Dynamic dual process account explaining the bias after outcome – An exploratory research on memory distortion hindsight bias

Liang Sheng

Dissertation thesis written at the Center for Doctoral Studies in the Social and Behavioral Sciences of the Graduate School of Economic and Social Sciences and submitted for the degree of Doctor of Philosophy (Ph.D.) of the Faculty of Social Sciences at the University of Mannheim

First supervisor: Prof. Dr. Rüdiger Pohl

Second supervisor: Prof. Dr. Edgar Erdfelder

Academic Director of the Center for Doctoral Studies in the Social and Behavioral Sciences:

Prof. Dr. Thomas Bräuninger

Disputation Date: 02.12.2015

Acknowledgments

I would like to express my deep gratitude to my supervisors Prof. Dr. Rüdiger Pohl and Prof. Dr. Edgar Erdfelder for offering me the precious opportunity to pursue my PhD thesis and for their great patience during the writing-up of my dissertation. I must offer special appreciation to Prof. Dr. Rüdiger Pohl for his patient guidance and encouragement throughout my PhD study and for his constructive suggestions and critique of this research work.

My sincere thanks go to Dr. Sandra Mattern and Dr. Janina Hoffmann for their help in conducting the statistical data analyses and for their valuable advice. Gratitude is also extended to Dr. Morten Moshagen, Dr. Andre Aßfalg, Dr. Lena Nadarevic, Dr. Carolina Küpper-Tetzel for their help and support of my study. I would also like to thank Judith Flöck for her assistance in programing and running the experiments, recruiting participants and collecting data.

I would also like to thank the department and the staff of Experimental and Differential Psychology at the University of Mannheim for offering me support in resources, technicians and laboratory rooms for my experiments.

For funding my PhD study, I would like to acknowledge the help provided by the following institutions and organizations

- Center for Doctoral Studies in Social and Behavioral Sciences (CDSS) of the Graduate school of Economic and Social Sciences (GESS) at the University of Mannheim for the scholarship funding of my PhD study.

- Deutsche Forschungsgemeinschaft (DFG) for funding for my whole research.

Finally, I wish to thank all those who took part in my experiments. They provided indispensable material for the current research.

Lastly, very special gratitude goes to my family – my parents, my husband and my two children – for their support and encouragement.

One important person I would like to thank is my friend Josh Dahlberg, who spent considerable time in helping me with the English language of my dissertation.

Abstract

When people attempt to recall their pre-feedback estimation of a general knowledge question, they are often biased toward the feedback. This is portrayed as hindsight bias (HB) memory distortion type. The current study is a theory-driven exploratory research that aims to discover the fundamental underlying mechanisms of such bias. Experiment 1 and 2 confirm and verify the appropriate interference task and depletion procedure used for the main experiments. Experiment 3 applies the depletion procedure on the HB memory paradigm and finds a descriptive increasing trend of HB after depletion. Experiment 4 combines the labeling effect (within-subjects manipulation) and depletion procedure on HB and further confirms the result of experiment 3. A single-dissociation pattern of HB magnitude after depletion with different labeling conditions is obtained, indicating a dynamic dual-process mechanism underlying HB, with knowledge-updating behavior and interference resistance action as the two controlled and flexible processes yielding HB. Results of HB are obtained from both traditional inference analyses and the multinomial modeling analyses. Comparisons are discussed demonstrating the superiority of the latter analyzing method applying HB.

Contents

Opening Words

From the moment I sit down to begin writing this dissertation, I am compelled to grapple with one of the most robust judgment-phenomena in Cognitive Psychology – the topic this dissertation focuses on – hindsight bias. I need to be conscious of hindsight bias throughout the writing process to ensure I provide sufficient explanation of concepts to readers who have limited prior knowledge of hindsight bias or Cognitive Psychology. This means not only deciding which material I should include in this dissertation, but also determining the extent to which I introduce and elaborate on specific points, ever conscious that as a writer I risk being “blindfolded” by the knowledge I already possess on subject matter my readers may be less familiar with.

Even being hyper-aware of this robust phenomenon, one finds it difficult to avoid falling into its traps. My aim is to manage the conundrum of hindsight bias, writing up this dissertation with every pertinent explanation while not burdening the reader with unnecessary detail and redundant information.

12 May 2011

1. Introduction

1.1. Hindsight Bias (HB) fundamental introduction

The concept

Hindsight Bias is among the most fascinating topics examined in Cognitive Psychology over the past three decades, since Fischhoff (1975) first conducted a study revealing an over-estimation of what people knew about events before receiving knowledge of the outcome of those events. Two meta-analysis studies of HB have also contributed to the literature in the field (Christensen-Szalanski & Willham, 1991; Guilbault, Bryant, Brockway, & Posavac, 2004). In everyday life, we often find ourselves using phrases such as "*I knew it*" or "*I told you so*" to those around us, whether they be friends, colleagues or partners. However, we typically do so without considering the possibility that we may in fact be experiencing an intriguing social and cognitive phenomenon – hindsight bias. The robustness of this phenomenon has garnered attention from psychologists in all areas from cognitive to economic. The concept may be observed in various eras, cultures and religions; it has been elaborated in Buddhist, Sanskrit and in Tibetan texts. In contemporary Western society, this bias is displayed in a multiplicity of milieu from legal decisions (Kamin & Rachlinski, 1995; Harley, 2007) and medical diagnoses (Arkes, Wortman, Saville, & Harkness, 1981; Arkes, 2013) to decisions in economic contexts (Zwick, Pieter, & Baumgartner, 1995) and in psychological attitudes such as paranormal beliefs (Kane, Core, & Hunt, 2010).

Although what we may colloquially call a "knew it all along" experience has been deemed the same fundamental phenomenon expressed in varied contexts, Blank, Nestler, von Collani, & Fischer (2008) argue that HB is not a unitary concept. They distinguish between three types of HB in the literature, each manifesting itself at different levels. These three types are *impression of necessity* (behavioral level); *impression of foreseeability* (affective level) and *memory distortion* (cognitive level). This core idea of this dissertation deals with the latter type of HB – *memory distortion* – on a cognitive level.

The main aim of this dissertation is to explore the fundamental underlying cognitive processes of HB. In introduction part, an overview to the relating aspects of HB is given, e.g., the plausibility effect of HB, the developmental findings. With respect to the statistical research method, both traditional HB measures and the relatively more recent multinomial processing models elaborated by Erdfelder and Buchner (1998) are represented since both methods are used in this dissertation. Then already existing theories and models of HB are reviewed, following which purpose and hypotheses of the current study are introduced. Then the empirical part of this study is presented with reporting and discussing 4 experiments. General concepts and theories about HB are firstly introduced in the next following sections.

The task and the design

All three types of HB can be tested experimentally in different task designs. Pohl (2007) summarizes the two main schemas employed in the literature: *memory* design and *hypothetical* design.

In the *hypothetical* design, subjects are given a question; they are then asked what they would have given as their estimating answer while simultaneously being presented with the correct answer to the question. The entire procedure is conducted within one experimental session without any retention interval. Conversely, in the *memory* design, subjects are given a question and asked for their first estimating answer; either immediately or after a certain retention interval, the correct answer is presented to them. In the final stage, subjects are required to recall their initial answer. Unlike the hypothetical design, there is a time delay involved; hence memory processes are relevant in the memory design. Subjects are susceptible in both designs to become biased by the outcome answer of the question when giving their own hypothetical answer or when trying to recall their initial estimation. Examples of both designs are demonstrated in Table 1.

Table 1

Examples of Hypothetical Design and Memory Design of HB Task

	Hypothetical design	Memory design
Question	*How high is the Mannheimer Watertower?*	*How high is the Mannheimer Watertower?*
First Estimation	(absent)	82 m
Retention Interval	(absent)	(present or absent)
Correct Answer	(later presented)	60 m
Retention Interval	(absent)	(present or absent)
Test	*The correct answer is 60m.* *What would you have guessed?*	*What was your answer initially?*

In both designs, different materials can be applied. The most commonly used materials for the memory distortion type of HB has been numerical stimuli, which are tested by general knowledge questions (or almanac questions) (e.g., Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988; Hertwig, Gigerenzer, & Hoffrage, 1997; Bayen, Erdfelder, Bearden, & Lozito, 2006). The examples given in Table 1 belong to the category of numerical memory. Another stimulus type for memory design of HB examines 'event outcome probability', using forced choice questions to obtain a subject's level of confidence that they have offered correct answers (Winman, Juslin, & Bjorkman, 1998). For instance, Fischhoff (1977) provided participants with a fact (e.g., that the story of Aladdin originated in Persia rather than China) after initially asking them to choose between two alternatives. Participants showed HB by reporting a higher confident memory about the correct answer than they had had prior to learning about it. In addition, there are also findings of HB in the form of behaviour knowledge and action intentions (Pieters, Baumgartner, & Bagozzi, 2006). Empirical studies employing the memory design have typically been conducted within subjects, whereas those using hypothetical design have been carried out between subjects. Although both designs measure the extent to which the post-outcome judgment is biased toward the feedback, there may be different underlying mechanisms involved. Researchers have also differed in their findings with regard to cultural variance of HB in hypothetical design and memory design (see Choi & Nisbett, 2000; Heine & Lehman, 1996 and Pohl, Bender, & Lachmann, 2002). Thus, it is necessary to differentiate between the two HB designs. Hertwig et al. (1997) contend the designs have discrete terminologies. They argue the term *Hindsight Bias* should only be applied in memory design, whilst *knew-it-all-along* should be used in hypothetical design. I shall adopt this differentiation for the current study; as I use only memory design in this work, all subsequent discussion concerns memory design exclusively, with hypothetical design being put aside.

In more recent studies, HB has been confirmed within the domain of visual perception. Harley, Carlsen, & Loftus (2004) showed evidence of participants overestimating the level of visual degradation at which they or a naïve peer are able to recognize visual targets, demonstrating HB through their own knowledge of target identity (using celebrity faces as stimuli). Similarly, Bernstein, Atance, Loftus & Meltzoff (2004) and Bernstein, Atance, Meltzoff & Loftus (2007) provided empirical evidence among both children (3-5 year olds) and adults that overestimation exists when they are asked to predict their own or a naïve peer's ability to identify a visual object that had been previously presented to them.

Moreover, short text stories and videos have also been used as materials to test the foreseeability component of HB (Fessel, Epstude, & Roese, 2009). This stream of research not only indicates that HB is a very robust phenomenon (e.g., in Harley et al., 2004) that is not effectively reduced by explicit instructions to avoid the bias, but also tells us that HB can be displayed in various modalities rather than simply being an intellectually involved concept. These pronounced characteristics of HB have triggered numerous empirical studies in the last four decades; using different methods of measurement and manipulations, this varied work has sought to uncover the underlying mechanism(s). In the next two sub-sessions, methodological issues pertaining to HB research will be discussed.

How to measure HB?

In the literature of memory HB, 3 methods have been applied to measure the magnitude of HB with continuous responses. Hell et al. (1988) first proposed to quantify HB as a percentage in the form of the Hell-Index: $HB = 100\ (OJ - ROJ)/(OJ - CJ)$, with *OJ* denoting the numerical value of the Original Judgment, *ROJ* denoting the numerical value of Recall of Original Judgment, *CJ* denoting the numerical value of the Correct Judgment (these abbreviations will be used henceforth). Excluding perfectly recalled items, the Hell-Index calculates the percentage of HB with a range of normal HB from 0% to 100% (higher values indicating a larger bias).

A negative value indicates that ROJ is further removed from the CJ than the OJ is. Values larger than 100% indicate that ROJs are on the opposite side of the CJs to the OJs. The advantage of the Hell-Index is that it is not dependent on the absolute values of the unit used, or of the numerical distance between the OJ and CJ. However, there are also cases of HB that the Hell-Index cannot detect. For example, when the mean of OJ in a memory design is the same as the CJ of that item, the mean of ROJ would likely also equal CJ; in this instance, HB would be zero using the Hell-Index computation. However, when looking at individual responses, there may in fact be more ROJs moving toward the CJ, indicating the existence of HB that cannot be detected using Hell-Index and the response's mean of OJ and ROJ. Thus, in this case individual responses must be used instead of the mean response with Hell-Index computation. Another problem with the Hell-Index concerns extreme values. For instance, when OJ is much closer to CJ than the ROJ is (i.e., ROJ is much further away from both OJ and CJ), the HB magnitude would in turn be extremely large. To overcome this problem embedded in Hell-Index calculation, extreme data points must be excluded; alternatively, median (instead of mean) HB values can be used.

