Explaining individual differences in cognitive processes underlying hindsight bias

Alisha Coolin, Edgar Erdfelder, Daniel M. Bernstein, Allen E. Thornton & Wendy Loken Thornton

Psychonomic Bulletin & Review

ISSN 1069-9384
Volume 22
Number 2

Psychon Bull Rev (2015) 22:328-348
DOI 10.3758/s13423-014-0691-5

Springer

Psychon Bull Rev (2015) 22:328–348
DOI 10.3758/s13423-014-0691-5

THEORETICAL REVIEW

Explaining individual differences in cognitive processes underlying hindsight bias

Alisha Coolin · Edgar Erdfelder · Daniel M. Bernstein · Allen E. Thornton · Wendy Loken Thornton

Published online: 22 August 2014

Abstract After learning an event's outcome, people's recollection of their former prediction of that event typically shifts toward the actual outcome. Erdfelder and Buchner (*Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 387–414, 1998) developed a multinomial processing tree (MPT) model to identify the underlying processes contributing to this hindsight bias (HB) phenomenon. More recent applications of this model have revealed that, in comparison to younger adults, older adults are more susceptible to two underlying HB processes: recollection bias and reconstruction bias. However, the impact of cognitive functioning on these processes remains unclear. In this article, we extend the MPT model for HB by incorporating individual variation in cognitive functioning into the estimation of the model's core parameters in older and younger adults. In older adults, our findings revealed that (1) better episodic memory was associated with higher recollection ability in the absence of outcome knowledge, (2) better episodic memory and inhibitory control and higher working memory capacity were associated with higher recollection ability in the presence of outcome knowledge, and (3) better inhibitory control was associated with less reconstruction bias. Although the pattern of effects was similar in younger adults, the cognitive covariates did not significantly predict the underlying HB processes in this age group. In sum, we present a novel approach to modeling individual variability in MPT models. We applied this approach to the HB paradigm to identify the cognitive mechanisms contributing to the underlying HB processes. Our results show that working memory capacity and inhibitory control, respectively, drive individual differences in recollection bias and reconstruction bias, particularly in older adults.

Keywords Multinomial processing tree models · Hindsight bias · Individual differences · Cognitive functioning

A. Coolin · A. E. Thornton · W. L. Thornton (✉)
Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada
e-mail: wthornto@sfu.ca

E. Erdfelder
Department of Psychology, University of Mannheim, Mannheim, Germany

D. M. Bernstein
Department of Psychology, Kwantlen Polytechnic University, Surrey, British Columbia, Canada

A. E. Thornton
BC Mental Health and Addictions Research Institute, Burnaby, British Columbia, Canada

Our current knowledge state influences the certainty with which we view the past. This *hindsight bias* (HB) phenomenon typically results in a retrospective tendency to overestimate the predictability of prior events. Research has indicated that HB has important implications for practical and professional decision-making (e.g., Arkes, Wortman, Saville, & Harkness, 1981; Harley, 2007; Leary, 1981). For example, HB may result in people overlooking reasons for why an outcome occurred, potentially resulting in a failure to learn from negative outcomes and overconfidence in future decision-making (see Pezzo & Pezzo, 2007). Given literature suggesting age-related declines in deliberate decision-making processes (Peters, Hess, Vastfjall, & Auman, 2007), older adults may be particularly vulnerable to decision-making errors that stem from HB. However, HB has been relatively unexplored within the context of aging, with only a handful of studies to date (Bayen, Erdfelder, Bearden, & Lozito, 2006; Bernstein, Erdfelder, Meltzoff, Perria, & Loftus, 2011; Coolin, Bernstein, Thornton, & Thornton, 2014; Groß & Bayen, in press). Understanding older adults' susceptibility to HB and how HB impacts independent decision-making will be critical to the development of tools to facilitate decision-making in older adults.

In the existing aging literature, researchers have assessed HB using a memory judgment task (Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988), which involves participants providing original judgments (OJ) to almanac questions (e.g., "How long is the Nile River?"). Later, they learn the correct answers (referred to as correct judgments or CJ) to half of the questions (experimental items), but not the other half (control items), and then recall their original judgments (ROJ) to all the questions. A HB response is characterized by the ROJ typically being closer to the CJ for experimental than for control items (for a review of HB measures, see Pohl, 2007). Across studies, older adults demonstrated a greater tendency to exhibit HB as compared to younger adults (Bayen et al., 2006; Bernstein et al., 2011; Coolin et al., 2014; Groß & Bayen, in press).

One explanation for increased HB in older adults is that declines in episodic memory and executive functioning may influence susceptibility to outcome knowledge when making hindsight judgments (e.g., Bayen, Pohl, Erdfelder, & Auer, 2007; Coolin et al., 2014). Indeed, we previously found that older adults' susceptibility to HB was partly due to age-related declines in episodic memory and inhibition (Coolin et al., 2014). Age-related declines in episodic memory (Hedden & Gabrieli, 2004) may result in outcome knowledge interfering with older adults' recollection of the OJ, creating a recollection bias: better recollection of the OJ for control than for experimental items. If the OJ is not recollected, then it must be reconstructed. During this reconstruction stage, age-related declines in inhibition (Hasher & Zacks, 1988) may result in older adults being unable to suppress the CJ. Subsequently, this information may bias the reconstruction of the forgotten OJ, creating a reconstruction bias: ROJ shifts toward the CJ relative to the OJ.

Erdfelder and Buchner (1998) developed a multinomial processing tree (MPT) model for memory judgment data to estimate the relative contributions of recollection and reconstruction biases to HB (for reviews of MPT models see Batchelder & Riefer, 1999; Erdfelder et al., 2009). The model has 13 parameters and thus is referred to as the *HB13 model*. Each parameter represents a different psychological process that together form the underlying processing tree, which describes the multiple ways in which HB arises. The biases and errors that can occur in each judgment stage are captured by parameters r_C, r_E, b, and c—the four core parameters of the HB13 model. Parameters r_C and r_E represent the probabilities of recalling the OJ for control and experimental items, respectively. Recollection bias is the difference between the recollection probabilities (i.e., $r_C - r_E$). Parameters b and c represent underlying processes that can occur in the reconstruction stage. Parameter b is the probability of reconstruction bias, represented by a shift of the ROJ toward the CJ relative to the OJ (e.g., OJ < ROJ < CJ). Parameter c is the probability of a verbatim CJ adoption, represented by a complete shift of the ROJ to the CJ (e.g., OJ < ROJ = CJ). The two prior studies that have used the HB13 model to investigate age differences in HB have shown that older adults tend to have an overall larger reconstruction bias and recollection bias than do younger adults (Bayen et al., 2006; Bernstein et al., 2011). With regard to parameter c, older adults tend to show more CJ adoptions, particularly when the CJ is accessible during ROJ (Bayen et al., 2006).

A limitation of previous HB13 model applications has been their assumption of fixed process contributions across individuals. When data are aggregated the model provides a single set of 13 parameters for the entire sample and thus cannot account for individual differences. The underlying parameter homogeneity assumption may easily be violated, especially when modeling a cognitive phenomenon. Parameters quantify different cognitive aspects of task performance and people demonstrate variability in their cognitive skills, particularly in samples with large within-group differences (e.g., for older adults, see Smith & Batchelder, 2008). Ignoring this heterogeneity in analyses of aggregated data is also problematic from a statistical point of view, as parameter heterogeneity may distort results of parameter estimation and goodness-of-fit tests (Batchelder & Riefer, 1999; Erdfelder et al., 2009; Klauer, 2006, 2010; Smith & Batchelder, 2008; Stahl & Klauer, 2007). Furthermore, models that adequately describe the response structure at an individual level will often be rejected at the aggregate group level because the parameters vary across individuals (Stahl & Klauer, 2007).

In response to these issues, researchers have developed formal approaches to incorporate individual variability into MPT models (e.g., Klauer, 2006, 2010; Smith & Batchelder, 2010). These approaches are based on hierarchical extensions of traditional MPT models and capture parameter heterogeneity by specifying a distribution of the parameters across individuals. For example, Klauer (2006) proposed a latent-class MPT model that is based on a discrete distribution of parameters. This approach assumes that each participant falls into one of a fixed set of mutually exclusive latent classes. Although the parameter values of all individuals in a certain latent class are assumed to be homogeneous, the model allows for variation in parameter values between classes (Klauer, 2006). In a subsequent article, Stahl and Klauer (2007) provided a computer program (HMMTree) to implement latent-class hierarchical MPT models.

More recently, researchers have argued that the latent-class model's assumption of a discrete distribution, in which participants are sampled from several distinct but homogeneous groups, may be inadequate in certain applications and that a continuous distribution of parameter values across individuals is more reasonable (Smith & Batchelder, 2010) . Therefore, Smith and Batchelder (2010) developed a class of hierarchical models that allows parameters to vary continuously over individuals. The assumption is that each individual's model parameters are sampled independently from a multivariate

Beta distribution. Furthermore, Klauer (2010) proposed a latent-trait approach based on continuous latent variables with an underlying multivariate Gaussian structure. A major advantage of this model is that it can conceivably be extended to include covariates that explain variability in the model parameters.

Although researchers have made considerable progress in implementing interindividual heterogeneity into the estimation of MPT models, these hierarchical models remain quite complex and statistically intensive. Furthermore, the models are limited by distributional assumptions for the parameters. That is, the validity of this family of models hinges on subscribing to a certain distributional assumption for the model parameters, which may not be warranted in all situations. To date, no one has attempted to apply any of these approaches to the HB13 multinomial model, and thus, our understanding of the abilities that impact the processes underlying HB remains incomplete.

For these reasons, we present a variation of Klauer's (2010) latent-trait model that does not require distributional assumptions for the parameters. In addition, our model incorporates individual difference data (i.e., cognitive ability scores) into the HB13 multinomial model to explain variability in the underlying processes. We were particularly interested in modeling HB in older adults because the large variation in cognitive functioning in this population permits us to examine whether individual differences in cognitive functioning could explain heterogeneity in the underlying HB processes. Although there is less variation in cognitive functioning in younger adults, we also applied our model to a younger-adult comparison group to examine whether the general pattern of findings observed in older adults would also hold for this age group.

The objectives of this study were twofold: Our first aim was to present an alternative, easily implemented approach to incorporating heterogeneity into the estimation of MPT models. Although we aimed to explain heterogeneity in terms of cognitive functioning data, our model can be generalized to include any continuous or categorical covariate (e.g., personality, demographic variables, etc.), and thus will have broad applications in the MPT and HB literatures. Our second aim was to extend prior work in the HB literature by identifying the cognitive traits associated with the core underlying HB processes in older and younger adults. We will now briefly review cognitive theories of HB, and follow this with hypotheses regarding the cognitive abilities associated with the underlying processes and a description of our model.

The recollection–reconstruction theory of HB

The HB13 model is based on a recollection–reconstruction theory of hindsight judgments, similar to the two-stage memory judgment theories suggested for other paradigms (cf. Dehn & Erdfelder, 1998; Hell et al., 1988; McCloskey & Zaragoza, 1985; Stahlberg & Maass, 1998). The model assumes that when prompted for a hindsight judgment, individuals first try to recollect their own OJ. For control items, in which outcome knowledge is absent, recollection of one's OJ should depend on episodic memory functioning in the first place, such as the ability to successfully encode, consolidate, and retrieve the OJ from long-term memory. For experimental items, in which outcome knowledge is present, recollection of one's OJ may not only depend on memory functioning but also on the ability to inhibit outcome information. If this information is not inhibited, then it may interfere with the memory trace of the OJ, resulting in poorer recollection of the OJ for experimental than for control items (i.e., recollection bias). Conditional on a failed OJ recollection, individuals then enter a second judgment stage in which they try to reconstruct the OJ on the basis of available context information. In this reconstruction stage, overreliance on outcome knowledge may result in a biased reconstruction of the OJ (i.e., reconstruction bias).

