 631–651. https://doi.org/10.1037/0278-7393.28.4.631

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York, NY: Springer. Retrieved from `http://ggplot2.org`

Wulff, L., & Kuhlmann, B. G. (2020a). Is knowledge reliance in source guessing a cognitive trait? Examining stability across time and domain. *Memory & Cognition*, *48*(2), 256–276. https://doi.org/10.3758/s13421-019-01008-1

Wulff, L., & Kuhlmann, B. G. (2020b). *Is knowledge reliance in source guessing automatic or controlled? Evidence from divided attention and aging*. Manuscript submitted for publication.

Table 1

Group-Level Mean Parameter Estimates and Correlations With Contingency Judgment

	Mean				Correlation with Contingency Judgments			
Parameter	Estimate	*SD*	2.5%-BCI	97.5%-BCI	Estimate	*SD*	2.5%-BCI	97.5%-BCI
D	.42	.02	.38	.47	-.05	.07	-.18	.08
d_{I}	.32	.12	.09	.55	-.16	.10	-.36	.05
d_{E}	.39	.09	.20	.53	-.26	.14	-.47	.11
d_{U}	.39	.11	.15	.60	-.16	.16	-.42	.19
b	.37	.03	.31	.44	-.13	.05	-.22	-.03
a_{I}	.24	.05	.15	.33	.18	.13	-.09	.41
a_{E}	.50	.09	.31	.67	.40	.08	.22	.54
e_{I}	.81	.13	.47	.98	-.30	.04	-.38	-.21
e_{E}	.90	.12	.56	>.99	.30	.12	.03	.48
g_{OI}	.32	.04	.25	.40	-.16	.08	-.33	.003
g_{OE}	.68	.05	.59	.77	.32	.08	.15	.47
g_{NI}	.80	.09	.60	.94	-.34	.04	-.40	-.27
g_{NE}	.92	.05	.79	.99	.23	.12	-.03	.43

Note. Mean parameter estimates and their correlation with the item-source contingency judgments requested at the end of the experiment. Parameters: D = probability of recognizing an item that was presented by any source or was new (assuming that item memory does not differ between item types/trees); d_{E} / d_{U} / d_{I} = probability of remembering the source of an item that was presented by the expected (E), unexpected (U), or irrelevant (I) source; b = probability of guessing that an unrecognized item was old; a_{E} / a_{I} = probability of guessing the expected (E) or irrelevant (I) source given that the item was recognized in the first place; e_{E} / e_{I} = probability of guessing the expected (E) or irrelevant (I) source given that the item was detected to be new; g_{OE} / g_{OI} = probability of guessing the expected (E) or irrelevant (I) source given that the item was guessed to be old; g_{NE} / g_{NI} = probability of guessing the expected (E) or irrelevant (I) source given that the item was guessed to be new. *SD* = Standard Deviation, BCI = Bayesian Credibility Interval. Contingency judgements were averaged for expected-doctor and expected-lawyer items and included in the Bayesian-hierarchical MPT model as covariate.