A more elegant response to deficiencies in the Hell-Index became available with the introduction of the Proximity Index ($HB_{prox.}$) (Fischer & Budescu, 1995; Pohl, 1992).

$$HB_{prox.} = | OJ\text{-}CJ | - | ROJ\text{-}CJ |$$

The *Proximity Index* is 0 when ROJ is the same as OJ, indicating an absence of HB. A positive *Proximity Index* occurs when ROJ is closer to CJ than OJ is. If ROJ is further away from the CJ than OJ is, then the *Proximity Index* is negative. In empirical studies, standardized values (*z* values) are employed to enable comparison between HB items using different units and between experiments using different HB items (Pohl, 1992, 2007; Pohl & Hell, 1996). Thus,

$$\Delta z = | z_{OJ} - z_{CJ} | - | z_{ROJ} - z_{CJ} |$$

The techniques described above represent the conventional methods of measuring HB. However, they carry with them both methodological and theoretical drawbacks (Pohl, 1995). Methodologically, there is a regression tendency in that the ROJ tends toward the mean of the OJ. If a CJ happens to be close to the mean of OJ, this regression effect can lead to HB. By including a control group that is given the same HB items without CJ presentation, we are able to obtain a comparison with the experimental group (with CJ presentation). In turn, the true magnitude of HB can be derived. Even with this solution applied, conventional methods of measuring HB cannot distinguish whether the ROJ results from direct recall from memory, a reconstruction of the estimation process or even a pure guessing process. In light of this, Erdfelder & Buchner (1998) constructed a multinomial-processing model that separates different processes that may lead to the final ROJ. The next section describes this model.

Erdfelder & Buchner (1998)'s HB-13 model

Besides the general analysis technique, Erdfelder, 2000 and Erdfelder & Buchner (1998) apply the stochastic model analysis for HB data – the multinomial processing tree (MPT) model. As we have seen, traditional descriptive and inference statistics are not sophisticated enough to differentiate the underlying processes of HB, rendering them unable to reveal which specific process is influenced in different situations.

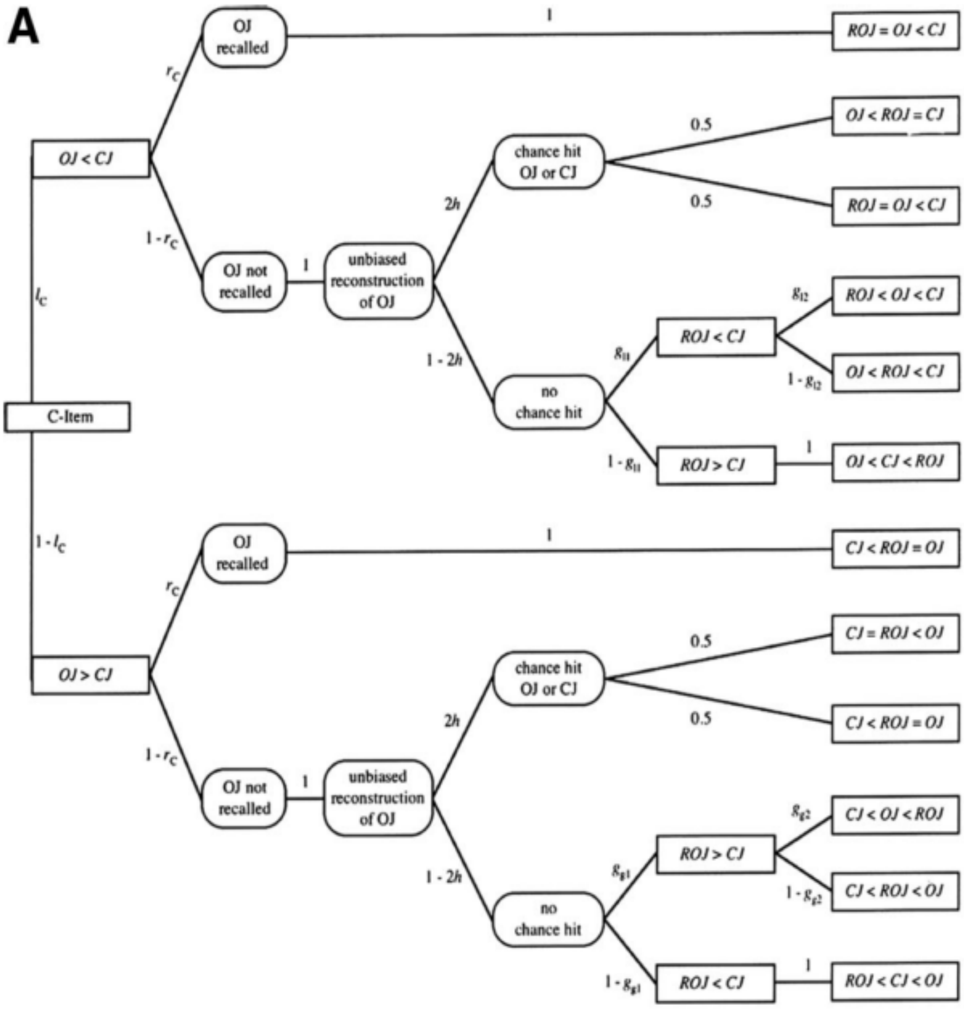
The MPT models provide insight into cognitive processes that are not readily seen via analysis of the observable frequencies of different categorical events. Until the 1980s, MPT models were rather underused. However, since the study undertaken by Riefer & Batchelder

(1988), MPT models have been widely adopted in areas of cognitive psychology concerned with dimensions of cognitive phenomena; for a review, see Erdfelder et al., (2009). However, MPT models only explain categorical data, requiring for example yes/no or correct/false responses. Our HB memory paradigm yields continuous data requiring numerical judgments. Erdfelder & Buchner (1998) solved this dilemma by sorting the HB memory task data into 10 categories according to the magnitude of OJ, CJ and ROJ (for a detailed order of categories see Appendix). Subsequent to this, their HB-13 model has been extensively used in a range of HB empirical studies (e.g., Bayen, Erdfelder, Bearden & Lozito, 2006).

The assumption of the HB-13 model is that participants in a HB task would first try to directly recall their initial judgments. Only when they fail do they then engage in a reconstruction action (Hoffrage et al., 2000). The 10 possible categories of responding orders are based on this reconstruction. For example, if the OJ of the question "How high is the Eiffel Tower?" is "280m", after being presented with the CJ (324m), participants may give a ROJ (320m) that is clearly biased toward the CJ. This response would then fall into the order category of OJ <ROJ<CJ. The frequencies of each order category are then calculated for later analyses, with cases of correct OJs being disregarded. The parameters denoting different underlying processes can then be estimated through the raw categorical data; the estimations consequently indicate the occurrence probability of each process. There are 13 parameters in Erdfelder & Buchner (1998)'s HB MPT model (listed in Table 2). The main HB-13 model parameters that are critical for most typical HB studies are r_C, r_E – denoting the probability of successful recollection in control and experimental conditions, respectively; and parameter b measuring the probability of a biased reconstruction in experimental conditions. It should be noted that the parameter c adopted for the current study based on Pohl, Bayen, & Martin (2010)'s study characterizes a different psychological interpretation to Erdfelder & Buchner (1998)'s classic HB-13 model (explanation of which is given in later tree description).

Table 2
HB-13 Model parameters and their Psychological Interpretations

Parameter	Interpretation
r_C, r_E	Probability of recalling the OJs for control and experimental items, respectively.
b	Probability of a biased reconstruction given a failed recall of the OJs.
c	Probability of adopting the CJs.
h	Probability of a chance hit of the OJs or the CJs in case of unbiased reconstructions.
g_{l1}, g_{g1}, g_{l2}, g_{g2}	Parameters denoting the ROJ distribution in case of unbiased reconstructions without chance hits (for OJ < CJ and OJ > CJ, respectively).
g_{l3}, g_{g3}	Parameters denoting the ROJ distribution in case of biased reconstructions without CJ adoption (for OJ < CJ and OJ > CJ, respectively).
l_C, l_E	Probability of OJ < CJ for control and experimental items, respectively.


A
C-Item
l_C
$1 - l_C$
OJ < CJ
OJ > CJ
r_C
$1 - r_C$
OJ recalled
OJ not recalled
1
unbiased reconstruction of OJ
2h
1 - 2h
chance hit OJ or CJ
no chance hit
0.5
ROJ = OJ < CJ
OJ < ROJ = CJ
ROJ < CJ
ROJ > CJ
ROJ < OJ < CJ
OJ < ROJ < CJ
OJ < CJ < ROJ
CJ < ROJ = OJ
CJ = ROJ < OJ
CJ < OJ < ROJ
CJ < ROJ < OJ
ROJ < CJ < OJ

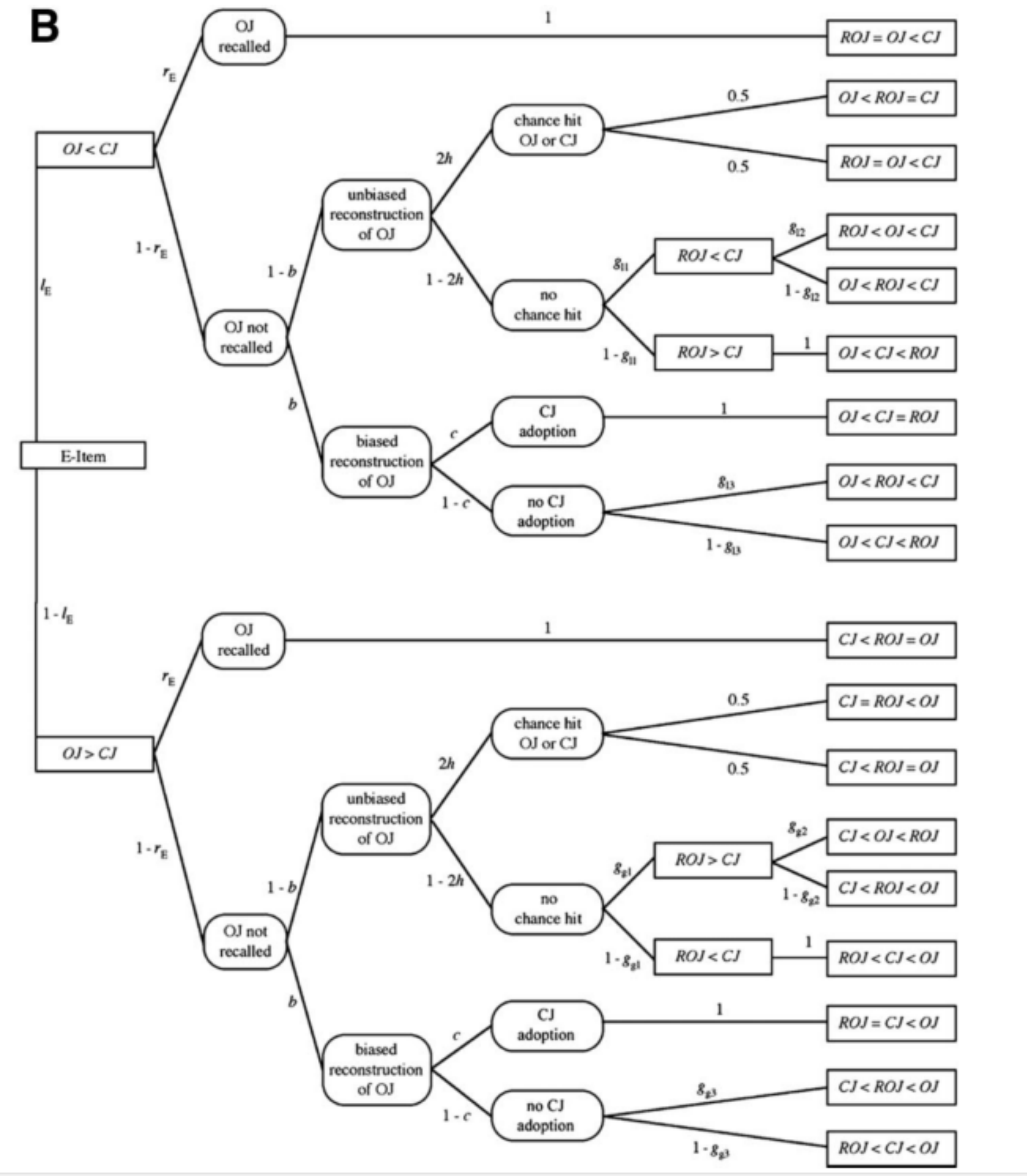


Figure 1. The HB-13 multinomial processing tree model by Erdfelder and Buchner (1998) for HB memory paradigm. A: Processing tree for control items. B: Processing tree for experimental items. For parameter interpretations see Table 2. The italicized *OJ*, *CJ* and *ROJ* represent the corresponding numerical values. 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, p. 392. Copyright 1998 by the American Psychological Association. Adapted with permission.