Reconstruction theories have received more attention than recollection theories, and researchers have developed several models to explain the mechanisms by which reconstruction bias can occur (see Christensen-Szalanski & Willham, 1991; Guilbault, Bryant, Brockway, & Posavac, 2004; Hawkins & Hastie, 1990, for reviews). Two popular reconstruction theories are the *anchoring-and-adjustment theory* and the *rejudgment theory*. The anchoring and adjustment theory posits that individuals reconstruct their OJ on the basis of their updated knowledge state (i.e., after learning the CJ), and then adjust this estimate to account for their naïve prior state (i.e., prior to learning the CJ). HB may result from an inadequate adjustment process. The rejudgment theory posits that individuals attempt to repeat the judgment process that they used to generate the OJ (Winman, Juslin, & Björkman, 1998). In this case, HB may result if the contextual information during the OJ stage differs from that of the ROJ stage, or if outcome knowledge alters (i.e., updates) one's knowledge base. Prior work generally supports reconstruction theories of HB (Erdfelder & Buchner, 1998; Hoffrage, Hertwig, & Gigerenzer, 2000; Schwarz & Stahlberg, 2003).

Although reconstruction bias is a major determinant of HB, other theorists have argued that recollection bias also plays a role (e.g., Erdfelder, Brandt, & Bröder, 2007; Nestler, Blank, & Egloff, 2010; Pohl, Bayen, & Martin, 2010). For example, in a review of 11 studies across 34 conditions, Erdfelder et al. (2007) found a small but reliable mean recollection bias estimate of .03 (range: –.05 to .22) in within-subjects manipulations of experimental and control items. This indicates that, on average, participants successfully recalled 3% more of their OJs when they were not shown the CJs. The recollection bias was found to be larger in studies using randomized between-subjects manipulations of outcome knowledge.

These findings are consistent with other reviews (Pohl, 2004), suggesting that recollection bias is a modest but significant contributor to HB. Broader integrative theories of HB have also proposed that both recollection and reconstruction biases are implicated in HB and that outcome knowledge interferes with the recollection and reconstruction of OJs (Blank & Nestler, 2007; Hoffrage et al., 2000; Pohl, Eisenhauer, & Hardt, 2003).

The role of cognitive functioning in HB

Cognitive abilities that have been implicated in age differences in HB include inhibition of irrelevant information, episodic memory, and working memory (e.g., Bayen et al., 2006; Bayen et al., 2007; Coolin et al., 2014; Groß & Bayen, in press). We have previously reported that inhibition and episodic memory partially mediated age-related increases in HB in a sample of older and younger adults (Coolin et al., 2014). Specifically, older age was associated with poorer inhibitory control and episodic memory, which were associated with a tendency to exhibit HB more often. In the present study, we applied an extended version of the HB13 multinomial model to assess whether these cognitive abilities influenced recollection or reconstruction processes, or both.

A potentially relevant theoretical framework is provided by the inhibitory-deficit theory (Hasher & Zacks 1988; see Lustig, Hasher, & Zacks, 2007), which proposes three functions of inhibition: (1) controlling access of irrelevant information from entering working memory, (2) suppressing irrelevant information that has gained access to working memory, and (3) restraining strong but inappropriate responses. Each inhibitory function is less efficient in older than in younger adults, and poorer inhibitory functioning is associated with difficulty in other areas of cognitive functioning (e.g., episodic memory, processing speed, attention; Lustig et al., 2007). Researchers have suggested that several inhibitory functions may be involved in the memory judgment HB task (see Bayen et al., 2006; Bayen et al., 2007). For example, access inhibition may be required to control access of irrelevant CJ information from entering working memory and interfering with recall of task-relevant OJ information. Given a failure in access control, the suppression function may be needed to suppress CJ information that has already gained access to working memory. Finally, restraint inhibition may be required to avoid responding using the highly accessible CJ information. Thus, poorer inhibitory functioning may result in the CJ interfering with recall of the OJ (i.e., recollection bias), and/or biasing the reconstruction of a forgotten OJ (i.e., reconstruction bias).

Erdfelder and colleagues (2007) found support for the role of inhibition in recollection bias in a sample of younger adults. The authors separated inhibition into two types of interference effects referred to as specific and generalized response competition (see Newton & Wickens, 1956). They defined specific response competition as new knowledge impairing memory for prior knowledge for a specific item (e.g., learning the length of the Nile River interferes with memory for one's OJ to this item), and generalized response competition as new knowledge interfering with memory for a set of items (i.e., learning the length of the Nile River interferes with the recall of the OJs to several other items—for example, "How long is the Rhine River?"). Erdfelder and colleagues demonstrated that both interference effects contribute to recollection bias and can be selectively manipulated. Whereas generalized recollection bias increased with the number (i.e., set size) and similarity (i.e., related knowledge domains) of items, specific recollection bias decreased when the encoding and retrieval contexts of experimental items were similar (i.e., enhanced ROJ).

Researchers have also hypothesized that episodic memory plays a role in the recollection stage (Erdfelder & Buchner, 1998; Hell et al., 1988). For example, according to the memory trace strength hypothesis, individuals with poorer episodic memory may have weaker memory representations of their OJs, and thus may be more susceptible to outcome knowledge interfering with retrieval of the OJ (Hell et al., 1988). Furthermore, if the OJ cannot be retrieved, individuals may overrely on the CJ during the reconstruction process (Hell et al., 1988; Pohl et al., 2003). Thus, poorer recall ability could be associated with an increased probability of recollection or reconstruction bias.

Finally, researchers have shown that the accessibility of outcome knowledge in working memory affects HB (Bayen et al., 2006; Groß & Bayen, in press; but see also Nestler, Blank, & von Collani, 2008). Bayen and colleagues (2006) found that when the CJs appeared immediately prior to recall with an instruction for participants to memorize them for a later test (Exp. 2), older and younger adults showed greater overall HB than when the CJs appeared minutes prior to recall with no encoding instructions (Exp. 3). MPT analyses revealed that both age groups demonstrated a significant reconstruction bias across experiments. In contrast, whereas younger adults did not show a recollection bias in either experiment, older adults demonstrated a significant recollection bias (but only in Exp. 2). One explanation may be that when outcome knowledge is available in working memory during recall, it decreases the ability to retrieve the OJ, possibly by creating interference between the OJ and CJ, resulting in a recollection bias.

The present study

We propose a variation of Klauer's (2010) latent-trait model that incorporates cognitive covariates into the core parameter estimation process using a logistic link function. Using the proposed model (referred to as the *logistic HB13 model*), we examined whether heterogeneity in recollection ability and reconstruction bias in older and younger adults can be accounted for by various cognitive abilities previously implicated in HB (e.g., Bayen et al., 2006; Bayen et al., 2007; Coolin et al., 2014; Erdfelder et al., 2007). By observing individual bias parameters, we examined the relationship between the parameters and cognitive abilities using appropriate regression models. For example, if one modeled the person-specific parameter for reconstruction bias, denoted as b_k, as a logistic function of inhibition, the function would be

$$b_k = \frac{\exp(\alpha + \beta \cdot \text{Inhibition}_k)}{1 + \exp(\alpha + \beta \cdot \text{Inhibition}_k)} \quad (1)$$

where α and β are real-valued constants to be estimated and k represents an individual, $k = 1, \ldots, n$. The rationale for this equation is that individuals with better inhibitory control should have a lower probability of reconstruction bias than would individuals with poorer inhibitory control. As we will describe later, higher scores on our inhibition measure reflect poorer performance. Thus, we would expect β to be positive in this equation. If the β coefficient for the inhibition term were significant, one would reject the null hypothesis that b_k does not vary as a function of inhibition. The logistic HB13 model presented here allows us to estimate α and β in the above regression equation and to test hypotheses about these parameters.

On the basis of prior work (Bayen et al., 2006; Bayen et al., 2007; Coolin et al., 2014; Erdfelder et al., 2007), we investigated whether recollection ability and reconstruction bias vary as functions of individual differences in (1) inhibition, (2) episodic memory, and (3) working memory capacity. These particular cognitive abilities often show age-related declines in older-adult samples, and larger interindividual variability is apparent within older than within younger samples (e.g., Christensen et al., 1994; Raz, Ghisletta, Rodrigue, Kennedy, & Lindenberger, 2010; Zelazo, Craik, & Booth, 2004). Such variability in cognition may contribute to differences in the underlying HB processes. We predicted that similar cognitive abilities would contribute to these processes across age groups; however, given the larger variability in cognitive functioning in older adults, we expected the effects to be larger in older than in younger adults.

Our working hypotheses were that r_C would depend on memory functioning, such that better memory would be associated with a higher recollection rate for control items, and r_E would depend on both memory and inhibitory functioning, such that better memory and inhibitory control would be associated with a higher recollection rate for experimental items. Our working hypothesis for reconstruction bias (b) was that each of the three cognitive abilities examined would affect reconstruction bias. First, we expected individuals with poorer inhibitory control to have more difficulty suppressing outcome knowledge. This information might then bias the reconstruction process. Second, we expected individuals with poorer episodic memory to have weaker memory traces of their OJs, and thus to overrely on the CJ to reconstruct forgotten OJs. Third, we expected individuals with lower working memory capacity to have more difficulty separating the multiple pieces of information that were in working memory during recall (e.g., OJ and CJ), and thus to be more likely to exhibit interference between the OJ and CJ.

Method

HB13 model (Erdfelder & Buchner, 1998)

Prior to detailing our proposed extension to the HB13 model, we will first review some of the basic considerations of Erdfelder and Buchner's (1998) HB13 model. The model requires discrete data. Thus, participants' continuous HB judgments on each item are assigned to one of ten possible categories. The categories are a set of all possible rank orders of the OJ–CJ–ROJ, allowing for ties between the OJ and ROJ but excluding ties between the OJ and CJ, because HB cannot be investigated when the OJ and CJ are identical (Erdfelder & Buchner, 1998). Five categories are created to encompass instances when the OJ underestimates the CJ– (1) ROJ < OJ < CJ, (2) ROJ = OJ < CJ, (3) OJ < ROJ < CJ, (4) OJ < ROJ = CJ, and (5) OJ < CJ < ROJ– and another five for instances when the OJ overestimates the CJ– (6) CJ < OJ < ROJ, (7) CJ < ROJ = OJ, (8) CJ < ROJ < OJ, (9) ROJ = CJ < OJ, and (10) ROJ < CJ < OJ. The categories capture whether the ROJ deviates in the direction of the CJ (Rank Orders 3, 4, 5, 8, 9, and 10), deviates in the opposite direction (Rank Orders 1 and 6), or is equal to the OJ (Rank Orders 2 and 7). On the basis of the frequency of responses that fit into each category, the probabilities of the sequence of processes that lead to certain OJ–CJ–ROJ rank orders can be calculated to decompose the HB observed in memory judgment experiments.

Illustration of the HB13 model

Erdfelder and Buchner's (1998) HB13 model contains 13 parameters, each reflecting a different psychological process (see Table 1), that together form the underlying processing trees. The HB13 model is illustrated in Fig. 1. The processing tree in Fig. 1A depicts the model for control items (C), in which the CJ is absent. The first branch identifies whether the OJ underestimates (l_C) or overestimates ($1 - l_C$) the CJ.[1] The

[1] Erdfelder and Buchner (1998) suggested that the size of the HB effect may differ depending on whether an individual's OJ underestimates or overestimates the CJ during the encoding stage. Although the values of the model's parameters (particularly the unbiased reconstruction parameters) may depend on whether the OJ underestimates or overestimates the CJ, the sequence of cognitive processes that occur during the retrieval stage is assumed to be independent of events duringencoding.

Table 1 HB13 model parameters and their psychological interpretations

Parameter	Interpretation
r_{Ck}	Probability of an individual recalling the OJ for a control item.
r_{Ek}	Probability of an individual recalling the OJ for an experimental item.
b_k	Probability of an individual making a biased reconstruction given a failure to recall their OJ.
c	Probability of a CJ adoption in the case of biased reconstructions.
h	Probability of a chance hit of the OJ or CJ in the case of unbiased reconstructions.
g_{11}, g_{g1}	Parameters characterizing the ROJ distribution in the case of unbiased reconstructions without chance hits (for OJ < CJ and OJ > CJ, respectively).
g_{12}, g_{g2}	Parameters characterizing the ROJ distribution in the case of unbiased reconstructions without chance hits (for OJ < CJ and OJ > CJ, respectively).
g_{13}, g_{g3}	Parameters characterizing the ROJ distribution in the case of biased reconstructions without a CJ adoption (for OJ < CJ and OJ > CJ, respectively).
l_C, l_E	Probability of OJ < CJ for control and experimental items, respectively.

upper tree represents processes that occur when the OJ initially underestimates the CJ. If the OJ is successfully recalled, with probability r_C, then the ROJ = OJ. However, if the OJ is not recalled, with probability $(1 - r_C)$, then a reconstruction is required. Because the CJ is absent for control items, the reconstruction process is unbiased, but could hit the OJ or CJ by chance, with probability $2h$. In the case of an unbiased reconstruction without a chance hit, with probability $(1 - 2h)$, parameters g_{11} and g_{12} denote the probabilities of the OJ underestimating the CJ (OJ < CJ). Specifically, if the reconstruction results in an ROJ that is smaller than the CJ (ROJ < CJ), with probability g_{11}, then the ROJ will be smaller or larger than the OJ, with probabilities g_{12} and $(1 - g_{12})$, respectively. In contrast, if the reconstruction results in an ROJ that is larger than the CJ (ROJ > CJ), with probability $(1 - g_{11})$, then the ROJ will be larger than the OJ with probability 1. The lower tree (cases in which OJ > CJ) corresponds to the upper tree described here.