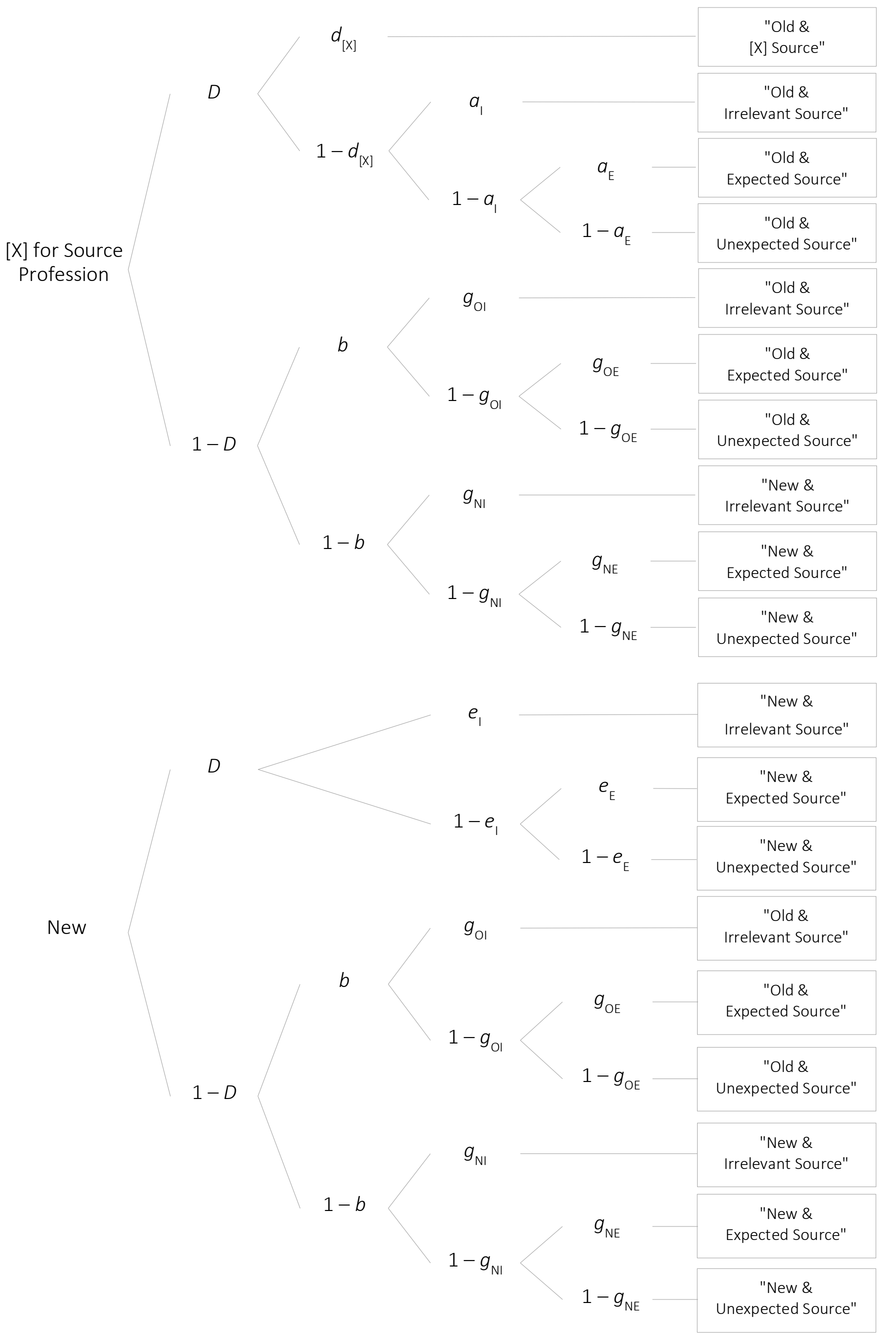
[X] for Source Profession
D
$d_{[X]}$
"Old & [X] Source"
$1 - d_{[X]}$
a_I
"Old & Irrelevant Source"
$1 - a_I$
a_E
"Old & Expected Source"
$1 - a_E$
"Old & Unexpected Source"
$1 - D$
b
g_{OI}
"Old & Irrelevant Source"
$1 - g_{OI}$
g_{OE}
"Old & Expected Source"
$1 - g_{OE}$
"Old & Unexpected Source"
$1 - b$
g_{NI}
"New & Irrelevant Source"
$1 - g_{NI}$
g_{NE}
"New & Expected Source"
$1 - g_{NE}$
"New & Unexpected Source"
New
D
e_I
"New & Irrelevant Source"
$1 - e_I$
e_E
"New & Expected Source"
$1 - e_E$
"New & Unexpected Source"
$1 - D$
b
g_{OI}
"Old & Irrelevant Source"
$1 - g_{OI}$
g_{OE}
"Old & Expected Source"
$1 - g_{OE}$
"Old & Unexpected Source"
$1 - b$
g_{NI}
"New & Irrelevant Source"
$1 - g_{NI}$
g_{NE}
"New & Expected Source"
$1 - g_{NE}$
"New & Unexpected Source"