Figure 1A illustrates the model tree of the processes and the resulting rank order categories of the control condition when no feedback is presented. When participants are asked for their OJs they have the probability of l_C to underestimate CJ, giving an OJ that is lower than CJ, consequently generating the upper part of the control tree. The lower portion of the tree describes the state of an overestimation of the CJ (i.e., giving an OJ that is higher than CJ), with the complementary probability $1 - l_C$. Later when participants are asked for their ROJs, they have a probability of r_C to correctly recall their OJs, leading to the rank order ROJ = OJ < CJ. When this direct recall fails, with a probability of $1 - r_C$, participants may engage in a reconstruction action that is not biased with control items due to the absence of CJ presentation. The unbiased reconstruction may generate a result of ROJ with a chance of hitting either the OJ or the CJ (with the probability of h for each possibility), or with no chance to hit OJ and CJ (with the complementary probability of $1 - 2h$). With the no chance-hit path, ROJs are obtained through a guessing process, with a probability of g_{l1} generating a ROJ that is lower than CJ. The complementary probability $1 - g_{l1}$ thus denotes the distribution of a higher ROJ than CJ, eventually resulting in the rank order of OJ < CJ < ROJ. The lower ROJ compared with CJ could be either even lower than their OJ, with a probability of g_{l2}; or higher than OJ, with the complementary probability $1 - g_{l2}$. The lower part of the tree is a mirrored version of the upper tree section, and thus follows the same logic.

Figure1B represents the processing of experimental items wherein CJs are presented to participants. The tree structure mirrors the tree for control items (Figure 1A) apart from the involvement of the CJ influence on the reconstruction process. More specifically, the corresponding parameters are l_E and r_E with the experimental items l_C and r_C with the control items described above. In the situation of a failed direct recall of the OJ, the reconstruction process may then be biased by the CJ, with a probability b. The biased process may either take the path of falsely adopting the CJ as their OJ, with a probability of c; or the path with no such mistaken adoption, this having the complementary probability $1 - c$. Note that in this study, I adopted parameter c (CJ adoption) as in Pohl et al. (2010) instead of parameter s (source confusion) as in Erdfelder & Buchner (1998) classic HB-13 model. In the current study, CJ and ROJ are presented on the same screen and at the same time. Thus, CJ adoption is more adaptable than source confusion. Subsequently, if the biased reconstruction process occurs with no adoption of the CJ as OJ, participants are to subsequently engage in a guessing process to generate a ROJ that is either in between the OJ and CJ with a probability of g_{l3}, or biased toward and higher than CJ, with a complementary probability $1 - g_{l3}$. The lower part of the tree for experimental items is once more the mirrored version of the upper tree section, as with the control items. The model equations are obtained by adding all probability branches (i.e., by multiplying the probabilities within one branch) that lead to the same rank order of a tree.

The estimations of the parameters can be obtained from the rank order frequencies via maximum-likelihood estimation (Hu & Batchelder, 1994). The validity of the model's assumptions is tested via the goodness-of-fit tests indicated by the likelihood ratio chi-square statistics G^2 (Hu & Batchelder, 1994).

The recollection bias can be tested through a comparison between the parameter estimations r_C and r_E. If the null hypothesis $r_C = r_E$ is supported, then memory for OJs is not influenced by the CJ presentation, indicating an absence of recollection bias. If however $r_C > r_E$, memory for OJs is then impaired by the CJ presentation, thus reflecting the presence of recollection bias. The HB-13 model by Erdfelder & Buchner (1998) has been successfully applied to many empirical HB studies. They conducted and presented 4 such successful experimental examples in their own paper for developing and validating the model. In almost all previous studies concerning HB memory paradigm, the model met the criteria of goodness-of-fit with parameter estimations being satisfactorily consistent with the hypotheses of corresponding manipulations. The current study also employed the multinomial processing models to analyse HB data. As described above, only a small modification was made comparing to Erdfelder and Buchner (1998)'s classic HB-13 model. That is, CJ adoption (parameter c) instead of source confusion (parameter s) was used, due to the simultaneous presentation of CJ during ROJ task. Thus, the CJ presentation timing here is also an influential issue of memory design of HB, which is discussed in the next session.

The way OJ CJ ROJ are presented

The critical element here is CJ presentation timing. In the empirical studies, CJ is either presented after OJ, during ROJ or before ROJ. The study by Hell et al. (1988) falls in the first and the third category at the same time. In fact, time of CJ presentation was one of the manipulating variables in their study. Subjects were given OJs of almanac questions in the first session. They were then asked to write down an exact numerical estimate for each question. In the same session, having completed all questions, subjects were given the second booklet with CJs of half of the questions (i.e., CJ immediately after OJ). One week later, at the second session, the same subjects were given a booklet with CJs of the other half of the questions. Immediately afterwards, the same 88 questions used in the OJ phase were presented and subjects' ROJs were obtained (i.e., CJ immediately before ROJ). The results of this work showed that time of CJ presentation was a significant manipulating factor. HB increased when CJ was presented later at the second session (i.e., immediately before ROJ).

Bayen et al. (2006) however, presented the CJs during the ROJ phase in their Experiment 1. For half of the items, the CJs were presented at the same time as corresponding questions, subjects being asked simultaneously to give their ROJs. These were served as experimental items, as opposed to control items with no CJ presentation. An experimental example would follow this form:

In what year was the Monkey Wrench invented?
The Monkey Wrench was invented in 1841.
What was your original answer?

Perhaps unsurprisingly, the effect of presenting CJ was significant despite the relatively irregular timing of presentation.

In their second experiment, Bayen et al. (2006) presented CJs immediately before ROJs rather than simultaneously with ROJs. The manipulation of whether or not CJ was presented in the physical retrieval environment was designed to examine the hypothesis that knowledge of CJ in working memory is able to yield significant HB; furthermore, this second experiment tested the age difference found in Experiment 1. In addition, they encouraged subjects to encode CJs deeply in their working memory by giving them a recollection of CJ task (RCJ). Results supported their hypothesis that HB was present, this being demonstrated by the comparison of HB index between experimental (with CJ) and control (no CJ) groups. In regard to age difference, results remain significant as with Experiment 1, indicating that older adults showed larger HB than younger adults did. Thus, they concluded that increased hindsight bias among older adults was not due to the salient physical presence of CJ in the retrieval environment.

While the retention interval between CJ presentation and ROJ attempt was short in Bayen et al. (2006)'s Experiment 2, i.e., immediate condition, they added a delay condition wherein subjects were required to carefully "read and study" CJs that were presented one at a time for

6 seconds. CJs are less accessible for older adults compared with younger adults as CJs are retrieved via the delay condition from long-term memory, which is evidently more difficult for older adults (Smith, 1996). As a result, they hypothesized that given the delay condition, older adults should not have significantly larger HB than younger adults, and perhaps even have smaller HB (i.e., a reduced or reversed age difference in reconstruction bias).

Although CJ presentation timing varies in previous empirical work, no cross comparison has been done to systematically examine the effect of when and how CJ is presented across all manipulations. This omission is largely due to the interactive influence of other factors on HB, for example the type and design of HB tasks used or the retention interval between the OJ and ROJ phase, etc. However, researchers did manipulate CJ presentation timing to test conflicting theories. As discussed above in regard to the Hell et al. (1988) example, keeping the time between OJ and ROJ constant varied the CJ presenting time to either immediately after OJ or immediately before ROJ. According to Fischhoff (1975)'s *assimilation* theory of HB, about the same amount of HB should occur under these two conditions, as the assimilation is supposed to be *immediate* and a stable memory trace is made when CJ is presented (i.e., immediate assimilation), irrespective of when this presentation occurs. However, if this assimilation is supposed to gradually increase with longer coexistent time of OJ and CJ memory traces, then HB should be larger in the earlier CJ presentation condition. Conversely, if this assimilation is supposed to gradually decrease (or decay) over longer periods of these two memory traces, then a larger HB should be seen with the later CJ presentation condition. Their results supported the decay theory. It should be noted however that this immediate assimilation theory is only one among various theories explaining HB, particularly the almanac question HB task. Hell and colleagues' (1996) results could also be explained by other HB theories, for example inhibition theory; that is, that later CJ presentation made the correct information more salient and harder for subjects to inhibit during ROJ, thus eliciting larger HB.

Differing theoretical approaches to explaining HB will be explored in the next section. Before discussing these theories, I shall turn to typical applications of HB in recent literature.

Applications

HB has been applied in domains ranging from scientific study to the activities of daily life (Henriksen & Kaplan, 2003). Berlin (2000) presented a malpractice case in radiology and discussed the critical influence of HB in such situations. In recent research on applications of

HB, Arkes (2013) presented the involvement and importance of HB in medical malpractice. Researchers have also given attention to the detrimental effects of HB in legal decision-making situations (Harley, 2007). More specifically, the robustness of HB has presented a difficult challenge for jurors tasked with ignoring the negative outcomes of a given crime in order to consider a defendant's actions in a fair manner. In such cases HB is considered to be more detrimental than beneficial in terms of reaching just and desirable outcomes. Thus, exploring how HB could be eliminated has been the main focus of discussion.

However, in other domains, HB could be considered benign or even valuable. Hölzl & Kirchler (2005) studied HB in an economic development environment, demonstrating that euro supporters were more hindsight biased for positive developments than for negative ones, while euro opponents showed the opposite pattern. The authors therefore argued that people are highly motivated to re-judge the current situation using new information supporting their attitudes, in turn serving to stabilize their subjective belief systems. Other positive consequences of HB were discussed in Tykocinski, Pick, & Kedmi (2002). Participants in their study showed more HB when outcomes were undesirable or unfavorable, such that in retrospect the outcome seemed almost inevitable. Through this retroactive pessimism, people could regulate their disappointment after events. Louie (1999) and Louie, Rajan, & Sibley (2007) also observed benefits as a result of HB, but from an inverse perspective to the above. They argued that HB possesses a self-serving function in that it is present in favorable outcomes (e.g., "I knew I would succeed") and absent with unfavorable outcomes (e.g., "My idea should have worked") in self-involving events. Through this asymmetric pattern, people could accept credit for favorable outcomes and deny blame for unfavorable outcomes. The inconsistency of these two arguments could be caused by whether or not a given setting or context led to personal behavior, i.e., the extent of self-relevance to the event.

In summary, HB has profound practical implications in an array of domains, both experimental and real-life; indeed, some of the latter being crucial to people's lives, e.g., in forming legal decisions and in medical diagnoses. It is thus important to study and understand the underlying mechanisms of HB for both scientific and practical purposes.

Developmental findings on HB

As with many other pervasive cognitive phenomena, researchers have been interested in developmental trends associated with HB. While most studies have been conducted among younger adults (predominantly university students), Bayen et al. (2006) were the first researchers investigating HB in older adults. Using almanac questions, they found no consistent form of age effect. The first two experiments exhibited larger HB in older adults compared with younger adults. In their 3rd experiment, no such difference was found. The authors argued that in Exp.1 & 2, participants were given an instruction to encode CJ, thus CJ was present in working memory. Older adults find it more difficult to suppress CJ as task-irrelevant information during ROJ, leading to larger HB. In Exp.3, not only was there no instruction to encode CJ, it also appeared minutes before ROJ. Under such conditions, CJ had to be retrieved from long-term memory to interfere with ROJ. Because older adults have relative difficulty in retrieval, leading to an equally exhibited "intact" function with regard to inhibiting CJ's access to working memory, no difference in HB task performance was found vis-à-vis their younger counterparts.

Further developmental study of HB has been conducted with children of different age groups. Smith's (1999) Experiment 2 could be seen as an anchoring design of the HB paradigm. In that experiment, third graders, fifth graders and college students were asked to estimate the number of jellybeans in a glass container after they were given a question sheet asking whether there were more or fewer than either 50 (low anchor group) or 250 (high anchor group) jellybeans in the container. While results showed an anchoring effect, no age difference was found in the extent of this effect. Due to material limitations, HB studies have not commonly been conducted with children.

Bernstein et al. (2004) developed a visual HB paradigm that can be easily conducted with pre-school children as well as adults. Participants were asked when a naïve peer would identify a drawing that gradually becomes more defined on a computer screen, under the condition that they either knew what object the drawing was (baseline condition) or they did not know (hindsight condition). Only when different objects were used in these two conditions was a declining age effect found such that adults showed smaller HB than their pre-school counterparts. When the same objects were used in both baseline and hindsight conditions, this age effect of HB disappeared. The authors did not offer any specific explanation for this disappearance of age effect. We may only infer that the declined HB in adults compared with younger children found in their Exp.1 was not sufficiently robust to be replicated in Exp.2.

Pohl & Haracic (2005) used a hypothetical design with almanac questions as materials investigating HB in fourth, sixth and eighth graders and in university students. They did find an age effect indicated by a significant anchor x age interaction; however, the effect size was relatively small ($f^2 = 0.03$, 1-ß = .70).

The more recent developmental study on HB was conducted with 9-year olds, 12-year olds and adults (Pohl et al., 2010). They compared the results obtained from traditional variance of analyses methods (ANOVAs) with results using the multinomial modeling analyses specifically tailored for HB (i.e., the HB-13 model by Erdfelder & Buchner, 1998). While the ANOVAs detected no difference in the shift measure of HB across each of the three age groups, the multinomial modeling analyses showed that only 9-year olds exhibited high recollection bias and more adoption of the feedback.