The processing tree in Fig. 1B depicts the model for experimental items (E), in which the CJ is present. This model follows the same basic structure as the control model, with several notable differences. First, the probability of recalling the OJ for experimental items (r_E) may differ from that for the control items, indicating a recollection bias ($r_C - r_E$). If no recollection bias is present, then recollection of the OJ would be equal across item conditions (i.e., $r_C = r_E$); however, if outcome knowledge interferes with recollection ability during recall, then recollection of the OJ may be better for control than for experimental items ($r_C > r_E$). The second difference to note in the experimental model is that given a failure to recall the OJ, the reconstruction process can be biased by outcome knowledge, with probability b. One way in which a biased reconstruction can result is from a verbatim CJ adoption (c). If there is no CJ adoption, with probability $(1 - c)$, a biased reconstruction can also occur if the ROJ falls between the OJ and CJ (OJ < ROJ < CJ or CJ < ROJ < OJ), with a probability of g_{13} or g_{g3}, or if the ROJ is larger or smaller than the CJ and OJ (OJ < CJ < ROJ or ROJ < CJ < OJ), with probability $(1 - g_{13})$ or $(1 - g_{g3})$.

Model equations are derived by summing all of the corresponding branch probabilities that lead to a particular rank-order event, with the probability of a branch being a product of all parameters belonging to that branch. For example, in the control model, the probability of perfectly recalling an OJ when it is an underestimate of the CJ (i.e., ROJ = OJ < CJ) is derived from the product of the corresponding parameters (i.e., $l_C \cdot r_C$). If more than one branch leads to the same rank-order event, then the branch probabilities are summed. The remaining model equations can be found in Table 3 of Erdfelder and Buchner (1998, p. 396). On the basis of these 20 model equations (ten for control, ten for experimental), the parameters are estimated by maximizing the model's likelihood function, and the likelihood ratio G^2 is typically used to test whether the model adequately fits the data (Hu & Batchelder, 1994).

The logistic HB13 model: Model derivation

For ease of comparison, we adapt the notation used in Erdfelder and Buchner's (1998) HB13 model. Let i denote the experimental or control item condition ($i = 1, 2$), j denote the rank order ($j = 1, \ldots, 10$), and k denote the participant ($k = 1, \ldots, n$). Let $Y_{i,j,k}$ represent the number of observations in condition i and rank order j for participant k, and $p_{i,j,k}$ the corresponding probability for the same event category. The model parameters for individual k are summarized in a parameter vector θ_k, and the vector of individual characteristics (e.g., cognitive ability test scores) for participant k is denoted by $\mathbf{X}_k$. In addition to the observed individual characteristics, the total number of HB rank-order observations for condition i and participant k is $N_{i,k}=\sum_{j=1}^{10} Y_{i,j,k}$. We adapt Erdfelder and Buchner's expression for the probability of observing a vector of sample frequencies to an individual level. For participant k in condition i, this probability follows a multinomial distribution, as follows:

$$\Pr\left(Y_{i,1,k}, \ldots, Y_{i,10,k}, \mathbf{X}_k\right) = N_{i,k}! \prod_{j=1}^{10} \frac{p_{i,j,k}(\theta_k, \mathbf{X}_k)^{Y_{i,j,k}}}{Y_{i,j,k}!}. \quad (2)$$

Apart from introducing separate multinomial distributions for each individual k, the main innovation in Eq. 2, as compared to Erdfelder and Buchner's (1998, p. 390) Eq. 1, is that

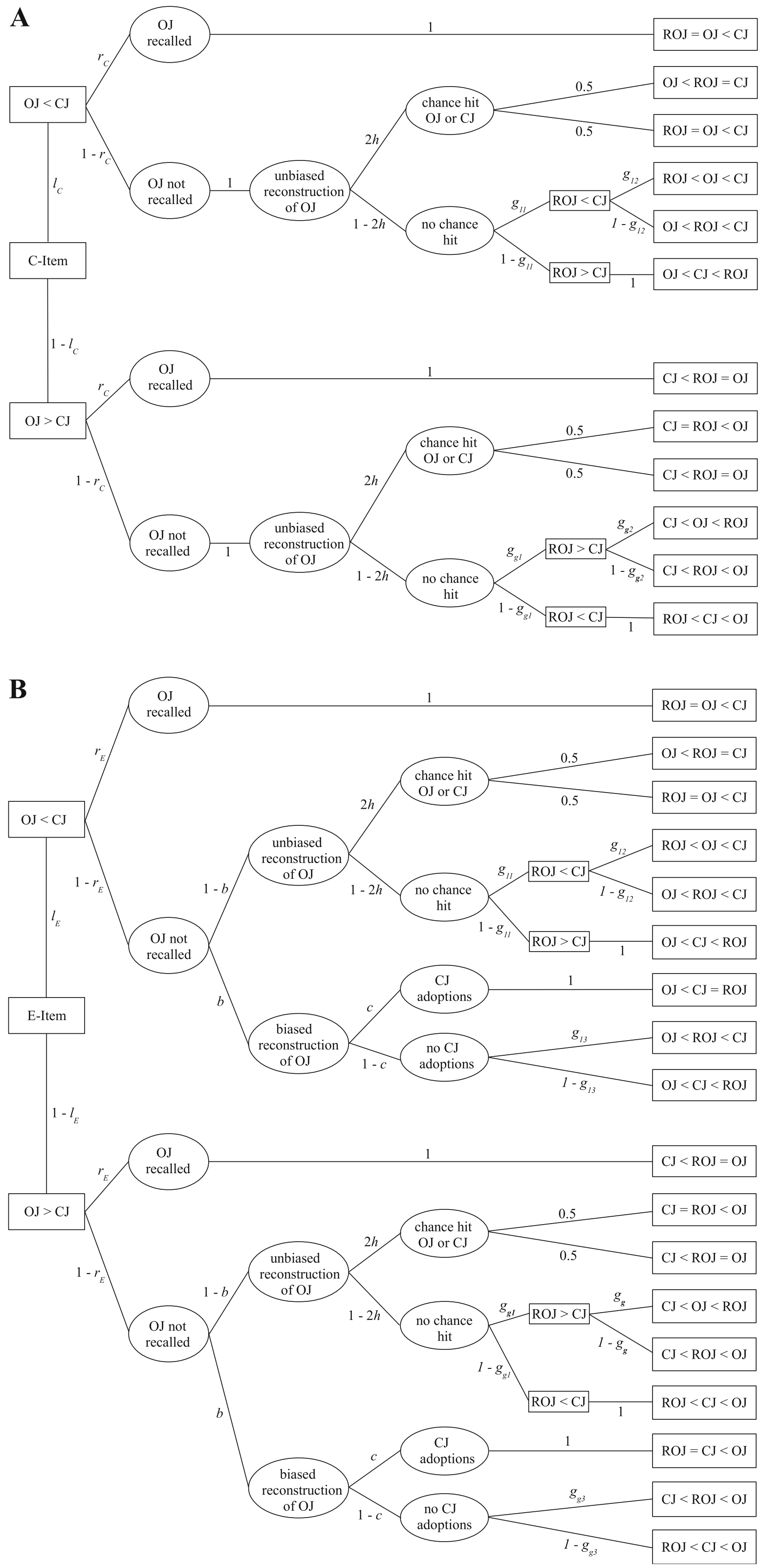

Fig. 1 Erdfelder and Buchner's (1998) multinomial processing tree model for hindsight bias. (A) Processing tree for control items. (B) Processing tree for experimental items. OJ = original judgment, ROJ = recall of the original judgment, CJ = correct judgment. From "Decomposing the Hindsight Bias: A Multinomial Processing Tree Model for Separating Recollection and Reconstruction in Hindsight," by E. Erdfelder and A. Buchner, 1998, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, pp. 392–393. Copyright 1998 by the American Psychological Association

$p_{i,j,k}$ is now not only a function of the parameter vector θ_k, but also of the individual characteristics $\mathbf{X}_k$. Note that individual variability in $\mathbf{X}_k$ makes the category probabilities $p_{i,j,k}$ vary between participants, even if the parameter vector θ_k happens to be constant across individuals. Because of this model property, we can allow heterogeneity into the model with only a small increase in the number of to-be-estimated parameters. For notational ease, we gather the individual sample frequencies for condition i into a vector $\mathbf{Y}_i = (Y_{i,1,1}, \ldots, Y_{i,10,1}, \ldots, Y_{i,1,n}, \ldots, Y_{i,10,n})$, where n represents the number of participants. Similarly, we gather the individual characteristics into a sample vector denoted as $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$. By experimental design, there is independence across the participants' data; thus, the joint probability of observing all participants' data in a given condition is

$$\Pr(\mathbf{Y}_i, \mathbf{X}) = \prod_{k=1}^{n} \Pr\left(Y_{i,1,k}, Y_{i,2,k}, \ldots, Y_{i,10,k}, \mathbf{X}_k\right). \quad (3)$$

Like Erdfelder and Buchner, we assume independence between the experimental and control observations. Thus, the probability of observing the entire sample is

$$\Pr(\mathbf{Y}, \mathbf{X}) = \prod_{i=1}^{2} \Pr(\mathbf{Y}_i, \mathbf{X}), \text{where} \quad \mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2). \quad (4)$$

Mapping individual characteristics to HB parameters

In accordance with the HB13 model, we model the category probabilities as being the product of an underlying multinomial processing tree. However, to incorporate the cognitive data, we conceive the process parameters as being derived from underlying latent variables that may vary between individuals. For example, we model reconstruction bias as occurring on item m for participant k if a latent variable b^*_{km} exceeds a threshold. More precisely,

$$\text{reconstruction bias} = \begin{cases} 1 & \text{if} \quad b^*_{km} > 0 \\ 0 & \text{if} \quad b^*_{km} \leq 0 \end{cases}. \quad (5)$$

By implication, the probability of reconstruction bias for participant k is $b_k = p(b^*_{km} > 0)$. We then need to model the relationship between the observed individual characteristics and the latent variables. To incorporate cognitive data, we follow Klauer (2010) and Ansari, Vanhuele, and Zemborain (2008, as cited in Klauer, 2010) by modeling the latent variable b^*_{km} as a function of the cognitive data and a random error component:

$$b^*_{km} = \alpha_b + \beta_{b1} X1_k + \beta_{b2} X2_k + \ldots + \beta_{bC} XC_k + \varepsilon_{km}, \quad (6)$$

where α_b and β_{bc}, $c = 1, \ldots, C$, are parameters to be estimated; Xc_k, $c = 1, \ldots, C$, denotes the value of variable Xc for participant k; and ε_{km} is the random error of participant k on item m. Whereas Klauer (2010) modeled ε_{km} as being normally distributed, Ansari et al. (2008, as cited in Klauer, 2010) modeled ε_{km} as being logistically distributed. We adopt the logistic distribution for its simplicity, since it allows for an analytical solution for the data likelihood.[2] However, we conducted additional analyses using a normally distributed error term. As we will show in Appendix 1, the results for the normal error model proved to be very similar, providing converging evidence for the logistic error model.