Figure 1. Graphical illustration of the source monitoring MPT model (adapted from Bell et al., 2020). One exemplar tree for items for a respective profession and one for new items are presented. [X] refers to the respective item type: items are either expected, unexpected, or irrelevant for the sources' profession. In total, the model thus consists of four trees. The trees for items that are expected, unexpected, or irrelevant for the sources' profession can be differentiated by the respective source-memory parameter d; the general structure and all other parameters are equivalent across trees. Parameters represent the following source-monitoring processes that were estimated based on individuals' source-attribution responses (in quotation marks): D = probability of recognizing an item that was presented by any source or was new (assuming that item memory does not differ between item types/trees); $d_{[X]}$ = probability of remembering the source of an item that was presented by the expected (E), unexpected (U), or irrelevant (I) source; b = probability of guessing that an unrecognized item was old; a_E / a_I = probability of guessing the expected (E) or irrelevant (I) source given that the item was recognized in the first place; g_{OE} / g_{OI} = probability of guessing the expected (E) or irrelevant (I) source given that the item was guessed to be old; g_{NE} / g_{NI} = probability of guessing the expected (E) or irrelevant (I) source given that the item was guessed to be new; e_E / e_I = probability of guessing the expected (E) or irrelevant (I) source given that the item was detected to be new.

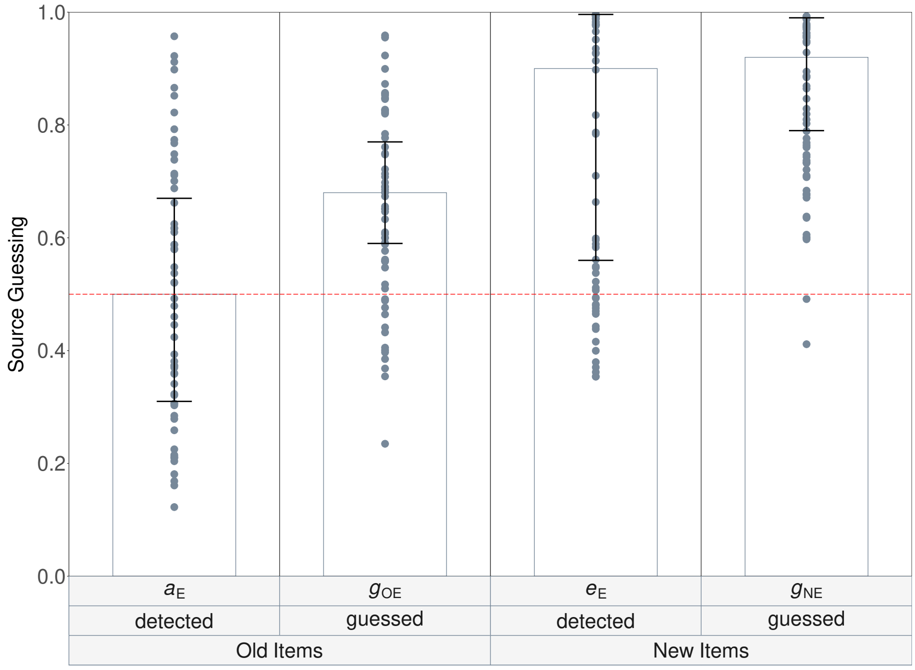

Figure 2. Source-guessing parameters of the source monitoring MPT model (adapted from Bell et al., 2020) for the expected (E) source depending on the recognition status (old vs. new) of items that were either detected or guessed. a_{E} = probability of guessing the expected source given that the item was recognized in the first place; g_{OE} = probability of guessing the expected source given that the item was guessed to be old; e_{E} = probability of guessing the expected source given that the item was detected to be new; g_{NE} = probability of guessing the expected source given that the item was guessed to be new. Red dashed line = chance-level, unbiased source guessing, parameters > .50 indicate schema-based source guessing, parameters < .50 indicate counter schema-based source guessing. Group-level parameter means are displayed as bars, individual parameter estimates are displayed as dots. The error bars represent the 95%-BCI.

This dissertation thesis was supported by the *Graduate School of Economic and Social Sciences* (GESS) at the University of Mannheim.

In order to compensate for the paper waste due to the printing of this thesis,
I donated to https://www.primaklima.org to support the planting of 30 trees in Nicaragua.