Most recently, Coolin, Bernstein, Thornton and Thornton (2014) compared young adults in their HB occurrence and magnitude with older adults. Age effect in general was confirmed in both aspects of HB. That is, older adults were more often to show HB than young adults. Moreover, HB magnitude was larger in older adults than in younger adults. More importantly, Coolin et al (2014) argued that whether HB occurred was dependent on participants' episodic memory and inhibition. However, these two cognitive variables did not predict how big HB was once it occurred. Regarding how could episodic memory and inhibition accounted for the age differences in these two aspects of HB, no answer was drawn due to the limitations of the research method. Nonetheless, inhibition was proposed to be critical in decrease of HB of older adults. Further study by Coolin, Erdfelder, Bernstein, Thornton and Thornton (2014) incorporated several cognitive covariates proved to be critical to HB into the core parameter estimation process on the basis of the classic HB13 model (Erdfelder & Buchner, 1998). The results indicated that while recollection bias of HB was influenced by three cognitive factors – episodic memory, working memory capacity and inhibition; reconstruction bias was mostly and only dependent on inhibition, but not on the other two factors. Although this pattern of individual differences were only found in their study with older adults, but not with younger adults, the importance of inhibition was again affirmed. Gross and Bayen (2015) further approved this notion. They found that the typical increase of HB in older adults compared with their younger counterparts disappeared when a longer retention interval (RI) was inserted in the HB task. In their study, young adults with a long RI performed even worse in recalling their OJs than older adults did. However, these young adults did not show a higher reconstruction bias than older adults. This was most probably due to their better inhibitory control ability that avoided ROJ was more biased by CJ. This finding further strengthened the important role of inhibition might play in memory paradigm of HB.

In general, the developmental findings have been rather mixed revealing several methodological problems. First, the experimental designs differ substantially. Some adopted an anchoring design, whereas others used hypothetical design and memory design, presented either by text questions or visual tasks. Although the main focus has been on the HB concept, disparate cognitive underlying mechanisms could be involved given the different experimental design and presentation formats. Second, most of the studies discussed above analyzed data using the traditional statistical method (ANOVAs). As such, apart from the age difference in HB at large, very little in the way of further results capturing the detailed processes underlying HB was obtained. In current HB literature, only five studies have utilized the MPT method to analyze developmental data, namely the above-mentioned Bayen et al. (2006), Coolin et al. (2015), Groß & Bayen (2015), Pohl & Haracic (2005) and Pohl et al. (2010). The former three studies compared HB of older adults and younger adults while the latter two compared children and younger adults. With the more sophisticated MPT analysis, these five studies suggested a U-shape developmental trajectory of HB, with younger adults being in the middle demonstrating the smallest degree of HB compared with children and older adults. The researchers explain this U-shape trajectory with cognitive aging account; more specifically, the inhibition-related function deficit with older adults and children. Moreover, there is also indirect evidence suggesting the relationship between HB and inhibition-related functions, e.g., the interaction between inhibitory control and children's eyewitness memory (Roberts & Powell, 2005), which is considered to operate within a similar paradigm to the memory type HB. However, the direct correlation between inhibition-related ability and HB was found not to be significant (Exp.1 in Bernstein et al., 2007; Pohl, 2008). There could be several explanations for the non-significant correlation between HB and inhibition. First, the correlational design is not powerful enough to find the relationship. Further, the low reliability of HB tasks using almanac questions may lead to the non-significant findings in a correlational study (Musch & Wagner, 2007). Thus, the current study employs a more powerful experimental design (for more detail see the method section). Second, in Bernstein et al (2007)'s correlational study, a visual HB task was used, instead of the memory text-based HB task in both Bayen et al (2006)'s and Pohl & Haracic's (2005) studies. The third reason, which is perhaps the most plausible and worthy of further investigation, is that apart from inhibition, there could be other essential cognitive mechanisms or processes involved in the HB paradigm. The current study is thus devoted to discovering the more fundamental underlying mechanisms. Before considering the current study in detail, other theories and models explaining HB in the literature will be summarized in the next section.

1.2. Theories and models explaining HB

In general

HB has been broadly demonstrated in retrospective tasks. However, these tasks may differ in a variety of aspects (see Pohl, 2004, 2007 for a review). First, there are different ways that the relevant information is presented. Some tasks having required participants to give true or false for statements (Fischhoff & Beyth, 1975; Wood, 1978; Pohl, Stahlberg, & Frey, 1999; Musch, 2003) while others have used *2-alternative-forced-choice* (2AFC) questions (Fischhoff, 1977; Hoffrage, Hertwig, & Gigerenzer, 2000). An example of the latter would be *"Is absinthe (a) a liqueur or (b) a precious stone?"* (Fischhoff, 1977). Participants are required to select the correct definition of the word absinthe and then give a confidence rating showing how confident they are in their choice. In my current study, I am only concerned with HB tasks that require exact numeric responses as presented above in the Task session.

Pohl (2007) argued that different types of HB task could lead to subjects using different processes to make their initial judgments or estimates. For example, the cognitive processes involved in the almanac question task for subjects to formulate their numeric estimates may be of a different order to the processes in the confidence-rating task in which metacognitive assessment is engaged (see also Ash, 2009). Even with the same type of HB task, the almanac trivia question being an example, Hertwig et al (1997) argued that different cognitive processes might be involved in different designs (hypothetical vs. memory, discussed earlier) of the same almanac question task. In fact, they suggested the term *hindsight bias* only for memory design and *knew-it-all-along effect* for hypothetical design.

Moreover, the nature of the feedback a person receives also varies. The anchoring design used in Smith (1999)'s Experiment 2, for example, involves the estimation of the number of jellybeans in a jar. Although this task shares a many similarities with the almanac general knowledge question, the nature of subsequently presented information is different. In the HB task of current study, information presented later is the *correct* answer of the question. Thus, subjects have a greater tendency to learn about this information than in the jellybean-anchoring task, which resembles more a pure retroactive interference paradigm.

In sum, when considering theories explaining HB, any given theory should only be used to explain a specific and defined type of HB as there are different processes involved in every stage of each type of HB (Blank, Musch, & Pohl, 2007). The current study limits its focus to memory design using numeric almanac trivia question tasks and comparison between subjects. Thus, the metacognitive lines of particular theories (e.g., expectation- and experience-based adjustment theories by Ash, 2009; Pezzo, 2003; Renner, 2003; Sanna & Schwarz, 2007) will not be included for discussion in this dissertation. Moreover, subjective attitudes and causal attributions do not figure in the "cold material" of the current study, being relevant only to more applied situations, e.g., economics (Hölzl & Kirchler, 2005; Pezzo & Beckstead, 2008). Now I shall turn the attention to the main streams of theory explaining the HB task in this study.

Cognitive reconstruction theories

Cognitive reconstruction theories suggest that people tend to re-estimate the question when asked for their previous judgment, after having failed to directly access their initial estimation in episodic memory (Stahlberg, 1994; Stahlberg & Maass, 1997). The critical outcome information conceivably exerts its effect in both the access failure of OJ and the re-estimation process. The memory trace of OJ could be destroyed or disturbed, which makes it harder for subjects to directly recall their OJ in episodic memory. Thus, the re-estimation process is engaged; however, this process is influenced by the existence of outcome information leading to a deviation from their initial judgment (Guilbault et al., 2004).

The *Anchoring and adjustment* notion also falls into the category of cognitive reconstruction theory. This notion proposes that the outcome information acts as an anchor for people to reconstruct their initial judgment. They give their retrospective judgment adjusting away from this anchor through heuristic cues.

The *updating and re-judging* notion asserts that the outcome information updates the very knowledge-base upon which an individual's re-judgment is based (for a review, see Hawkins & Hastie, 1990). Automatic association theory by Fischhoff (1975), for instance, suggests that the presentation of outcome information induces an automatic assimilation with the existing knowledge units. Therefore, the later re-judgment produces a deviated ROJ based on the changed knowledge units.

Both lines of theory contend that there is a reconstruction stage when people are asked for their previous estimation of a question or a situation. There is little agreement however on when and how the outcome information applies its effect. Both theoretically and empirically, researchers have differed as to: 1) whether or not the outcome information erases the OJ memory trace and if yes, to what extent. That is, whether the memory traces of OJ and CJ co-exist, rather independently, or if the latter blends in the former (Fischhoff, 1975; Fischhoff & Beyth, 1975; Hell et al., 1988; Pohl, Eisenhauer, & Hardt, 2003). Thus, in turn, 2) in the ROJ process, whether people tend to engage in resisting the interference of CJ as they try to directly access OJ; or they instead attempt to reconstruct their initial estimation (Ash, 2009) and the re-judgment is deviated by the outcome information. While the former is empirically supported by developmental studies in HB (discussed earlier), the latter finds evidence in same age studies. It should be noted, however, that there may be no absolute answer for the discrepancy. Instead, due to the high variance of HB tasks, some strands of theory explain certain situations better, while simultaneously failing to account for other situations. More specifically, whether or not the memory trace of OJ and CJ is rather independent or merging could be dependent on several conditions: CJ presentation timing or subject characteristics (e.g., learning-motivated or indifferent; expert or layman in the task material; age, etc.). Erdfelder & Buchner (1998) postulate an account reconciling the second argumentation, i.e., HB could be explained by both recollection bias and reconstruction bias (see also Erdfelder, Brandt, & Bröder, 2007).

Recollection bias and Reconstruction bias

The typical result of the numeric memory HB task is that subjects give some correct ROJs that are identical to their OJs. However, the percentage of correct ROJs is lower when outcome information is presented (as in the experimental condition) than when no outcome information is presented (as in the control condition). Heuristically, subjects first engage in direct recall of their OJ from memory when asked what they have answered previously. Failure in this process leads to a reconstruction. As it is unclear whether these correct OJs are due to a successful direct recall, matching results of re-judgment, or perhaps simply guessing, the presentation of outcome information could exert its influence on the memory of OJ resulting in less successful direct recall; or on the re-judging process resulting in unmatched ROJ. Erdfelder & Buchner (1998) postulate that two forms of biases build up HB, the *recollection bias* (when CJ presentation affects the direct recall process) and the *reconstruction bias* (when the re-judging process is impacted). Thus, previous theories explaining HB could be categorized as being one of these two types (Table 3).

Table 3
Accounts Explaining HB Based on Recollection Biases and Reconstruction Biases

Recollection bias *could be due to -*		
	CJ altering the OJ memory by	destroying the OJ memory completely by assimilation, or
		adding to OJ memory so that both OJ and CJ co-exist
	Retrieval barrier of the OJ due to CJ	
Reconstruction bias *could be due to -*		
	Anchoring and adjustment	
	Re-judgment based on stronger accessibility of CJ-congruent information	
	Reconstruction based on the updated knowledge after CJ presentation (e.g., sense-making theory, Pezzo & Pezzo, 2007).	
	Reconstruction based on the influenced metacognitive inference of	the feeling about the answer, or
		the feeling about one's ability in a certain knowledge domain

Empirically, some studies support only the reconstruction bias view that OJ memory is not impaired by CJ (Schwarz & Stahlberg, 2003, Exp.1); whilst others argue that recollection bias and reconstruction bias work together *in a graduated fashion, rather than all-or nothing* (Pieters et al., 2006). Moreover, as demonstrated in Table 2, recollection bias and reconstruction bias should be viewed as types of biases in HB or components of HB, and are *not* detailed explanations of HB. Thus, researchers began to construct more systematic models to explain HB. The most influential ones are: the Selective Activation, Reconstruction and Anchoring (SARA) model (Eisenhauer, 2000; Pohl et al., 2003), the Reconstruction After Feedback with Taking the best (RAFT) model (Hoffrage et al., 2000; Hertwig, Fanselow, & Hoffrage, 2003) and Causal Model Theory (CMT) (Blank & Nestler, 2006; Blank & Nestler, 2007; Nestler, 2007). However, only two of these (SARA and RAFT) concern the memory paradigm of HB. Furthermore, only SARA attempts to explain HB generated by way of almanac questions requiring numerical judgment. Thus, I will only discuss the SARA model in detail.

Selective Activation, Reconstruction, and Anchoring (SARA)

The SARA model (Eisenhauer, 2000; Pohl et al., 2003) systematically elucidates rather detailed cognitive processes incorporating the two types of HB discussed above. SARA (based on the memory model *SAM*: Search of Associative Memory, Raaijmakers & Shiffrin, 1980) assumes that people possess a network consisting of associative *image sets* which they attach to certain knowledge. Concerning the typical memory HB paradigm, the relevant associative *image sets* (e.g., a, b, c, d, &e) are chosen to generate OJ. When CJ is presented, the *image sets* that are more congruent to the correct answer (e.g., c, d, &e) are *selectively* strengthened. Later in the ROJ phase, these strengthened *image sets* are thus more activated, leading to a biased ROJ toward CJ. Moreover, SARA proposes that CJ could also induce a *biased sampling* by acting as a cue that facilitates the access to the adjacent *image sets*. SARA not only clarifies HB theoretically, but the model also successfully simulates the recollected judgment in empirical studies (Pohl et al., 2003).