A bias will occur whenever $\varepsilon_{km} > -\alpha - \beta_{b1} X1_k - \beta_{b2} X2_k - \ldots - \beta_{bC} XC_k$. Thus, the probability of reconstruction bias for participant k, b_k, is the probability that this event occurs. Because ε_{km} is logistically distributed with expectation zero, this probability is

$$b_k = \frac{\exp(\alpha_b + \beta_{b1} X1_k + \beta_{b2} X2_k + \ldots + \beta_{bC} XC_k)}{1 + \exp(\alpha_b + \beta_{b1} X1_k + \beta_{b2} X2_k + \ldots + \beta_{bC} XC_k)}. \quad (7)$$

For parameters that are not modeled as functions of cognitive data, the logistic function simplifies to a constant, as in the aggregated HB13 model. Another possibility is to allow the probabilities to vary freely across individuals. In this case, the latent variable would be modeled as $b^*_{km} = \alpha_{bk} + \varepsilon_{km}$, and the probability of bias would be $b_k = \frac{\exp(\alpha_{bk})}{1 + \exp(\alpha_{bk})}$.

Given that (a) a larger b_k indicates a larger probability of reconstruction bias and (b) higher episodic memory and working memory scores indicate better performance, we would expect the coefficients for episodic memory and working memory to be negative. Conversely, because higher scores on the inhibition measure indicate poorer performance (i.e., increased latencies), we would expect the coefficient for inhibition in the function for b_k to be positive. This would indicate that better episodic memory, higher working memory capacity, and better inhibitory control are associated with a lower probability of reconstruction bias. Analogously, we would expect the coefficient for episodic memory in the corresponding logistic functions for r_{Ck} and r_{Ek} to be positive, and the effect of inhibition on r_{Ek} to be negative. This would indicate that better episodic memory and better inhibitory control, respectively, are associated with a higher probability of recalling one's OJ.

Model-based data analysis

The logistic model outlined in the previous section can only be used to test our primary hypotheses when it provides a good

[2] Note that another major difference from Klauer's (2010) latent-trait approach is that we make a distributional assumption about the latent error term only, whereas he makes an additional assumption about the joint distribution of the latent model parameters (i.e., the assumption of a multivariate normal distribution).

account of the empirical data. To check the empirical adequacy of this model, we compared it to a series of alternative models incorporating individual variability. More precisely, we introduce four candidate models: (1) the general multinomial model, (2) the unconstrained HB13 model, (3) the simplified HB13 model, and (4) the logistic HB13 model. Table 2 presents a summary of each of these models. In addition, to facilitate comparisons with previous research on age differences in HB, we also provide analyses based on the original HB13 model when it is applied to data aggregated across individuals (see Appendix 2). To anticipate, we observed age differences in the core HB parameters that are quite similar to those obtained in prior work (Bayen et al., 2006; Bernstein et al., 2011).

The general multinomial model was our reference model and represents the $2 \cdot 10$ probabilities of the experimental and control HB data, respectively, for each of the n individuals within an age group without any restriction. In contrast, the unconstrained HB13 model assumes that the HB13 model holds for each individual and allows for unconstrained individual variability in each of its 13 parameters. We first assessed the fit of the unconstrained HB13 model to the probabilities of the observed HB data (i.e., the general multinomial model) for the n individuals in the respective age groups. Then we introduced the simplified HB13 model, which is a more parsimonious version of the unconstrained HB13 model with fewer parameters. More precisely, we conducted heterogeneity tests on each of the 13 parameters to determine whether any of the unconstrained parameters could be simplified to a constant. On the basis of the results of these heterogeneity tests, in the simplified HB13 model, homogeneous parameters were modeled as constants and heterogeneous parameters were allowed to vary between individuals. Finally, we introduced the logistic HB13 model, which replaces the unconstrained core parameters (b_k, r_{Ck}, and r_{Ek}) of the simplified HB13 model with logistic functions of the cognitive covariates, allowing us to test the role of the cognitive covariates in these parameters. We first applied the above series of model steps to our older-adult data set and then applied the same model steps to our younger-adult comparison group.

Prior to presenting tests of our substantive hypotheses, we present a model selection analysis to determine which of our four candidate models best approximated the "true" model underlying the data. One possibility was to examine which model provided the closest fit to the data, but this criterion did not take into account model complexity. Although complex models with many free parameters tend to be more flexible in fitting different sets of data, it is more desirable to select a model that balances model accuracy (i.e., adequately accounts for the data) and parsimony (i.e., using few parameters; Myung, 2000; Myung & Pitt, 1997; Wagenmakers & Farrell, 2004). The main reason for this is that the flexibility of complex models leads to a disadvantage in predicting future data, due to the increased probability of sampling error influencing the parameter estimates (e.g., Klauer, Stahl, & Erdfelder, 2007). To overcome this problem, we chose the Akaike information criterion (AIC; Akaike, 1974) for model selection, because this criterion (1) penalizes nonparsimonious models, and thus balances descriptive accuracy and parsimony, (2) assesses the goodness-of-fit of the model for predicting future data from the model as fitted from the observed data, and (3) is commonly used for choosing between stochastic models of cognition (e.g., Ashby, Prinzmetal, Ivry, & Maddox, 1996; Burnham & Anderson, 2002; Klauer et al., 2007; Myung, 2000).

For each of the candidate models q, $q = 1, 2, \ldots, Q$, the AIC is defined as follows:

$$\mathrm{AIC}_q = -2 \cdot \ln(L_q) + 2P_q, \tag{8}$$

where L_q is the maximized likelihood for candidate model q, and P_q is the number of parameters in candidate model q. As can be seen in Eq. 8, AIC penalizes for lack of simplicity of the model, such that AIC values increase with the number of

Table 2 Summary of the various multinomial processing tree models

Model	Interpretation	Number of Free Parameters
General multinomial	Raw category frequencies ($p_{i,j,k}$)	$2 \cdot 9 \cdot n$
HB13	HB13 model with aggregated data (Erdfelder & Buchner, 1998)	13
Unconstrained HB13	HB13 model applied to n individual data sets. Parameters are $r_{Ck}, r_{Ek}, b_k, \ldots, k = 1, \ldots, n$	$13 \cdot n$
Simplified HB13	HB13 model with core parameters b_k, r_{Ck}, and r_{Ek} and ancillary parameters g_{11k} and g_{g1k} unconstrained, and c and the remaining ancillary parameters constant	$5 \cdot n + 8$
Logistic HB13	HB13 model with b_k, r_{Ck}, and r_{Ek} constrained as (four-parametric) logistic functions of the three individual-difference variables, g_{11k} and g_{g1k} unconstrained, and c and the remaining ancillary parameters constant	$4 + 4 + 4 + 2 \cdot n + 8$

model parameters. The AIC-best model is the candidate model that provides the best balance between goodness-of-fit and parsimony, and is identified by the lowest AIC value. Notably, it can be shown that choosing the model with the lowest AIC value is asymptotically equivalent to choosing the model with the smallest expected information loss (minimizing the Kullback–Leibler discrepancy) when approximating a true model (Wagenmakers & Farrell, 2004). This is a highly desirable property of our model selection criterion. However, it does not provide a simple interpretation of raw AIC scores.

Given the difficulty in interpreting the statistical importance of raw AIC differences among candidate models, Wagenmakers and Farrell (2004) developed a method to easily transform raw AIC values into Akaike weights (Akaike, 1978; Burnham & Anderson, 2002), which represent the relative likelihood of each model and can be interpreted in terms of conditional probabilities. To obtain the Akaike weights for our four candidate models, we first calculated their ΔAIC scores, where

$$\Delta_q(\mathrm{AIC}) = \mathrm{AIC}_q - \mathrm{minAIC}, \tag{9}$$

with minAIC being the smallest AIC value and, by implication, the best model having a ΔAIC score of 0. Next, we calculated the relative likelihood of each candidate model, that is

$$\exp\{-0.5\Delta_q(\mathrm{AIC})\}. \tag{10}$$

Finally, we calculated the Akaike weights w_q(AIC) by normalizing the relative model likelihoods, where

$$w_q(AIC) = \frac{\exp\{-0.5\Delta_q(AIC)\}}{\sum_{q=1}^{Q}\exp\{-0.5\Delta_q(AIC)\}}, \tag{11}$$

such that $\sum w_q(\mathrm{AIC}) = 1$. Weight w_q(AIC) can be interpreted as a probability estimate that M_q is the best model in terms of minimizing information loss, given the data and the set of candidate models (Wagenmakers & Farrell, 2004). Thus, we identified the AIC-best model for older adults and the AIC-best model for younger adults on the basis of the candidate model with the smallest AIC score and the highest Akaike weight in each respective age group.

Statistical analyses

We programmed the likelihood functions for all candidate models in MATLAB and estimated different versions of the HB13 model using MATLAB's optimization toolbox. The *p* values and 95% confidence intervals reported for the model fit tests and for the logistic HB13 hypothesis tests were based on 500 bootstrapped samples. Given the length of time required to complete the heterogeneity tests on each parameter, these analyses were based on 100 bootstrapped samples. We used the parametric bootstrap exclusively to determine *p*-values, because *p* values based on the asymptotic chi-square distribution of G^2 under H0 can be severely misleading when the data include many zero cells, which was the case for our data. Thus, we dispensed with asymptotic *p* values and replaced them with estimates based on the exact distribution of G^2 under H0 using the parametric bootstrap (Efron & Tibshirani, 1993). For clarity, we denote all *p*-value estimates based on the parametric bootstrap method by p_b. In accordance with prior work (e.g., Bayen et al., 2006), we set an alpha level of .01 for model fit tests because we did not want to reject a model that only slightly differed from the comparison model. Importantly, even at this alpha level we had sufficient power to detect moderate (i.e., w = .3; Cohen, 1988) but not small (i.e., w = .1) deviations from the comparison model.[3] Thus, setting alpha to .01 provided a balance between rejecting small deviations from the comparison model and detecting moderate model misfit issues. We provide more detailed power information in the Results section. We used the standard .05 alpha level for all other analyses. We chose this conventional alpha level for heterogeneity tests because we were less concerned with Type I errors, since this would simply result in a homogeneous parameter being modeled freely.

Participants

We present a reanalysis of data from a previously conducted study examining HB differences between older and younger adults (Coolin et al., 2014). We recruited 60 healthy community-dwelling older adults and 64 college-aged younger adults. Community-dwelling adults over the age of 65 were recruited through newspaper advertisements in the Metro Vancouver area, the university staff union e-mail list, and academic aging seminars conducted by the corresponding author. The younger adults were introductory psychology students who received course credit for participating.

Within each age group, we examined the distributions of observed values from the general multinomial model and predicted values from the logistic HB13 model in order to identify individuals who had predicted values that were substantially different from their observed values. We used the following criterion to identify extreme values: An individual was an outlier if the residual $y_i - \hat{y}_i$ was larger than three times the standard deviation of the $n-1$ remaining residuals within that age group. Our outlier analyses revealed no outliers in our older-adult group, but two outliers in our younger-adult group. Thus, after excluding these two outliers, the younger-adult sample consisted of 62 individuals (M_{age}= 20.10 years, range = 18 to

[3] Power analyses were computed with the G*Power 3 program (Faul, Erdfelder, Lang, & Buchner, 2007).

25; 45 female, 17 male), and the older-adult sample remained at 60 (M_{age} = 72.50 years, range = 65 to 87; 35 female, 25 male).

Participants met the following inclusion criteria: (a) English fluency; (b) a minimum of grade 7 education (i.e., completion of primary school; older, M = 14.28 years of education, SD = 2.96; younger, M = 13.39 years, SD = 1.94); (c) no major visual (corrected vision ≤ 20/50) or hearing impairments; (d) absence of self-reported diagnosis of major psychotic illness by a physician, concurrent acute illness that might affect testing, neurological disorder, major organ failure, severe traumatic head injury, history of a stroke that affected daily living activities, and history of dementia; and (e) alcohol consumption of less than 3 ounces/day. All older-adult participants received a score greater than 24 on the Mini Mental State Examination (Folstein, Folstein, & McHugh, 1975). We included performance on a receptive vocabulary test (the Peabody Picture Vocabulary Test–3; Dunn & Dunn, 1997; scores: older, M = 188.33, SD = 9.42; younger, M = 180.39, SD = 7.65) to index general knowledge of word meanings. Please refer to Coolin et al. (2014) for descriptive statistics regarding the traditional HB indices (e.g., Hell's et al., 1988) and correlations between these indices and the cognitive ability data (see Table 4 below).