Despite the success of SARA both on a theoretical level and in terms of empirical prediction, the model has limitations, specifically in the ***selective strengthen*** phase when feedback is presented. SARA implies that this *selectivity* is merely at the *objective* level wherein only feedback-congruent image sets are strengthened. This fails to observe the *subjective* level in which only plausible feedback could provoke this strengthening process. Although the theoretical model overlooks this subject level, it has often been demonstrated in empirical studies as the plausibility of feedback information.

Plausibility effect in HB

Researchers have attempted to eliminate HB magnitude by discrediting feedback information or manipulating the plausibility of feedback. Findings have been mixed in both HB and other similar anchoring effect literature. Hasher, Attig, & Alba (1981) failed to discredit feedback information in their Exp.1. Strack & Mussweiler (1997) found no difference in the anchoring effect when both implausible and plausible anchors were presented. What is more, no elimination of HB was found when the feedback was labeled as “another person’s estimate” and values that were incorrect by a large margin were provided as feedback in Pohl (1998). In contrast, successful discrediting can be found in Choi & Choi (2010), Erdfelder & Buchner (1998) Exp.3, Hardt & Pohl (2003), Hasher et al. (1981) Exp.2 and Wassermann, Lempert &

Hastie, 1991). This inconsistency could in large part be due to 1) the mixed type of anchoring or HB tasks used, and 2) the level of the manipulation (implicit vs. explicit). Concerning the first point, a large body of research on HB plausibility effect has focused on the event probability aspect (for a review see Choi & Choi, 2010). Moreover, Hasher et al. (1981) used a true-false stating task and successfully discredited the feedback information in their Exp.2. Others used factual knowledge questions requiring absolute numeric judgment (Strack & Mussweiler, 1997; Pohl, 1981; Erdfelder & Buchner, 1998 and Hardt & Pohl, 2003). Regarding the level of the manipulation, almost all feedback discrediting was only manipulated at a rather implicit level, i.e., feedback information was presented as an anchor exerted by answering a comparative question (e.g., whether Brandenburg Gate is taller or shorter than 150m) (Strack & Mussweiler, 1997). Although highly incorrect values were provided as feedback, Pohl (1998) labeled the values as "another person's estimate" instead of "false". Only two cases explicitly told participants that the presented feedback was wrong (Exp.2 in Hasher et al., 1981 and Exp.3 in Erdfelder & Buchner, 1998).

Taken together, when feedback information is presented in a rather explicit manner (instead of an implicit exertion) in an almanac question task, the plausibility of the feedback does have a significant effect on HB.

The SARA model assumes feedback information strengthens the congruent image sets, regardless of its plausibility. It therefore lacks an evaluation phase that would allow for selective strengthening (at the subjective level). With such an evaluation phase, only plausible feedback could be encoded leading to a strengthening of congruent image sets. When, however, implausible feedback is presented, the evaluation process could initiate an inhibition activity that blocks the strengthening. The "kicking-in" of the inhibition process is also highly consistent with the developmental findings of HB discussed above. Notwithstanding this, as stated in the section "developmental findings on HB", previous studies have failed to establish a direct correlation between inhibition and HB. The puzzle now is thus: it is highly probably that inhibition functions are involved in the HB memory paradigm, this despite the fact they have not been observed in empirical studies. The primary aim of this study is to solve this problem using a more powerful experimental manipulation: depletion procedure.

1.3. The current study

Why this study?

Despite fruitful findings in previous studies of HB, we lack a systematic theory to explain the primary lines of results stemming from them, these being the developmental trajectory and the rather stable and robust plausibility effect (again keeping in mind that in this study, only memory design using factual knowledge questions requiring absolute numeric judgment is under consideration). Moreover, there has been considerable debate concerning whether HB is an automatic process that is part of our evolutionary heritage or a controlled process. More specifically, if it is automatic, how could this automaticity be involved? And if it is controlled, what particular cognitive function plays an important role? To sum up, the present picture forming around memory HB can be described as:

- A U-shaped developmental trajectory
- Difficulty in preventing HB from occurrence, even under overt warning to avoid the bias
- Implausible feedback resulting in clear reduction of HB
- No direct evidence of a significant correlation between inhibition and HB

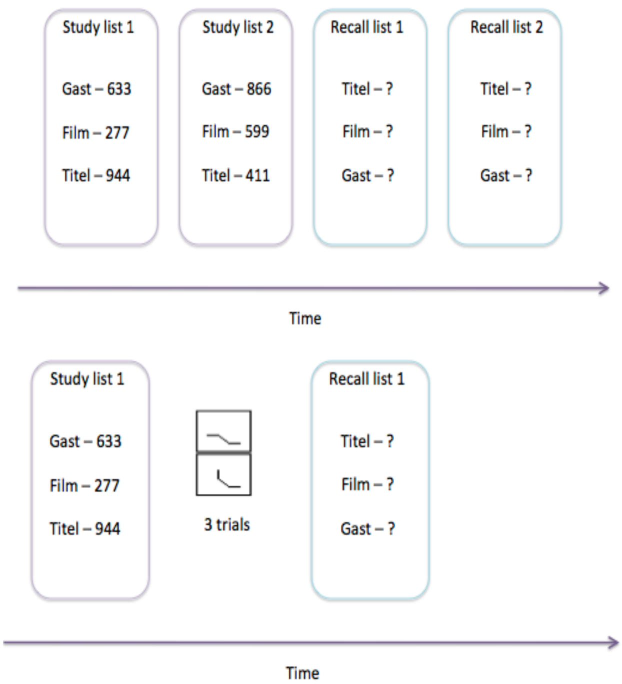
As discussed, no previous theory or model can reasonably clarify these quite incongruous fragments of evidence. Therefore, further investigation using a different manipulation is necessary to give genesis to more progressive theory, one which might help us to better understand the underlying mechanisms of this robust phenomenon.

Introduction to the current study and hypotheses

I intend to apply a depletion procedure on the memory paradigm of HB. Based on previous findings and suggestions, this depletion procedure will not be on general cognitive resource, but on a related specific cognitive function: inhibition. In order to be more targeted, it is first necessary to establish an appropriate depletion task for the current study. In prior work examining the relationship between inhibition and HB, different types of inhibition tasks have been used. Bernstein et al. (2007) applied two different inhibitory control tasks: Day/Night and Bear/Dragon tasks (Carlson & Moses, 2001). Pohl (1998) and Pohl, Auer, Bayen, & Martin (in prep.) used a battery of inhibition tasks, such as word recall (testing proactive interference), a pattern comparison task, an embedded figures test and stroop tasks (digit count and color words) to examine the correlation between inhibition and HB. They found significant correlations for some of the inhibitory tests, e.g., the pattern comparison task and embedded figures test. Bernstein et al. (2007) chose inhibition tasks so that their requirements were similar to theory of mind tasks, with which they intended to investigate the relationship with inhibition. Likewise, however, they failed to find a relationship between inhibition and HB in one of their two experiments, most probably because they ignored whether these inhibition tasks are similar in nature to HB tasks. Thus, I shall select inhibition tasks for the depletion procedure ensuring they are related in nature to the memory HB paradigm of current study. Reviewing the inhibition literature, the classic retroactive interference design (by Müller & Pilzecker, 1900, cited in Anderson & Neely, 1996) is chosen as the best candidate for the inhibition tasks for 2 reasons, as follows:

1. The core structure of this design is based on the paired-associate methodology to test the interfering effect of learning new information on recalling previously-learned materials. This is very similar in nature to the current HB memory task.
2. Previous findings of significant correlation between children's retroactive inhibition skills and similar cognitive paradigm (eyewitness memory) (Roberts & Powell, 2005).

For the current study, a slightly modified AB-AD task is applied to further suit the current HB memory task (see Figure 2).



Experimental condition

Control condition

Figure 2. Illustration of the modified AB-AD task based on Müller & Pilzecker (1900)'s classic retroactive interference design.

As shown in Figure 2 which illustrates how semantic questions are presented and numeric answers are required in the current HB task, I shall change word-word pairs to word-number pairs. Moreover, empirically, the length of the word pairs in each study list is crucial in order to yield an appropriate amount of retroactive interference; of additional benefit, participants will not be fatigued by an protracted list. Taking into account these methodological factors, Experiment 1 aims to develop the best combination of characteristics of the AB-AD task for the purpose of the current study. In turn, Experiment 2 applies the depletion procedure to ensure that retroactive inhibition can indeed be depleted after intensive similar inhibition-related tasks, and to validate operative tasks used for the depletion procedure. Subsequently, in experiment 3, this effective depletion procedure will be applied on top of the HB task to discern whether the critical underlying process of HB is indeed retroactive interference resistance. The core logic behind this methodology is based on the notion of ego depletion (R F Baumeister, Bratslavsky, Muraven, & Tice, 1998); that is, if two tasks share a common specific cognitive resource, then intensive performance of one task (interference tasks in the current study) should deteriorate the performance of the task succeeding it (HB memory task in the current study). Surprisingly, the results of experiment 3 are contrary to my expectation. According to the results of experiment 3, I introduced a new manipulation of HB design in experiment 4, that is, CJ is labeled either as “correct answer” or “wrong answer” of the question. Combining this labeling manipulation with the depletion procedure, experiment 4 aims to find out the more specific underlying cognitive processes, apart from the single interference resistance explanation. That is, both knowledge updating behavior and interference resistance ability are involved to decide the magnitude of HB.

Hypotheses

1. Regarding Experiment 1, participants should perform significantly worse in the experimental condition, wherein high retroactive interference is involved by studying list 2 (the same words paired with different numbers), compared with the control condition. Moreover, the best length of word-number pairs in one list would be confirmed for further use.
2. Regarding Experiment 2, participants should perform significantly worse in the AB-AD task after having engaged intensively in interference tasks in which similar cognitive processes are involved and high interference resolution ability is demanded (i.e., only with the high interference condition). In contrast, no such depletion effect

should occur after having done tasks that demand minimal interference resolution ability (i.e., only with low interference condition).

3. Regarding Experiment 3, HB should be significantly larger after depletion of retroactive interference resistance ability. There should be no such increase after having performed same number of tasks that require minimal interference resolution ability.
4. Regarding Experiment 4, there should be a plausibility effect in which HB is significantly larger when CJ is labeled as “correct answer” as opposed to when it is labeled as “wrong answer”. After depletion, HB should be significantly decreased in the “correct answer” labeling condition, as the knowledge-updating behavior would be exhausted via depletion. In contrast, HB should be significantly increased after depletion in “wrong answer” labeling condition, because the interference resistance ability is depleted.

2. Experiments

Experiment 1

In Experiment 1, participants were presented with a paired-associate ABAD task consisting of either 3, 4 or of 5 word-number pairs.

Method

Participants

Participants for all experiments in the current study were recruited from the University of Mannheim. They were paid either 12 Euro or a 1.5-hour credit to complete their Bachelor of Psychology course. A total of 75 students participated, of which thirteen were male. The participants ranged in age from 17 to 29 years (M = 20.57, SD = 2.33). All participants were native German speakers.

Design

The design was a 2 (interference level) x 3 (length of the word-number pairs within one list) x 3 (task order) mixed factorial design with interference level (high-interference or experimental vs. low-interference or control conditions) and length of word-number pairs (3, 4 or 5) as two within-subject variables and task order (presented with 3,4 or 5 pairs first) as a between-subjects variable. The dependent variables were the participants' response accuracy and time.

Materials

All materials in the current study were presented via computer. As mentioned, a modified AB-AD task (Müller & Pilzecker, 1900, cited in Anderson & Neely, 1996) was used (Figure 1). Each AB-AD task consists of an experimental (i.e., high interference) condition and a control (i.e., low interference) condition, with 3, 4 or 5 word-number pairs in each list. Words were chosen from commonly used nouns with 4 or 5 letters. They words were paired with 3 digit numbers, for example, 633, 277 or 844, so that 1) there could be more variations than using 1 or 2 digit numbers; and 2) participants were required to make some effort to remember them (being more difficult than single digit numbers); meanwhile 3) being easier to recall than entirely random 3 digit numbers, this level of effort would not over-occupy participants' cognitive resource, thereby prompting fatigue. In addition, a figure comparison task (Salthouse & Babcock, 1991) was applied to fill in the time gap between learning a list and recollecting a list in the control condition due to the absence of the second study list. This filling task (see Figure 1 control condition) was designed to be rather easy to complete so that minimal cognitive resources were demanded.

Procedure

All participants in the current study were tested in a seminar room with 2 computers. E-PRIME software was used for all experiments (Psychology Software Tools, Pittsburgh). Participants were asked to sign a consent form at the beginning of all experiments in this study. They were then given instructions about the experiment they were to take part in. Practice trials were then presented to help them understand the main task.