Measures

We assessed HB using a memory judgment design, which consisted of an Original Judgment (OJ) questionnaire and a Recall of the Original Judgment (ROJ) questionnaire. The OJ questionnaire included 54 almanac questions (see Appendix 3) adopted from Bayen et al. (2006) and Hardt and Pohl (2003). One question, "When was Socrates born?" was excluded from the analyses because some participants responded in BC and others in AD, which prohibited the aggregation of responses across participants. We provided participants with the metric-system unit with which they had to respond. We randomized the order of the questions, and then presented questions in a fixed order to all participants. The ROJ questionnaire required participants to recall their OJs to the 54 almanac questions. Participants were told that they would learn the CJs to half of the questions (experimental items), but not the other half (control items), and that their task was to recall their OJs to all of the questions. To control for the content of specific questions, we counterbalanced the control and experimental items by randomly assigning participants to one of two versions. In Version 1, the first 27 items were experimental and the last 27 were control, and vice versa in Version 2.

We measured inhibition using the Color-Word Inteference test, or "Stroop test," from the Delis–Kaplan Executive Function System (D-KEFS; Delis, Kaplan, & Kramer, 2001). This test involves inhibiting a dominant verbal response (word reading) in favor of a less dominant response (color naming). Because performance is measured by time to completion, we used the recommended procedure for minimizing the effect of processing speed by subtracting the baseline (color naming condition) from the inhibition condition (Delis et al., 2001). Although this "Stroop" test is typically labeled as a prepotent response inhibition task, like many tests of inhibition, it is not a process-pure measure, and likely involves several inhibitory processes (see Lustig et al., 2007). For example, the access function of inhibition may be involved in preventing task-irrelevant word information from entering working memory. Given a failure in access control, suppression and restraint functions may be involved in avoiding a dominant reading response. Each of these inhibitory processes may be relevant to the demands of the memory judgment task.

We measured episodic memory using the long-delay free recall trial (number of words recalled) of the California Verbal Learning Test–2 (CVLT-II; Delis, Kramer, Kaplan, & Ober, 2000), which assesses the ability to recall a list of 16 words after a 20-min delay. Finally, we measured working memory capacity using raw scores obtained on the Letter–Number Sequencing and Backward Digit Span tests of the Wechsler Adult Intelligence Scale–III (WAIS-III; Wechsler, 1997). In the former task, the examiner reads different combinations of numbers and letters and the participant must recall the number(s) in ascending order, followed by the letter(s) in alphabetical order. In the latter task, the examiner reads a string of digits and the participant reproduces the digits in the reverse order. Given the strong correlations between these two measures in our sample (older, r = .64, p < .001; younger, r = .47, p < .001), we created a working memory composite by converting the data on both variables to z scores and then summing them (Edgington, 1995, p. 183).

Procedures

Trained research assistants tested participants individually in a single 2-h session at the Simon Fraser University Cognitive Aging Laboratory. The session began with the OJ questionnaire of the memory judgment HB task. Following the OJ questionnaire, participants completed the battery of cognitive tests, which lasted approximately 90-min. Immediately following this retention period, participants completed the ROJ questionnaire. Completion of both the OJ and ROJ questionnaires was self-paced; participants took 10–20 min to complete each questionnaire.

Results

Older-adult model tests

Goodness-of-fit of the unconstrained HB13 model with respect to the general multinomial model Our first statistical test concerned the fit of the unconstrained HB13 model to the probabilities of the observed HB data for each of the n individuals in the older-adult group. We compared the unconstrained HB13 model against the general multinomial model for $2 \cdot 10$ data categories and n individuals using parametric bootstrapping. Each model contains $S \cdot n$ parameters, where S denotes the number of unconstrained parameters in the respective HB model. For $n = 60$ participants, the unconstrained HB13 model has $13 \cdot 60 = 780$ parameters, and the general multinomial model has $18 \cdot 60 = 1{,}080$ parameters. With an alpha level of .01, $N = 3{,}135$ (60 participants · 53 items – 45 missing responses), and $df = 300$, we had adequate power (>.99) to detect moderate deviations ($w = .3$) from the general multinomial model, but insufficient power (.15) to detect small deviations ($w = .1$). On the basis of 500 bootstrapped samples, we initially failed to find an acceptable model fit, $\Delta G^2(300) = 337.95$, $p_b < .001$. Following the procedures of Bayen et al. (2006), we identified two items that had a disproportionately high number of observations in the ROJ = OJ < CJ category relative to the ROJ = OJ > CJ category, and one item that had a disproportionately high number of observations in the CJ < ROJ < OJ relative to the OJ < ROJ < CJ category. Thus, these three items violated the symmetry assumption of the model. After excluding these problematic items, the remaining data set consisted of 50 items and had an acceptable model fit, $\Delta G^2(300) = 309.18$, $p_b = .01$.

Simplification of the unconstrained HB13 model To determine the model specification of the simplified HB13 model, we tested for heterogeneity in each of the 13 model parameters. This involved testing the unconstrained HB13 model against a null model in which the parameter of interest was a constant and the remaining 12 parameters were unconstrained. If we rejected the null that the parameter was equal across individuals, then in the simplified HB13 model the corresponding parameter would be left unconstrained. Each null model had $12 \cdot 60 + 1 = 721$ parameters. With an alpha level of .05, $N = 2{,}955$, and $df = 59$, we had adequate power (>.99) to detect moderate deviations ($w = .3$) of the unconstrained HB13 model from each of the null models, but low power (.75) to detect small deviations ($w = .1$).

Table 3 depicts the results of the heterogeneity tests, as well as summary statistics on these parameters. On the basis of 100 bootstrapped samples, we rejected the null model that the ancillary parameter was a constant for two of the nine ancillary parameters—namely, g_{11k} and g_{g1k}. These parameters affect the distribution of unbiased OJ reconstructions (i.e., reconstructions given no perfect OJ recollection) and represent probabilities of reconstructing ROJs that deviate from the CJ in the direction of the OJ. With regard to the core parameters, we rejected the null models that the core parameters b_k, r_{Ck}, and r_{Ek} were equal across individuals. As expected, participants differed in how often they shifted their ROJ toward the CJ relative to the OJ (i.e., reconstruction bias), as well as in their ability to recollect their OJs perfectly, in both the presence (r_{Ek}) and the absence (r_{Ck}) of outcome knowledge. Conversely, we accepted the null model that parameter c_k was equal across individuals.

Table 3 Heterogeneity tests of the HB13 parameters based on the older-adult data set ($n = 60$)

Parameter	Unconstrained Coefficient			
	Mean	*SD*	ΔG^2	p_b
Ancillary Parameters				
g_{11k}	.81	.15	87.04	.03
g_{g1k}	.86	.11	88.53	.046
g_{12k}	.48	.20	69.87	.46
g_{g2k}	.49	.17	69.38	.41
g_{13k}	.67	.29	60.34	.89
g_{g3k}	.65	.32	55.16	.62
l_{Ek}	.44	.11	73.75	.15
l_{Ck}	.45	.09	50.89	.76
h_k	.01	.02	22.15	>.99
Core Parameters				
b_k	.46	.26	81.95	<.001
r_{Ck}	.23	.10	108.35	<.001
r_{Ek}	.20	.13	153.12	<.001
c_k	.03	.10	22.65	>.99

The means and standard deviations are based on the older-adult unconstrained HB13 model. The ΔG^2 statistics represent the model tests comparing the unconstrained HB13 model to a null model in which the parameter of interest was a constant and the remaining 12 parameters were unconstrained. The p value (denoted as p_b) for ΔG^2 is based on 100 bootstrapped samples. A significant ΔG^2 indicates heterogeneity in the parameter, whereas a nonsignificant ΔG^2 indicates homogeneity

On the basis of the results of these heterogeneity tests, we defined the simplified HB13 model as follows: The ancillary parameters g_{11k} and g_{g1k} and the core parameters b_k, r_{Ck}, r_{Ek} were unconstrained across individuals, and parameter c and the remaining seven ancillary parameters were constrained to be equal across individuals. The simplified model has $S \cdot n + z$ parameters, where S denotes the number of unconstrained parameters and z denotes the number of parameters that were simplified to constants. Thus, the simplified HB13 model had $5 \cdot 60 + 8 = 308$ parameters. With an alpha level of .01, $N = 2{,}955$, and $df = 472$, we had adequate power (>.99) to detect moderate deviations ($w = .3$) from the unconstrained HB13 model, but insufficient power (.09) to detect small deviations ($w = .1$). As expected, on the basis of 500 bootstrapped samples, the simplified HB13 model had an acceptable model fit, $\Delta G^2(472) = 425.76$, $p_b = .19$.

Logistic HB13 model We then introduced the logistic HB13 model by replacing the unconstrained core parameters b_k, r_{Ck}, and r_{Ek} in the simplified HB13 model with logistic functions of inhibition, episodic memory, and working memory capacity, respectively. The logistic HB13 model has $S \cdot n + (3 \cdot (C + 1)) + z$ parameters, where S is the number of unconstrained HB13 parameters, C is the number of cognitive variables in each bias function (three, in our case), the 1 accounts for the constant terms in each bias function, and z is the number of parameters that were simplified to a constant. Thus, the number of parameters in the logistic HB13 model was $2 \cdot 60 + (3 \cdot (3 + 1)) + 8 = 140$.

Model selection for the older-adult sample We compared the previously described four candidate models (general multinomial, unconstrained HB13, simplified HB13, and logistic HB13) to identify which model provided the best approximation of the true model underlying our older-adult data. The AIC values and Akaike weights for each of the models are shown in Table 4. The model associated with the smallest AIC value and the highest Akaike weight was the logistic HB13 model. In fact, the probability estimate that the logistic HB13 model was the best of our four candidate models was very close to 1. Also note that the AIC values decrease with each model introduced, indicating that the models became progressively better in terms of the compromise between descriptive accuracy and parsimony as we progressed through our model hierarchy.

Hypothesis testing Using the logistic HB13 model, we tested our primary hypotheses regarding the role of cognitive functioning in the underlying HB processes. This analysis involved bootstrapping the 95% confidence intervals for the beta coefficients on each cognitive covariate for each of the core parameters.[4] As is shown in Table 5, our analyses of older adults revealed that episodic memory was the only significant predictor of r_{Ck}, $\beta = 0.06$, 95% CI = [0.02, 0.10], $p_b < .001$, such that one standard deviation increase in episodic memory scores was associated with a 3.6% increase in the mean estimated r_{Ck}. Yet both episodic memory, $\beta = 0.07$, 95% CI = [0.03, 0.11], $p_b < .001$, and inhibition, $\beta = -0.02$, 95% CI = [–0.04, –0.01], $p_b < .001$, were significant predictors of r_{Ek}: One standard deviation increase in episodic memory scores was associated with a 4.3% increase in the mean estimated r_{Ek}, and one standard deviation increase in inhibition scores was associated with a 7.9% decrease in mean estimated r_{Ek}. Furthermore, in the older-adult sample, working memory capacity was a marginally significant predictor of r_{Ek}, $\beta = 0.06$, 95% CI = [–0.003, 0.13], $p_b = .056$, such that one standard deviation increase in working memory scores was associated with a 1.9% increase in mean estimated r_{Ek}. The significant contribution of inhibition and the marginally significant contribution of working memory capacity to r_{Ek} but not to r_{Ck} suggest that these abilities explain the reduction in recollection rates due to the presentation of the CJ (i.e., recollection bias). Finally, inhibition was the only significant predictor of b_k, $\beta = 0.04$, 95% CI = [0.02, 0.07], $p_b < .001$, such that one standard deviation increase in inhibition scores was associated with a 11.5% increase in mean estimated b_k.

[4] Given that general slowing of processing speed often accounts for age-related changes in other cognitive functions (Salthouse, 2000), we conducted an additional analysis to identify whether our findings held after including a measure of processing speed (the Digit Symbol Coding subtest of the WAIS-III). The findings from our main analyses held, suggesting that processing speed did not account for the effects of the cognitive covariates on the core parameters. Furthermore, processing speed did not significantly contribute to the core HB processes.

Younger-adult model tests

Goodness-of-fit of the unconstrained HB13 model with respect to the general multinomial model Following the same model hierarchy as in older adults, we began by assessing the goodness-of-fit fit of the unconstrained HB13 model in our younger-adult comparison group. We based our analyses on the 50-item data set that provided an acceptable model fit in our older-adult group. For $n = 62$ participants, the unconstrained HB13 model has $13 \cdot 62 = 806$ parameters, and the general multinomial model has $18 \cdot 62 = 1{,}116$ parameters. With an alpha level of .01, $N = 3{,}085$ (62 participants · 50 items – 15 missing responses), and $df = 310$, we had adequate power (>.99) to detect moderate deviations ($w = .3$) from the general multinomial model, but insufficient power (.14) to detect small deviations ($w = .1$). On the basis of 500 bootstrapped samples, using the 50-item data set we found an acceptable model fit, $\Delta G^2(310) = 287.49$, $p_b = .82$.

Simplification of the unconstrained HB13 model Rather than conducting separate heterogeneity tests for younger adults, we assessed whether the simplified HB13 model that we defined for older adults was also an acceptable model for younger adults. This was necessary to avoid varying results across age groups due to different model specifications. The simplified model has $S \cdot n + z$ parameters, where S denotes the number of unconstrained parameters and z denotes the number of parameters that were simplified to constants. Thus, the younger-adult simplified HB13 model had $5 \cdot 62 + 8 = 318$ parameters. With an alpha level of .01, $N = 3{,}085$, and $df = 488$, we had adequate power (>.99) to detect moderate deviations ($w = .3$) from the unconstrained HB13 model, but insufficient power (.09) to detect small deviations ($w = .1$). As expected, on the basis of 500 bootstrapped samples, the simplified HB13 model for the older-adult sample also fit the younger-adult data, $\Delta G^2(488) = 425.43$, $p_b = .02$. Thus, we used the same model specification for the simplified HB13 model in older and younger adults.

Logistic HB13 model We then introduced the logistic HB13 model by replacing the unconstrained core parameters b_k, r_{Ck}, and r_{Ek} in the simplified HB13 model with logistic functions of

Table 4 AIC values and Akaike weights for each of the four candidate hindsight bias (HB) models by age group

Model	AIC_q		Δ_q (AIC)		w_q (AIC)	
	Older	Younger	Older	Younger	Older	Younger
General multinomial	12,868	13,321	841	909	<.001	<.001
Unconstrained HB13	12,577	12,988	550	577	<.001	<.001
Simplified HB13	12,059	12,438	32	26	<.001	<.001
Logistic HB13	12,027	12,411	0	0	>.99	>.99

AIC = Akaike information criterion; $\Delta_q(\text{AIC}) = [\text{AIC}_q - \min(\text{AIC})]$; w_q(AIC) = rounded Akaike weights interpreted as the probability that Model q is the best model given the data set and set of candidate models

inhibition, episodic memory, and working memory capacity, respectively. The logistic HB13 model has $S \cdot n + (3 \cdot (C + 1)) + z$ parameters, where S is the number of unconstrained HB13 parameters, C is the number of cognitive variables in each bias function, the 1 accounts for the constant terms in each bias function, and z is the number of parameters that were simplified to a constant. Thus, the number of parameters in the younger-adult logistic HB13 model was $2 \cdot 62 + (3 \cdot (3 + 1)) + 8 = 144$.

Model selection for the younger-adult sample We compared our four candidate models following the same rationale as for the older-adult data set. The AIC values and Akaike weights for each of the models are summarized in Table 4. Again, the model associated with the smallest AIC value and an Akaike weight close to 1 was the logistic HB13 model. We also replicated the decrease in AIC values from the most complex model (i.e., the general multinomial model) to the least complex model (i.e., the logistic HB13 model) for the younger-adult sample. Thus, the logistic HB13 model provided the best balance between model fit and parsimony for both the older and younger adults.

Table 5 Model tests for the logistic HB13 model by age group

Cognitive Function	Coefficient		p_b (two-tailed)	
	Older	Younger	Older	Younger
Recollection: Control (r_{Ck})				
α_{C0}	–1.39	–0.47		
β_{C1} (Inhibition)	–0.01	–0.01	.32	.12
β_{C2} (Episodic memory)	0.06	0.01	<.001	.81
β_{C3} (Working memory)	0.03	0.001	.33	.84
Recollection: Experimental (r_{Ek})				
α_{E0}	–1.34	–0.77		
β_{E1} (Inhibition)	–0.02	0.002	<.001	.80
β_{E2} (Episodic memory)	0.07	0.001	<.001	.96
β_{E3} (Working memory)	0.06	0.04	.06	.29
Reconstruction Bias (b_k)				
α_{b0}	–1.49	0.54		
β_{b1} (Inhibition)	0.04	–0.02	<.001	.22
β_{b2} (Episodic memory)	–0.004	–0.06	.96	.27
β_{b3} (Working memory)	0.08	–0.04	.24	.62

p values (denoted p_b) for the cognitive covariate tests are based on bootstrapping the 95 % confidence intervals for the beta coefficients on each cognitive covariate for each of the core parameters

Hypothesis testing In line with these model selection results, we tested our primary hypotheses regarding the role of cognitive functioning in the underlying HB processes using the logistic HB13 model. As is shown in Table 5, on the basis of 500 bootstrapped samples, the 95% confidence intervals for the beta coefficients revealed that none of the cognitive covariates significantly predicted parameters b_k, r_{Ck}, or r_{Ek} in the younger adults. Nonetheless, the pattern of effects of the cognitive covariates on the HB parameters was similar to that observed in older adults. In fact, the directions of all but one of the six beta coefficients required to predict recollection of the control (r_C) and experimental (r_E) items were the same across age groups. The picture is a bit more confusing for reconstruction bias b, where only the direction of the beta coefficient for episodic memory was the same across age groups. Most likely, the effects of inhibition and working memory capacity on b were very weak, or even absent, in younger adults, so that sampling error caused deviations of the regression coefficient estimates in the unexpected directions. The lack of statistical significance of all coefficients for younger adults is consistent with this explanation.

Discussion

We have proposed a novel logistic HB13 multinomial model to assess the cognitive abilities that contribute to the underlying HB processes in older and younger adults. Our model selection analysis indicated that, given our data and a set of four candidate models (see Table 4), the logistic HB13 model was clearly the best approximation to the true model underlying the data for both the older- and younger-adult groups, and it was the model that provided the best balance between model accuracy and parsimony. Consistent with our predictions, our findings revealed that in older adults, (1) individuals with better episodic memory have better recollection of their OJ in the absence of outcome knowledge, (2) individuals with better episodic memory and inhibitory control and higher working memory capacity have better recollection of their OJ in the presence of outcome knowledge, and (3) conditional on a failure to recall their OJ, individuals with better inhibitory control are less likely to be influenced by outcome knowledge when reconstructing their forgotten OJs (i.e., reconstruction bias). In younger adults, descriptively, most of the regression coefficients for the cognitive covariates were in the same direction as those of older adults, but none of the effects attained statistical significance.

Ours is the first study to model interindividual variation in the underlying processes that contribute to HB. This is an important advancement over prior applications of the HB13 multinomial model, which estimated a fixed set of 13 parameters for the entire sample (Erdfelder & Buchner, 1998). The approach presented in this article is a variation of Klauer's (2010) latent-trait model that can be easily implemented and adapted to any experimental paradigm. The primary difference between Klauer's model and our logistic model is that the former requires the assumption of joint distribution of the model parameters and covariates, whereas the latter only requires a distributional assumption about the error in predicting the model parameters from the covariates. Furthermore, in our logistic model we assume that the error term associated with the model parameters is logistically rather than normally distributed. Subsequently, we arrive at a model in which the processes (e.g., reconstruction bias, b_k) are modeled as a logistic function of covariates rather than an ogive function. In Appendix 1, we demonstrate that replacing logistically distributed errors with normally distributed errors does not change the substantive conclusions. The other difference is in the estimation method: Whereas Klauer used a Bayesian approach, we use a maximum likelihood approach. The latter approach requires fewer distributional assumptions and less computational effort, and is easier to implement.

Given the large interindividual variation in cognitive functioning in older adults, our primary interest was in modeling individual differences in the underlying HB processes in an older-adult population; however, we also included a younger-adult comparison group to assess whether the findings would generalize to this age group. For both age groups, we followed a series of model steps to arrive at our final logistic HB13 model. One of our model steps involved examining whether there was significant heterogeneity in the HB13 parameters. Given that we would expect older adults to have the largest variability in their underlying processes, we performed the heterogeneity tests on this age group. We then demonstrated that the resulting model was also acceptable for our younger-adult group. The results of the heterogeneity tests suggest that ancillary parameters g_{l1k} and g_{g1k} might be conceived of as heterogeneous. These parameters represent probabilities of reconstructing ROJs that deviate from the CJ in the direction of the OJ. Although this "partial OJ memory" does not suffice for a perfect recollection (ROJ = OJ), it allows for a reconstructed ROJ that is "close" to the OJ. Our findings thus indicate that participants differ in their ability to generate good, unbiased reconstructions that are close approximations of the OJ. With regard to the core parameters, heterogeneity tests revealed that parameters b_k, r_{Ck}, and r_{Ek} were heterogeneous, whereas parameter c_k was homogeneous. Given that reconstruction bias is one of the primary contributors to HB, we expected individual variation in the frequency with which individuals shift their ROJ toward the CJ. Our finding that participants differed in their ability to recollect the OJ was also expected, given the large variation in recall ability in older adults (e.g., Christensen et al., 1994; Riddle, 2007). Furthermore, the finding of homogeneity in CJ adoptions was not surprising, given prior reports of a low probability of CJ adoptions in adults (e.g., Bayen et al., 2006; Bernstein et al., 2011).

In our final model step, we used the logistic HB13 model to test whether variability in cognitive functioning explained individual variation in the core HB processes in older and younger adults. In older adults, as expected, better episodic memory predicted better OJ recall in the absence of outcome knowledge (r_{Ck}). Moreover, better episodic memory and inhibitory control each predicted better OJ recall in the presence of outcome knowledge (r_{Ek}). It is also noteworthy that working memory capacity was a marginally significant predictor of recollection rates in the presence ($p_b = .056$), but not the absence ($p_b = .33$), of outcome knowledge. The significant contribution of inhibition and the marginally significant contribution of working memory capacity to r_{Ek} but not to r_{Ck} suggest that the ability to recall a prior judgment in the presence of new knowledge depends on one's ability to suppress irrelevant information (i.e., CJ), and potentially depends on one's ability to discriminate multiple pieces of information in mind during recall. Thus, older adults with poorer inhibitory control and lower working memory capacity may be more susceptible to recollection bias.

Our finding that inhibition contributes to recollection bias in older adults is consistent with Erdfelder et al.'s (2007) finding and with the hypothesis that inhibition would be needed to suppress outcome knowledge so that it does not interfere with recall of the OJ (e.g., Bayen et al., 2007; Groß & Bayen, in press). One mechanism may be that older adults with poor inhibition are unable to suppress the CJ, and consequently this information alters their memory representation of the OJ. This would be consistent with adaptive learning or immediate outcome assimilation theories, which postulate that HB results from outcome knowledge automatically updating one's knowledge base (e.g., Fischhoff, 1975; Hoffrage et al., 2000). An alternative possibility is that the CJ interferes with rather than changes one's memory representation of the OJ. This explanation is consistent with the relative-trace-strength hypothesis proposed by Hell et al. (1988), which posits that weaker memory trace strength of the OJ predicts a larger recollection bias. Although the present study implicates inhibition in recollection bias in older adults, further research will be needed to tease apart the precise role that inhibition plays in this process.

Our finding that working memory capacity was a marginally significant contributor to recollection bias in older adults is consistent with the finding from Bayen and colleagues' (2006, Exp. 2) study, which demonstrated that older adults exhibited a significant recollection bias when they were shown the CJs immediately prior to the ROJ stage and were instructed to remember them for a later memory test. Presumably, the instruction to encode the CJs taxed participants' working memory, which decreased overall working memory capacity and increased recollection bias. Calvillo (2012) also found evidence for poorer working memory capacity contributing to a greater overall magnitude of HB (but see also Nestler et al., 2008). However, because the effect of working memory capacity fell just short of statistical significance (p_b = .056) in the present study, future studies should clarify the role of working memory capacity in older adults' recollection biases.