In Exp.1, participants were randomly assigned into three groups, and began with a study list consisting of either 3, 4 or 5 word-number pairs. Participants in each group completed 4 blocks of AB-AD tasks, with 10 trials of the figure comparison task carried out as refreshing tasks in between. Each block consisted of 4 sets of variations. Each variation comprised all 3 types of AB-AD task starting with different numbers of study pairs (3, 4 or 5). The variations were partly counterbalanced in the sense that only 4 combinations (3.4.5; 4.3.5; 5.4.3; and 3.5.4) were included. Thus, apart from the practice trials, each participant completed 48 AB-AD tasks and 174 trials of figure comparison tasks.

Figure 3 shows the detailed procedure within each AB-AD task. Each slide was displayed on the screen for 2 seconds, except for the fixation point slides, response slides and the figure comparison task slides. The latter two types were presented until participants typed in their responses; in the case of the figure comparison task however, a warning to speed up was presented after 5 seconds of each trial to ensure the total duration of the filling task was no longer than the time taken to study list 2 in the experimental condition. The whole experiment session took approximately 70 minutes with some variation depending on individual pace.

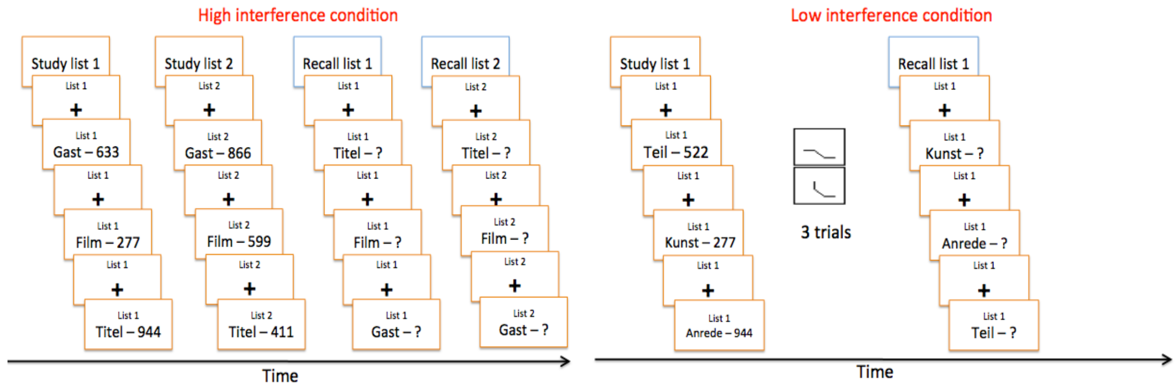


Figure 3. Procedure of one trial of AB-AD task.

Results

To minimize the effect of extreme values on the distribution, medians instead of means for each participant were computed for all experiments in this study. I also report the partial eta-squared (η^2) and Cohen's d as measures of effect size where appropriate (Cohen, 1988). For the final data analyses of Experiment 1, data from 2 participants were excluded due to too much missing data. For the remaining data from 73 participants, analyses of variance (ANOVAs) were conducted. I set the confidence alpha level (α) to .01 for the significant tests of all experiments unless noted otherwise. As stated above, both response accuracy (percentage of the correct recollection) and response time (mms) were analyzed via ANOVAs. Results showed that there was no significant main effect of task order both in response accuracy (starting with 3 pairs M = 38.32, starting with 4 pairs M = 38.93, starting with 5 pairs M = 39.52) and in response time (M_3 = 3159,48, M_4 = 3089,44, M_5 = 3226,44). Thus, I merged the data from all 3 groups and conducted a 2 (interference condition: high vs. low) x 3 (length: 3, 4 vs. 5 word pairs) ANOVA. My interest here focused upon the interaction of these two variables; specifically, which length of word-number pairs could yield the largest interference effect. Results showed a significant length by interference interaction in response accuracy only with the 5 word-pairs condition, F (2, 61) = 4.83, $p <$ .01, partial η^2 =.073 (see Figure 4a). Further paired sample t-test showed that participants' response accuracy was significantly lower in the high interference condition than in the low interference condition with word-number pairs being 5 within one list, t (63) = 3.15, $p <$ .01, d = .40. In terms of response time, no factor had a main effect or exhibited interaction of significance, F (2, 61) = .324, p = *.724*, partial η^2 =.005 (see Figure 4b). However, further paired sample t-test showed that when there were 5 word-number pairs within one study list, response time was significantly shorter in the low interference condition than that in the high interference condition, t (63) = 1.62, $p < .05$, d = .20.

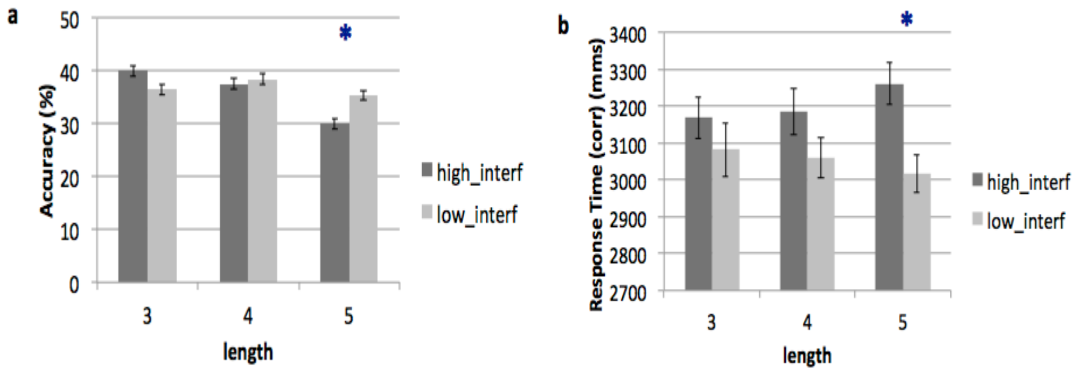


Figure 4. Response accuracy (a) and response time (b) of the AB-AD task in high (experimental) and low interference (control) condition with 3, 4 or 5 word-number pairs within one study list of Exp.1. Error bars indicate 99% confidence intervals.

Discussion

The main purpose of Exp.1 was to implement the modified AB-AD task and ensure the replication of the interference effect while applying the modifications discussed previously. Moreover, the number of pairs within one list eliciting the most significant interference effect would be detected. The current results successfully replicated the interference effect with the number of pairs within one study list being 5. When there were 3 or 4 word-number pairs in one study list, participants' recollection performance of the first list did not differ whether there was a second list to study or not. However, 5 word-number pairs within one study list could yield a significant interference effect by way of recollection of the first list when there was a second list to study. Participants successfully recalled substantially more numbers from study list 1 when there was no need to study a second list. Moreover, in regard to response time, significantly more time was needed to recall the first list when there was a second list to study (see Figure 4). In light of these findings, I decided to include 5 word-number pairs for the AB-AD task in subsequent experiments.

Experiment 2

In Exp.2, the depletion procedure was examined to ensure its effectiveness before the manipulation was applied on HB. I tailored the depletion procedure based on the ego depletion notion of Baumeister et al. (1998) and the depletion technique used by (Jonas Persson, Welsh, Jonides, & Reuter-Lorenz, 2007).

Ego Depletion

Baumeister et al. (1998) asserts that a limited resource may be involved in some cognitive activities; and that a prior use of such resource may result in a reduced amount of cognitive strength in subsequent activities. Muraven & Baumeister (2000) further argue that the executive functions – in particular inhibition-related ones – consume a limited cognitive resource. Employing tasks demanding inhibition ability may have a detrimental impact on succeeding inhibition tasks. More research on ego depletion revealed 3 critical characteristics that should not be overlooked. First, this reduction of cognitive resources following prior use does not last long. It is instead *temporary* and can be replenished (Baumeister, Gailliot, DeWall, & Oaten, 2006). Baumeister and colleagues made an apt comparison to illustrate this, suggesting that cognitive resource resembles muscle in as much as ego depletion, as with muscle, becomes tired after initial exertion (Baumeister, Vohs, & Tice, 2007; Gailliot, Plant, Butz, & Baumeister, 2007; Muraven & Baumeister, 2000). As with muscle fatigue, performance of subsequent tasks can deteriorate. Second, this reduction acts only on *controlled* activities that require active guidance in terms of self-control, but not on simpler activities in which no engagement of cognitive resource is demanded (Schmeichel, Vohs, & Baumeister, 2003). Third, this reduction is not likely to be general; that is, it cannot be transferred to any type of cognitive ability. Rather, the performance decrement after prior exertion is *specifically* applicable to types of cognitive behaviour within the same category as the prior behaviour (Muraven & Baumeister, 2000).

Depletion technique used by Persson et al. (2007)

Persson et al. (2007)'s study is a good example of the 3 points discussed above. They conducted 3 experiments on depletion of inhibition-related tasks and revealed a process-specific fatigue. More specifically, they argued and demonstrated empirically that if two tasks share a common type of cognitive resource that is limited, then intensive exertion of one task within one experimental session (i.e., training task, used during depletion procedure) should temporarily deteriorate the performance of a second task (i.e., transfer task, used for pre and post-test of the depletion effect) requiring the same cognitive resource that was depleted by prior use. In contrast, no such transfer effect after depletion should be observed if the transfer task and training task rely on functionally different and neutrally separate underlying mechanisms. In their Exp.1 and 3, the training task (letter recognition task) and the transfer task (verb generation task) involve a common specific cognitive process (interference resolution ability), thus revealing the negative transfer effect. Their Exp.2 however used a training task (stop signal task) that entails a distinct cognitive process (response inhibition) with the transfer task (verb generation task), thus showing no such transfer effect after depletion.

In Exp.2, I adopted the same depletion procedure used in Persson et al. (2007) and applied selected tasks from their study and Persson & Reuter-Lorenz's (2008) study. The latter is also a training study on inhibition-related abilities, but taking the other direction, i.e., regular exercise of a specific cognitive ability could enhance performance of tasks entailing the same cognitive ability. Detail of the procedure and tasks will be discussed in the method session below.

Method

Participants

A total of 51 students participated in Exp.2. Thirteen were male. The participants ranged in age from 19 to 27 years (M = 21.53, SD = 2.05). Four were non-native German speakers. They were paid either 8 Euro or a 1-hour credit toward the completion of their Bachelor of Psychology course.

Design

The design was a 2 (interference level) x 2 (group or depletion intensity) x 2 (time) mixed factorial design with interference level (high-interference or experimental vs. low-interference or control conditions) and depletion intensity (high vs. low depletion) as two between-subjects variables and time (pre vs. post-depletion) as a within-subjects variable. The dependent variables were response accuracy and response time of correct responses.

Materials

Three types of tasks were applied in this experiment, these being a transfer task, a training task and a filling task. The modified AB-AD task validated in Exp.1 was used as the transfer task. As in the Exp.1, a figure comparison task was used as the filling task. The training tasks during the depletion procedure were a verb generation task, a letter recognition task and the N-back task used in Persson et al. (2007) and Persson & Reuter-Lorenz (2008). All tasks could be manipulated to create a high interference condition and a low interference condition.

In the *verb generation task* (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997), participants were presented with a noun on a computer screen and were required to silently generate a verb that functions appropriately with that noun. In the high interference condition, the nouns could be connected to many different verbs (e.g., Ball – spielen, hüpfen, werfen etc.); as a dominant verb commonly associated with the noun is lacking, in order to choose a verb the high interference resistance ability is demanded. In contrast, in the low interference condition, the nouns are primarily partnered with a single, typically related verb (e.g., Schere – schnitten), thus conferring minimal interference resistance ability. The nouns in Exp.2 were from Persson et al. (2007)'s study and were translated into German.

In the *letter recognition task* (a verbal working memory task, Persson et al., 2007), participants were first briefly presented with 4 letters. A 'yes' or 'no' recognition response was then required to indicate whether a subsequent probe letter appearing onscreen was among the 4 letters. In the high interference condition, the probe letter was not from the current set, but from the previous (two) trials. Participants had to resist the interference elicited due to the familiarity of seeing it in previous trials, and to reject it by giving a 'no' response. In the low interference condition, the negative probe letter was not present in at least the previous two trials, thus requiring minimal interference resistance process.

In the *N-back task* (Persson & Reuter-Lorenz, 2008), participants were asked to press the left mouse button as quickly and accurately as possible when a word matched one of the *n* words presented earlier (n = 3 or 1). The words were presented on the computer screen at a rate of 3 seconds per word, with an additional 3-second interval between 2 adjoining words. I slightly modified the N-back task used in Persson & Reuter-Lorenz (2008)'s study. For the current Exp.2, sixty-four words were presented in total. Among all these words, half matched the previous item presented 3 items earlier (target words) while half did not (non-target words).