With regard to reconstruction bias, our findings revealed that better inhibitory control predicted less reconstruction bias in older adults, suggesting that the accuracy of reconstructive processes depends on the ability to suppress outcome knowledge that might otherwise influence the reconstruction process. This finding is compatible with Bayen and colleagues' (2006, 2007) inhibitory-deficit explanation of increased susceptibility to HB in older adults. Bayen and colleagues reviewed the inhibitory-deficit theory of cognitive aging (Hasher & Zacks, 1988), which posits that older adults have difficulty inhibiting task-irrelevant information. As a result, this information gains access to working memory and, if not suppressed, interferes with task-relevant performance. Applying this to the memory judgment paradigm, Bayen and colleagues proposed that older adults are more susceptible to HB because they are unable to suppress the CJ (task-irrelevant information) once it gains access to working memory, resulting in this information biasing the reconstruction process. Thus, we have provided further evidence for the role of inhibition in the reconstruction stage of HB and demonstrated that inhibition is more important in this stage than both episodic memory and working memory capacity. Neither of the latter abilities contributed significantly to explaining individual variability in reconstruction bias.

In younger adults, with the exception of the effects of inhibition on the r_{Ek} and b_k parameters and the effect of working memory capacity on the b_k parameter, the pattern of effects of the cognitive covariates on the HB parameters was similar to that observed in older adults; however, none of these effects reached statistical significance. Although the direction of the effects was often in agreement across age groups, the question remains why we obtained strong effects in older adults and weaker effects in younger adults. We propose two possible answers to this question: (1) There is little variability in HB in younger adults, and/or (2) there is little variability in cognitive functioning in younger adults. To test the first possibility, we conducted post-hoc heterogeneity tests on each of the core HB parameters in younger adults. On the basis of 100 bootstrapped samples, our findings revealed that parameter r_{Ck} [$G^2(61)$ = 115.69, p_b < .001], parameter r_{Ek} [$G^2(61)$ = 136.85, p_b < .001], and parameter b_k [$G^2(61)$ = 83.99, p_b < .001] were heterogeneous. Thus, these findings do not support the first possibility. To test for the second possibility, we conducted F tests to assess the equality of variances of the cognitive covariates in older and younger adults. The variances in older adults were about twice as large as those in younger adults. Subsequently, we rejected the null hypothesis that the variances were equal across age groups for inhibition [$F(59, 61) = 2.86$, $p < .001$], episodic memory [$F(59, 61) = 1.87$, $p = .01$], and working memory [$F(59, 61) = 1.55$, $p = .046$]. Thus, we believe that the weaker effects in our younger-adult logistic HB13 model were most likely due to younger adults having less variability in cognitive functioning, as compared to older adults. Indeed, our younger-adult sample was composed of a high-functioning sample of university students. Future research should assess the cognitive mechanisms underlying the HB processes in younger adults while using a more representative sample of the general population of younger adults, which would likely increase variation in cognitive functioning. Another possibility for future research will be to investigate other individual-difference variables (e.g., additional cognitive variables or personality factors) that might explain heterogeneity in younger-adult HB processes that were not addressed here.

Taken together, the present findings advance our understanding of HB by identifying why older adults are more prone to it. We revealed that older adults with poorer inhibitory control, and potentially those with lower working memory capacity, are more likely to forget their original predictions in the presence of outcome knowledge. Perhaps poorer ability to suppress the CJ results in this information interfering with retrieval of the OJ. Subsequently, these individuals must reconstruct their forgotten original predictions. During this reconstruction process, older adults with poorer inhibitory control are more likely to rely on outcome knowledge to guide this reconstructive process. Thus, older adults with poor inhibitory control may be especially vulnerable to HB. Given the abundance of complex and potentially risky financial, health, housing, and safety decisions encountered in later life, combined with the recent social trend to maintain independence in decision-making as we age (Peters et al., 2007), it will be imperative to determine the extent to which proneness to HB impacts real-world decisions.

In summary, we investigated age differences in HB and extended the MPT model of HB to identify the relationship between cognitive functioning and the underlying HB processes. The advantages of our logistic model are that it is easily implemented and can incorporate individual differences in the underlying processes without ascribing to any distributional assumptions of the model parameters. In additional analyses, we demonstrate in Appendix 1 that our findings hold when we assume that the error terms in the underlying latent variables are normally distributed rather than logistically distributed. Future research will be needed to elucidate the mechanisms underlying individual variation in HB processes in younger adults.

Author note This research was supported by the Social Sciences and Humanities Research Council (SSHRC), Canada Research Chairs, the Natural Science and Engineering Research Council (NSERC), and the German Research Foundation (GRF, Er 224/2-2)..

Appendix 1: Additional analyses using a normally distributed error term

To test whether our results would hold when we assumed that the error terms of the underlying latent variables (e.g., b^*_{km}) were normally rather than logistically distributed, we reconducted our primary analyses with older adults and younger adults using an ogive link function. We modeled the core parameters (b_k, r_{Ck}, r_{Ek}) as the following ogive functions of the cognitive covariates:

$$\begin{aligned} b_k &= \Phi(\alpha_{\mathrm{b}} + \beta_{\mathrm{b1}}X1_k + \beta_{\mathrm{b2}}X2_k + \ldots + \beta_{\mathrm{bC}}XC_k) \\ r_{\mathrm{C}k} &= \Phi(\alpha_{\mathrm{con}} + \beta_{\mathrm{con1}}X1_k + \beta_{\mathrm{con2}}X2_k + \ldots + \beta_{\mathrm{conC}}XC_k), \quad \text{(A1)} \\ r_{\mathrm{E}k} &= \Phi\left(\alpha_{\mathrm{exp}} + \beta_{\mathrm{exp1}}X1_k + \beta_{\mathrm{exp2}}X2_k + \ldots + \beta_{\mathrm{expC}}XC_k\right) \end{aligned}$$

where α and β_c, $c = 1, \ldots, C$, are the parameters to be estimated, Xc_k, $c = 1, \ldots, C$, denotes the value of participant k on the variable Xc, and $\Phi(.)$ represents the cumulative distribution function of the standard normal distribution. We thus refer to this model as the *ogive HB13 model* with core parameters b_k, $r_{\mathrm{C}k}$, $r_{\mathrm{E}k}$.

Model evaluation

Our model evaluation comparing the AIC values and Akaike weights for the ogive HB13 model against the remaining three candidate models (cf. Table 4 above) revealed that the ogive HB13 model was the preferable model in both age groups. In older adults, the rounded AIC value for the ogive HB13 model was only slightly worse than the AIC value for the logistic HB13 model (AIC logistic HB13 = 12,027; AIC ogive HB13 = 12,029). In younger adults, the rounded AIC values for the ogive HB13 model and the logistic HB13 model were the same (AIC = 12,412). For both age groups, the probability estimate that the ogive HB13 model was the best of our four candidate models was very close to 1. Thus, the ogive model and the logistic model provide equally good approximations to our data, and both outperform the other three candidate models.

Older-adult ogive HB13 model tests

To test whether any of the cognitive variables independently predicted recollection ability or reconstruction bias in older adults, we bootstrapped the 95% confidence intervals for the beta coefficients on each cognitive covariate for each of the core parameters (r_{Ck}, r_{Ek}, and b_k) in the ogive HB13 model. As is shown in Table 6, the findings from our logistic HB13 model held: Episodic memory was the only significant predictor of r_{Ck}, $\beta = 0.03$, 95% CI = [0.01, 0.06], $p_{\mathrm{b}} < .001$; both episodic memory, $\beta = 0.04$, 95% CI = [0.02, 0.06], $p_{\mathrm{b}} < .001$, and inhibition, $\beta = -0.01$, 95% CI = [–0.02, –0.01], $p_{\mathrm{b}} < .001$, were significant predictors of r_{Ek}; working memory capacity was a marginally significant predictor of r_{Ek}, $\beta = 0.04$, 95% CI = [–0.001, 0.07], $p_{\mathrm{b}} = .10$; and inhibition was the only significant predictor of b_k, $\beta = 0.03$, 95% CI = [0.01, 0.04], $p_{\mathrm{b}} < .001$.

Younger-adult ogive HB13 model tests

To test whether any of the cognitive variables independently predicted recollection ability or reconstruction bias in younger adults, we bootstrapped the 95% confidence intervals for the beta coefficients on each cognitive covariate for each of the core parameters (r_{Ck}, r_{Ek}, and b_k) in the ogive HB13 model. As is shown in Table 6, we once again replicated the findings from our logistic HB13 model: On the basis of 500

Table 6 Model tests for the ogive HB13 model by age group

Cognitive Function	Coefficient		p_{b} (two-tailed)	
	Older	Younger	Older	Younger
Recollection: Control (r_{Ck})				
α_{C0}	–0.86	–0.29		
β_{C1} (Inhibition)	–0.003	0.01	.38	.84
β_{C2} (Episodic memory)	0.03	–0.01	<.001	.35
β_{C3} (Working memory)	0.02	0.004	.32	.71
Recollection: Experimental (r_{Ek})				
α_{E0}	–0.83	–0.48		
β_{E1} (Inhibition)	–0.01	0.02	<.001	.88
β_{E2} (Episodic memory)	0.04	0.001	<.001	.89
β_{E3} (Working memory)	0.04	0.001	.10	.50
Reconstruction Bias (b_k)				
α_{b0}	–0.92	0.31		
β_{b1} (Inhibition)	0.03	–0.03	<.001	.21
β_{b2} (Episodic memory)	–0.0002	–0.01	.92	.24
β_{b3} (Working memory)	0.05	–0.04	.38	.56

p values (denoted p_{b}) for the cognitive covariate tests are based on bootstrapping the 95 % confidence intervals for the beta coefficients on each cognitive covariate for each of the core parameters bootstrapped samples, the 95% confidence intervals for the beta coefficients revealed that none of the cognitive covariates significantly predicted b_k, r_{Ck}, or r_{Ek}. Taken together, these results suggest that replacing logistically distributed errors with normally distributed errors does not change the substantive conclusions.

Appendix 2: Aggregate analysis using the HB13 model

We analyzed the data using the HB13 multinomial model developed by Erdfelder and Buchner (1998). To assess model fit, we compared the aggregate HB13 model against the general multinomial model for 2 · 10 data categories using parametric bootstrapping. In accordance with prior work (e.g., Bayen et al., 2006), we set an alpha level of .01 for the model fit test because we did not want to reject a model that only slightly differed from the comparison model. The model evaluation was based on 6,040 data points (122 participants · 50 items – 60 missing responses). Power analysis indicated that we had high power (.99) to detect even small deviations (w = .1; Cohen, 1988) from the general multinomial model. On the basis of 500 bootstrapped samples, we found an acceptable model fit, $G^2(10) = 31.27$, $p_b = .01$.

Table 7 presents the results of the parameter tests across age groups, and Table 8 reports the parameter estimates and their standard errors by age groups. We used the conventional alpha level of .05 to test for statistical differences in the parameters. As expected, on the basis of 500 bootstrapped samples, our analysis revealed significant age differences in overall recollection ability—that is, parameters r_C and r_E. In comparison to younger adults, older adults had poorer recollection of the OJs

Table 7 Parameter tests using the aggregate HB13 model

Null Hypothesis	ΔG^2	p_b	df
Across-Age-Group Tests			
r_C constant across age groups	45.31	<.001	1
r_E constant across age groups	59.71	<.001	1
b constant across age groups	1.37	.23	1
c constant across age groups	5.04	<.001	1
Within-Age-Group Tests: Younger Adults			
$r_C = r_E$	2.77	.11	1
$b = 0$	56.24	<.001	4
Within-Age-Group Tests: Older Adults			
$r_C = r_E$	6.68	.001	1
$b = 0$	76.08	<.001	4

p values (denoted p_b) are based on 500 bootstrapped samples. r_C = probability of recalling the original judgment (OJ) for control items; r_E= probability of recalling the OJ for experimental items; b = probability of reconstruction bias

Table 8 HB13 model parameter estimates (and standard errors) by age group

Parameter	Parameter Estimate (Standard Error)	
	Older	Younger
r_C	.25 (.01)	.36 (.01)
r_E	.21 (.01)	.33 (.01)
$r_C - r_E$	.04	.03
b	.42 (.04)	.35 (.04)
c	.01 (.01)	<.001 (.01)

for both control and experimental items. Furthermore, older but not younger adults demonstrated a significant recollection bias, defined as poorer recollection of the OJ for experimental as compared to control items. Descriptively, the recollection bias was slightly larger in older adults (.04) than in younger adults (.03). Bayen et al. (2006) also observed recollection biases for both younger and older adults, but the bias was not statistically significant for younger adults and fell short of statistical significance for older adults. Perhaps our significant recollection bias finding in older adults was due to our relatively large sample of 62 younger and 60 older adults, which was more than double the sample of Bayen et al. (2006; 26 younger and 26 older adults).

Older adults demonstrated a larger reconstruction bias than did younger adults; however, this difference did not reach statistical significance (p_b = .23). Although Bayen et al. (2006) found a significant age difference in reconstruction bias, this difference did not reach statistical significance in Bernstein et al. (2011, p = .18). Nevertheless, in all three studies, there was at least a descriptive trend toward older adults demonstrating larger reconstruction bias than did younger adults. Finally, we observed a significant age difference in parameter c, with older adults demonstrating significantly more CJ adoptions than younger adults. This is consistent with Bayen et al.'s (2006, Exp. 2) finding of increased CJ adoptions when the CJ is accessible during ROJ. Overall, our findings of age differences in the core HB parameters are generally consistent with those of prior work.

Appendix 3: List of questions (and correct answers in metric units) used in the memory judgment task

1. At what temperature does copper melt? (2,415 Celsius)
2. How high is the Statue of Liberty including its base? (93 meters)
3. What year did the mutiny on the Bounty occur? (1790)
4. What is the distance between New York and Los Angeles (by road)? (4,546 kilometers)
5. In what year was the monkey wrench invented? (1841)

6. In what year was the harmonica invented? (1821)
7. How long is the Rhine River? (1,320 kilometers)
8. What year did the Hundred Years' War begin? (1339)
9. What year was the lightning rod invented? (1752)
10. How long is the Great Wall of China? (3,460 kilometers)
11. What year were X-rays discovered? (1895)
12. At what speed must wind blow to be classified as a Moderate Gale Force? (51 kilometers per hour)
13. What is the average depth of the Pacific Ocean? (3,940 meters)

*14. At what temperature does tin melt? (2,930 Celsius)

15. On average, how many days is a female elephant pregnancy? (631 days)
16. How long is the Amazon River? (6,556 kilometers)
17. How long is the Mississippi River? (3,779 kilometers)
18. What year did William Herschel discover the planet Uranus? (1781)
19. In what year was Jane Austin's *Pride and Prejudice* first published? (1813)
20. What is the average temperature of the Antarctic winter? (-68 Celsius)
21. What is the highest temperature ever measured on Earth? (57 Celsius)
22. What percentage of the world's population was under the age of five in 1995? (7.7%)
23. What year was Leonardo da Vinci born? (1452)
24. How long is the world's longest bridge? (38.42 kilometers)
25. What year did Sir James Dewar, an English chemist, invent the thermos flask? (1873)
26. When was the first reflecting telescope developed? (1671)
27. How many carats is the world's largest reported diamond? (3,106 carats)

*28. What is the official land speed record for a land vehicle? (1,019 kilometers per hour)

29. How many days does the planet Mercury take to make one trip around the sun? (88 days)
30. How long is an international nautical mile? (1,852 meters)
31. What percentage of the world's population lived in Africa in 1994? (12.4%)
32. How many plays did William Shakespeare write? (37 plays)
33. When travelling 97 kilometers per hour in a car, how much room should you allow yourself to brake? (83 meters)
34. What is the distance between Tokyo and Chicago (by air)? (10,137 kilometers)
35. What year was the parking meter invented? (1935)
36. What year was radiotelegraphy invented? (1899)
37. What year did Leonardo da Vinci create *Mona Lisa*? (1503)

*38. In what year was Harvard University founded? (1686)

39. What year did Franz Joseph I, the emperor of Austria, die? (1916)
40. What year did Albert Einstein formulate the theory of relativity? (1903)
41. What is the diameter of the planet Mars? (6,787 kilometers)
42. How high is the highest point on Mount Kilimanjaro? (5,895 meters)
43. What year were the first modern-day Olympic games celebrated? (1896)
44. What percentage of the world's population lived in Europe in 1994? (9%)
45. How many muscles does the human body have? (639 muscles)
46. What percentage of the human body is composed of nitrogen? (8.5%)
47. What year was the first mailbox invented? (1653)
48. When was slavery officially abolished in the United States? (1865)
49. How many films did Alfred Hitchcock direct? (56 films)
50. In what year was William Shakespeare's *The Tragedy of King Lear* first published? (1608)
51. In what year was Socrates born? (470 BC)
52. In what year was Daniel Defoe's "Robinson Crusoe" first published? (1719)
53. What year was the mechanical loom invented? (1785)
54. How many detective books did Agatha Christie write? (67 books)

Note: The items marked by an asterisk were excluded from the analyses because they violated the model's symmetry assumption.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, AC-19,* 716–723. doi:10.1109/TAC.1974.1100705

Akaike, H. (1978). A new look at the Bayes procedure. *Biometrika, 65,* 53–59.

Ansari, A., Vanhuele, M., & Zemborain, M. (2008). *Heterogeneous multinomial processing tree models*. Unpublished manuscript. Columbia University, New York.

Arkes, H. R., Wortman, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology, 66,* 252–254.

Ashby, F. G., Prinzmetal, W., Ivry, R., & Maddox, W. T. (1996). A formal theory of feature binding in object perception. *Psychological Review, 103,* 165–192. doi:10.1037/0033-295X.103.1.165

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

Bayen, U. J., Erdfelder, E., Bearden, J. N., & Lozito, J. P. (2006). The interplay of memory and judgment processes in effects of aging on hindsight bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 1003–1018. doi:10.1037/0278-7393.32.5.1003

Bayen, U. J., Pohl, R. F., Erdfelder, E., & Auer, T.-S. (2007). Hindsight bias across the lifespan. *Social Cognition, 25,* 83–97.

Bernstein, D. M., Erdfelder, E., Meltzoff, A. N., Perria, W., & Loftus, G. R. (2011). Hindsight bias from 3 to 95 years of age. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 378–391.

Blank, H., & Nestler, S. (2007). Cognitive process models of hindsight bias. *Social Cognition, 25,* 132–146.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. New York, NY: Springer.

Calvillo, D. P. (2012). Working memory and the memory distortion component of hindsight bias. *Memory, 20,* 891–898. doi:10.1080/09658211.2012.706309

Christensen, H., Mackinnon, A., Jorm, A. F., Henderson, A. S., Scott, L. R., & Korten, A. E. (1994). Age differences and interindividual variation in cognition in community-dwelling elderly. *Psychology and Aging, 9,* 381–390.

Christensen-Szalanski, J. J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes, 48,* 147–168.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coolin, A., Bernstein, D. M., Thornton, A. E., & Thornton, W. L. (2014). Age differences in hindsight bias: The role of episodic memory and inhibition. *Experimental Aging Research, 40,* 357–374.

Dehn, D. M., & Erdfelder, E. (1998). What kind of bias is hindsight bias? *Psychological Research, 61,* 135–146.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan executive function system*. San Antonio, TX: Psychological Corp.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *California Verbal Learning Test–2nd ed. (CVLT-II)*. San Antonio, TX: Psychological Corp.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test 3rd ed. (PPVT-III)*. Circle Pines, MN: American Guidance Service.

Edgington, E. S. (1995). *Randomization tests* (3rd ed.). New York, NY: Marcel Dekker.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Journal of Psychology, 217,* 108–124.

Erdfelder, E., Brandt, M., & Bröder, A. (2007). Recollection biases in hindsight judgments. *Social Cognition, 25,* 114–131.

Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 387–414. doi:10.1037/0278-7393.24.2.387

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,* 175–191. doi:10.3758/BF03193146

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288–299.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini Mental State": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12,* 189–198. doi:10.1016/0022-3956(75)90026-6

Groß, J., & Bayen, U. J. (in press). Hindsight bias in younger and older adults: The role of access control. *Neuropsychology, Development, and Cognition.*

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology, 26,* 103–117.

Hardt, O., & Pohl, R. F. (2003). Hindsight bias as a function of anchor distance and anchor plausibility. *Memory, 11,* 397–394.

Harley, E. M. (2007). Hindsight bias in legal decision making. *Social Cognition, 25,* 48–63.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 22, pp. 193–225). San Diego, CA: Academic.

Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin, 107,* 311–327. doi:10.1037/0033-2909.107.3.311

Hedden, T., & Gabrieli, J. D. E. (2004). Insights into the ageing mind: A view from cognitive neuroscience. *Natures Review Neuroscience, 5,* 87–96.

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition, 16,* 533–538. doi:10.3758/BF03197054

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 566–581.

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

Klauer, K. C. (2006). Hierarchical multinomial processing tree models: A latent-class approach. *Psychometrika, 71,* 7–31.

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

Klauer, K. C., Stahl, C., & Erdfelder, E. (2007). The abstract selection task: New data and an almost comprehensive model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 680–703.

Leary, M. R. (1981). The distorted nature of hindsight. *Journal of Social Psychology, 115,* 25–29.

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view". In C. M. MacLeod & D. S. Gorfein (Eds.), *Inhibition in cognition* (pp. 145–162). Washington, DC: American Psychological Association.

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against the memory impairment hypothesis. *Journal of Experimental Psychology: General, 114,* 1–16. doi:10.1037/0096-3445.114.1.1

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44,* 190–204. doi:10.1006/jmps.1999.1283

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4,* 79–95. doi:10.3758/BF03210778

Nestler, S., Blank, H., & Egloff, B. (2010). Hindsight ≠ hindsight: Experimentally induced dissociation between hindsight components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1399–1413.

Nestler, S., Blank, H., & von Collani, G. (2008). Hindsight bias doesn't always come easy: Causal models, cognitive effort, and creeping determinism. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1043–1054.

Newton, J. M., & Wickens, D. D. (1956). Retroactive inhibition as a function of the temporal position of interpolated learning. *Journal of Experimental Psychology, 51,* 149–154.

Peters, E., Hess, T. M., Vastfjall, D., & Auman, C. (2007). Adult age differences in dual information processes: Implications for the role of affective and deliberative processes in older adults' decision making. *Perspectives on Psychological Sciences, 2,* 1–23.

Pezzo, M. V., & Pezzo, S. P. (2007). Making sense of failure: A motivated model of hindsight bias. *Social Cognition, 25,* 147–164.

Pohl, R. F. (2004). Hindsight bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement, and memory* (pp. 363–378). Hove, UK: Psychology Press.

Pohl, R. F. (2007). Ways to assess hindsight bias. *Social Cognition, 25,* 14–31.

Pohl, R. F., Bayen, U. J., & Martin, C. (2010). A multiprocess account of hindsight bias in children. *Developmental Psychology, 46,* 1268–1282.

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory, 11,* 337–356.

Raz, N., Ghisletta, P., Rodrigue, K. M., Kennedy, K., & Lindenberger, U. (2010). Trajectories of brain aging in middle-aged and older adults: Regional and individual differences. *NeuroImage, 51,* 501–511.

Riddle, D. R. (2007). *Brain aging: Models, methods, and mechanisms*. Boca Raton, FL: CRC Press.

Salthouse, T. A. (2000). Aging and measures of processing speed. *Biological Psychology, 54,* 35–54.

Schwarz, S., & Stahlberg, D. (2003). Strength of hindsight bias as a consequence of metacognitions. *Memory, 11,* 395–410.

Smith, J. B., & Batchelder, W. H. (2008). Assessing individual differences in categorical data. *Psychonomic Bulletin & Review, 15,* 713–731. doi:10.3758/PBR.15.4.713

Smith, J. B., & Batchelder, W. H. (2010). Beta-MPT: Multinomial processing tree models for addressing individual differences. *Journal of Mathematical Psychology, 54,* 167–183.

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

Stahlberg, D., & Maass, A. (1998). Hindsight bias: Impaired memory or biased reconstruction? In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 8, pp. 105–132). London, UK: Wiley.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11,* 192–196. doi:10.3758/BF03206482

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale–3rd ed. (WAIS-III)*. New York, NY: Psychological Corp.

Winman, A., Juslin, P., & Björkman, M. (1998). The confidence–hindsight mirror effect in judgment: An accuracy-assessment model for the knew-it-all-along phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 415–431. doi:10.1037/0278-7393.24.2.415

Zelazo, P. D., Craik, F. I. M., & Booth, L. (2004). Executive function across the life span. *Acta Psychologica, 115,* 167–183. doi:10.1016/j.actpsy.2003.12.005