In the high interference condition, the non-target words were intermixed with other non-target words that were presented 2, 4 or 5 items earlier (but not 3 earlier, as per the target words). Thus, this intermixing yielded a high interference effect to be resisted due to the recent exposure of the non-target words. In the low interference condition, participants were simply required to indicate whether or not a word was repeated (i.e., $n = 1$), and thus minimal interference existed in this condition.

Procedure

There were 3 phases in Exp.2: pre-depletion, depletion and post-depletion, with training tasks applied during the intermediate depletion phase and transfer tasks used in the pre and post-depletion phases. The schematic depiction of Exp.2 is shown in Figure 5.

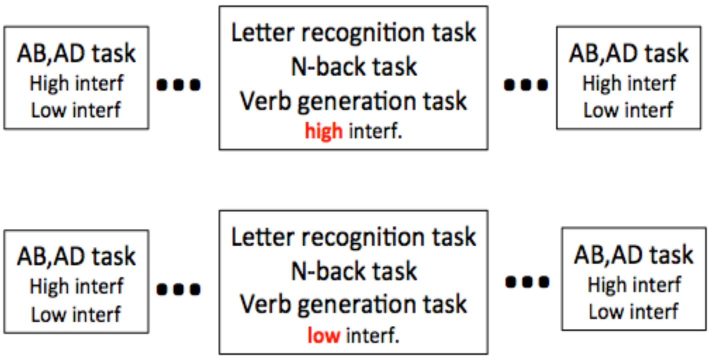


Figure 5. Schematic depiction of the procedure of Exp.2. From left to right, with tasks applied pre-depletion, during depletion and post-depletion. The first and second rows illustrate the high depletion group and low depletion group respectively.

The whole experiment session took approximately 60 minutes. The transfer task (AB-AD task) was performed during the first and last 15 minutes, with the intermediate depletion phase (approx. 30min) presenting the training tasks. Participants were given instructions at the beginning of each task. Practice trials were presented before the data collection of each task. Throughout the entire experiment I did not inform participants about the interference manipulations of both the transfer task and the training tasks.

Transfer Task

The detailed procedure of the transfer task (AB-AD task) is the same as in Exp.1.

Training Tasks

I adopted the procedure from Persson et al. (2007)'s study and Persson & Reuter-Lorenz (2008)'s study, with fewer trials of each task. A total of 160 trials of all 3 training tasks were performed, taking approximately 30 minutes.

In the letter recognition task, 32 trials were divided into two 16 trial blocks with a 1min rest period between blocks. At the beginning of each trial, four lowercase letters (target set) and a central fixation point were presented onscreen for 1.5 seconds, followed by a 3-second delay with only a fixation point appearing on the screen. A probe letter (uppercase) was then presented for 1.5 seconds. Participants were required to respond "yes" when the probe letter was a member of the target set of the current trial by pressing the left mouse button. Otherwise, the right mouse button was pressed to indicate a mismatch, and thus a "no" response. The inter-trial interval was 1.5 seconds, and no more than 2 successive trials required the same response; this was in order to avoid contaminating results from lazy-button-pressing participants. The total time for the letter recognition task was approximately 8 minutes.

A detailed description of the N-back task procedure has been provided previously in the material session. The total time for the N-back task was approximately 8 minutes.

In the verb generation task, sixty-four nouns were divided into two 32 trial blocks, with an intermediate rest period of 2 minutes. Each noun appeared for 4 seconds onscreen. Participants were required to press the left mouse button as quickly as possible when they had mentally generated a verb. 32 nouns were included in the high interference condition and 32 in the low interference condition, presented in random order. The whole verb generation task took approximately 8 minutes.

In between these 3 training tasks, two rest periods were filled with figure comparison tasks, taking approximately 3 minutes each.

Results

Data from 8 participants were excluded due to missing data for the final analyses. As in Exp.1, response accuracy (percentage of the correct recollection) and response time (mms) were analyzed via ANOVAs. The interaction was shown to be insignificant, F (1, 34) = 1.133, p = .15, partial η^2 =.032 (Figure 6). Despite this non-significance, I noted a rather apparent decrement of the response accuracy in the high depletion group with the high interference (experimental) condition only, when comparing before depletion to after depletion. Thus, I further conducted a paired sample t-test, revealing this decrement to be indeed significant, t (17) = 2.34, p = .01, d = .55. I also analyzed the response time of correct responses. Results indicated no significant main effect and interaction (for descriptive data see Table 4).

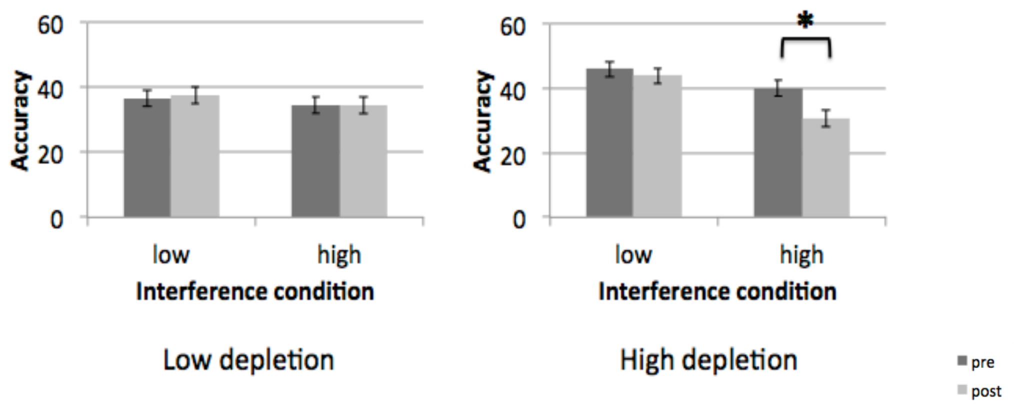


Figure 6. Response accuracy (percentage of correct responses) of the transfer task (AB-AD task), as a function of group (high vs. low depletion) and time (pre and post-depletion) of Exp.2.

Table 4
Response Times (mms) of the Correct Responses for the Transfer Task (AB, AD task)

Interference condition	Test phase	Depletion group	
		low	high
low	pre	2623.25 (426.77)	2907.22 (728.75)
	post	2924.81 (978.38)	2549.86 (415.91)
high	pre	3419.81 (1371.18)	3329.06 (1036.25)
	post	2816.39 (768.36)	2681.53 (585.46)

Discussion

Table 4 illustrates a trend in which participants performed faster after depletion compared with before depletion in 3 out of all 4 conditions (low interference condition after high depletion, high interference condition after both low and high depletion). This speed improvement trend ran counter to the expectation that participants' response time should be longer after depletion, indicating the depletion effect. However, this speed improvement could simply be due to practice with the AB-AD task. Moreover, I would argue that the response time in the current experiment could not be considered a proper measure in this regard. First, as demonstrated in Table 3, the response time baseline was highly scattered, indicating rather large individual differences. Second, while the current depletion procedure might work very well for some people, the response time may not be influenced much, as it can be easily affected by many other factors (e.g., motivation or one's own response behaviors). Therefore, I consider the response accuracy as a better measure to check the depletion effect. Additionally, the correlations between the response accuracy and the response time were not significant (Table 5). Thus, the significant decrement in response accuracy in the high interference condition in the high depletion group could not be due to speed improvement in response time, i.e., absence of the speed accuracy trade-off.

Table 5
Correlations between Response Accuracy and Response Time of Correct Responses (Exp.2)

Interference condition	Test phase	Depletion group	
		low	high
low	pre	.133	-.158
	post	-.381	-.352
high	pre	-.132	-.206
	post	-.177	-.054

The nature of Exp.2 was in fact a manipulation check testing whether the depletion effect worked under certain intensity. In turn, the effective depletion procedure would be applied on top of the HB memory paradigm. I need to clarify one point about the necessity of Exp.2 and the reasons for its design configuration. That is, I did not directly use the classic retroactive interference task (the modified AB-AD task validated from Exp.1) as a training task during the depletion procedure. Instead, several other tasks demanding resistance to interference were used. The AB-AD task was only used as a transfer task presented pre and post-depletion. Thus, in subsequent Exp.3, I determined not to apply the AB-AD task as a training task during depletion either. This was due to the material similarity between the AB-AD task and HB memory task. More specifically, both tasks involve semantic information paired with numbers and both require numeric responses. If I directly use the AB-AD task in between as a training task, participants' responses could be highly influenced by the numeric responses in the AB-AD task, both when they are asked to generate a numeric estimation in the HB OJ phase and when they are asked to recall their initial numeric estimation in the HB ROJ phase. Thus, the potential depletion effect of interference resistance ability obtained from the HB task could be confounded simply by depletion of repeated material (i.e., numerical) presentation. Instead, I used other interference tasks not involving numbers for the depletion procedure so that the expected depletion transfer effect on HB performance (in Exp.3) could reflect the authentic underlying relationship between interference and HB. This is precisely the value of Exp.2, in that I ensured the depletion procedure could be effectively transferred to the classic retroactive interference paradigm which I would argue highly resembles the HB memory paradigm.

Although the optimal result pattern (significant 3-way interaction) was not observed in this experiment, the depletion procedure could have worked, indicated by the results of the paired sample t-test. It was shown that participants performed significantly worse after depletion (percentage of correct responses M = 30.50, SE = 5.1) compared with before depletion (percentage of correct responses M = 40.26, SE = 5.2), only in the high interference condition and only after high depletion of interference resistance ability. Indeed, they did not show impaired performance after the low depletion procedure in the low interference condition in which minimal interference resistance ability was required. Thus, the deterioration of performance in the high interference condition after the high depletion process could not be due to fatigue caused by general cognitive ability performing one task successively, but rather a specific cognitive process involving interference resolution ability. As to the 3-way interaction, I presumably would need a bigger sample to reach a significant level. Thus, in the next experiment, a greater sample of participants were recruited to apply the depletion procedure validated in this experiment in addition to HB memory paradigm.

Experiment 3

In Experiment 3, the current depletion manipulation of interference resistance was applied to the memory HB paradigm. Specifically, the transfer task in Exp.3 is a HB memory task applying the same training tasks used in Exp.2 (i.e., a series of interference tasks). The purpose of Exp.3 is to investigate the main puzzle of the current study – what are the fundamental mechanisms underlying the HB memory paradigm. The underlying logic is that the magnitude of HB should become larger after participants' interference resistance ability is depleted, thus suggesting that people are normally hindsight biased due to their ineffective resistance to the retroactive interference induced by later presented information.

Method

Participants

A total of 61 students participated in Exp.3. Twenty-one were male. The participants ranged in age from 18 to 26 years (*M* = 21.09, *SD* = 3.00). Two were non-native German speakers. They were paid either 12 Euro or a 1.5-hour credit toward the completion of their Bachelor of Psychology course.

Design

The dependent variables in this experiment were HB task measures: the perfect recall percentage of ROJ and the proximity index of HB (see Pohl, 2007). I also conducted a manipulation test, with the dependent variable being the response accuracy of the AB-AD task.

The design was a 2 (item type) x 2 (group or depletion intensity) x 2 (depletion time) mixed factorial design with item type (whether the correct information was presented as feedback, i.e., experimental; or not, i.e., control condition for a particular question) and time (pre vs. post-depletion) as two within-subjects variables and depletion intensity (high vs. low depletion) as a between-subjects variable.

Materials

Three types of task were applied in this experiment: a transfer task, a training task and a filling task. The modified AB-AD task validated in Exp.1 was used as a task for manipulation

check; it was only presented once prior to and once following the depletion phase. As in Exp.2, a series of tasks involving interference resistance were used as training tasks for the depletion procedure. For the transfer task in Exp.2, an HB memory task was adopted.
A total of 88 almanac questions were presented to participants. Of these 88 questions, I developed 40 new question items. The other 48 questions were from previous HB studies undertaken by colleagues (Dr. Janina Hoffmann and Claudio Pichierri). An example of an HB item used in the current study is:

Wie viele Kapitel haben die drei "Herr der Ringe" Bände (ohne Anhänge) insgesmt?

_______ *Kapital*

All questions required numerical responses. The appropriate units were always presented. The answers were all two digit numbers ranging from 12 to 98, with no two digits being the same number. This was to avoid the possibility of mixing up numbers from the AB-AD task.
The questions were from various knowledge domains, including biology, history, literature, geography, physics and architecture. In order to increase the motivation of participants, I endeavored to use questions that were likely to be of interest or relevance to them. However, in each case, the exact answers to the questions would be unknown to the vast majority of participants.

In the OJ phase, all 88 items were presented to all participants (as shown in the above example). In the ROJ phase, all 88 items were divided into 2 sets: a pre-depletion set and a post-depletion set, with each set comprising 44 items. Within each set there were both experimental and control items. With experimental items, questions were presented with their correct answers (simultaneously shown on the screen), for example,

Wie viele Kapitel haben die drei "Herr der Ringe" Bände (ohne Anhänge) insgesmt?
Die richtige Antwort ist 62 Kapital.

Deine Antwort von Gestern war
_______ *Kapital*

With control items, questions were presented without the correct answers[1]:

Wie viele Kapitel haben die drei "Herr der Ringe" Bände (ohne Anhänge) insgesmt?

Deine Antwort von Gestern war

_______ *Kapital*

Procedure

As in Exp.2, there were 3 phases in Exp.3: pre and post-depletion phases and the depletion phase. However, considering the nature of HB memory tasks, i.e., there should be a retention interval between OJ and ROJ phases, we conducted the experiment across 2 days (Figure 7).

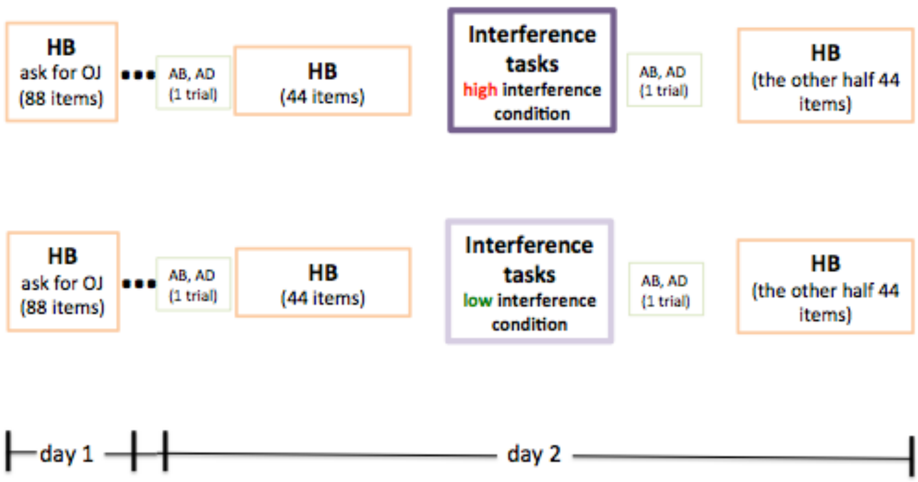


Figure 7. Schematic depiction of the procedure of Exp.3. From left to right, task applied pre-depletion (both in session 1 and in session 2) during depletion, and post-depletion. The first and second rows illustrate the high depletion group and the low depletion group, respectively.

Participants were randomly assigned into two groups: the high depletion group and the low depletion group. They were told that they would take part in a general knowledge test at the point of recruitment and again when given instructions at the beginning of the experiment.

[1] Each item was presented in *either* the experimental condition *or* in the control condition. No single item was presented in dual condition types as per this example, which is repeated in both forms here only to illustrate how questions were presented in the control condition.

They were also to give numerical estimations as best they could, given they would probably not know the exact answers to most of the questions. Thus, they were not informed about the ultimate purpose of the experiment, the HB paradigm or the later depletion procedure.

In day 1, all participants were provided with the HB questionnaire including all 88 items. Items were presented in a randomised order to each participant. For each item, a question and answer box were shown in the same screenshot so that participants would be able to review the question any time while they were giving their OJs. The process of answering questions was self-paced. Participants were asked to press the return key after they had given their estimation for each question. The next question then appeared onscreen. The whole OJ phase (session 1) took approximately 30 minutes for each participant to complete.

In day 2, each participant was presented with half of the 88 items and asked for their ROJs after one trial of the AB-AD task, followed by the depletion procedure. Of the 44 items before depletion (i.e., the first set, items 1-44), half were experimental items with correct answers presented simultaneously with ROJs requests, while the other half were control items with no provision of correct answers.

The depletion procedure was the same as that in Exp.2.

After depletion, one trial of the AB-AD task was given prior to the other 44 items being presented to participants (these being items 45-88, again half experimental, half control), who were again asked for their ROJs. As such, all items were partially counterbalanced across participants within each depletion group. Questions were identical to those in the first session, however, presented in a randomised order within each set.

For the ROJ phase, participants were told they would be presented with the same general knowledge questions from the previous day. Their task was to remember their answers to each question as accurately as they could. This was emphasized in the instructions using bold font. They were told that for some questions, the correct answers would be given under the questions, but I (as experimenter) was interested in how well participants were able to remember their previous estimations. Thus, the best answers were those closest to the answers provided the day before, irrespective of what the correct answers were. The second session took participants approximately 50 to 60 minutes to complete.

Results

All participants returned on the second day for the depletion procedure and HB ROJ task. Data from 8 participants was excluded due to too much missing data for final analyses. For the two experiments involving HB (Exp.3 and Exp.4), results would be reported from traditional AVOVAs of the perfect recall percentage and HB magnitude using the proximity index (Pohl, 1992) and results obtained using multinomial modeling (Erdfelder & Buchner, 1998). Comparisons across these two methods would be made and discussed. Before conducting a detailed analyse, data (both OJs and ROJs) were sorted as follows:

1. All missing data was excluded (440 cases out of 4664 = 53 participants × 88 questions).
2. Some participants were confused by a particular unit. Taking the item " Wie alt wurde Goethe?" as an example, participants were required to give the last 2 digit number (i.e., 68) while the first 2 digit number "19____" was provided. Some participants overlooked the already given "19" and wrote down the whole 4 digit response (e.g., 1962). Such clear cases of input error were corrected rather than being excluded before data analyses.
3. Extreme data was excluded. The exclusion criterion for the current experiments comprised those estimates falling outside the median plus or minus 4 times the interquartile range for each item (Tukey, 1977). As a result, 513 data cases were excluded out of all 8448 cases, this being a 6.07% exclusion rate.
4. The data was then standardized so that comparisons could be conducted across questions that were scaled differently (i.e., with different units).

ANOVAs

I conducted ANOVAs with four variables of Exp.3, namely response accuracy of the interference task (AB-AD task) and 3 measures of HB task: the quality of original judgments, the quality of recall of original judgments, and the shift in recalled judgments.

Response Accuracy of the Interference Task (AB-AD task)

The response accuracy, i.e., percentage of correct recall, was analyzed using ANOVAs. The results showed that response accuracy 1) was significantly lower with the high interference condition compared to the low interference condition across two depletion groups, $F\,(1, 51) = 15.306$, $p <. 01$, partial $\eta^2 =. 25$; and 2) was not significantly different between pre and post-depletion conditions $F\,(1, 51) = 1.102$, $p = .32$, partial $\eta^2 =. 189$.

Hindsight Bias measure: Quality of Original Judgments

Results showed that although the difference in the percentage of correct OJs between the high depletion group and low depletion group was readily discernible, it did not reach significance, $F\,(1, 51) = 1.998$, $p = .06$, partial η^2 =. 104. Furthermore, in total, 1.99% of the original judgments equaled the correct answers across the two groups. These correct original estimations were excluded for the further analyses, as HB is not defined for them.

Hindsight Bias measure: Quality of Recall of the Original Judgments

I first analyzed the percentage of perfect ROJs, i.e., cases in which participants gave exactly the same ROJs as their OJs. Results showed that there was a significant main effect based on group, $F\,(1, 51) = 9.306$, $p < .01$, partial $\eta^2 = .154$ (Figure 8). Due to this significance, and the approaching significance of the difference in correct original judgment percentages between two randomly assigned groups, I further tested the distribution normality of the two groups respectively. Results for the Kolmogorov-Smirnov goodness-of-fit test for normality (Field, 2007) showed that the distribution of both groups did not deviate significantly from a normal distribution ($D = .084$, $p = .211$) (for histogram see Figure 9). Thus, the normality of the current sample was met.

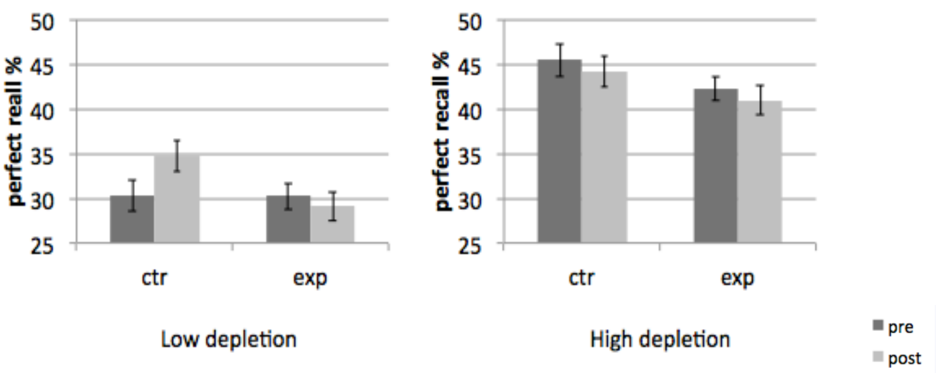


Figure 8. Response accuracy (percentage of ROJ = OJ) of the HB task, as a function of group (high vs. low depletion), time (pre and post-depletion) and item type (ctr: no feedback vs. exp: with feedback) of Exp.3.

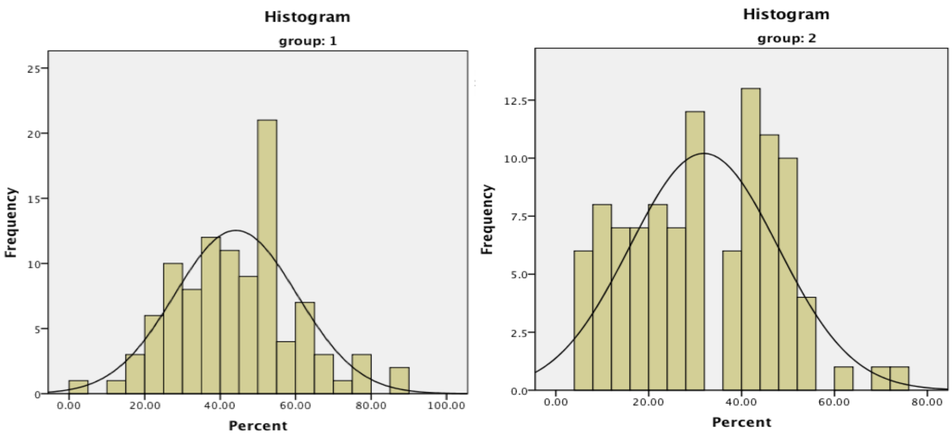


Figure 9. Histograms of the perfect recollections of the original judgments of the high depletion group (left) and low depletion group (right) of Exp.3.

In regard to the 3-way interaction of interest to us, it was found to be not significant, F (1, 51) = 1.123, p = .29, partial η^2 = .022.

Hindsight Bias measure: Mean Shift of the Recalled Judgments

The perfectly recalled OJ and ROJ pairs were excluded for the shift toward CJ measure. The 3-way ANOVA yielded an overall main effect of item type, F (1, 51) = 17.98, $p < .01$. The mean shift values were 0.35(±0.04) for experimental items, and only 0.02(±0.05) for control items, indicating an overall robust HB effect. However, regarding the variable depletion group, ANOVA results showed that there was neither a significant main effect of group, F (1, 51) = 1.149, p = .29, partial η^2 = .022, nor a significant 3-way interaction, F (1, 51) = 1.205, p = .28, partial η^2 = .023. Furthermore, paired sample t-tests comparing the delta-z mean shift between pre and post-depletion procedures with the experimental groups also showed no significant difference, t (26) = 1.57, p = .06, d = .31, 1-ß = .47; t (25) = 1.29, p = .10, d = .35, 1-ß = .54 for the low depletion group and the high depletion group, respectively (Figure 10).

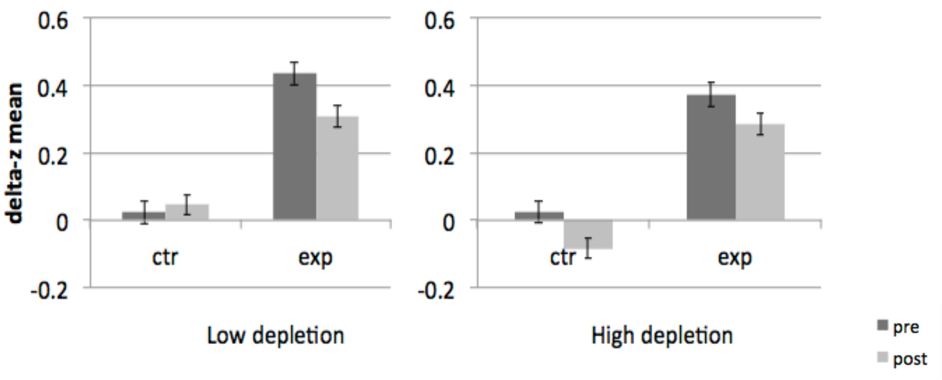


Figure 10. Mean shift of the recall of original judgments toward the correct answers from the incorrectly recalled pairs of the HB task, as a function of group (high vs. low depletion), time (pre and post-depletion) and item type (ctr: no feedback vs. exp: with feedback) of Exp.3.

To examine the extent to which (if at all) participants' bias toward the correct answers is associated with their perfect recollections of the original estimates, I further computed the correlations between the two measures of HB (Table 6).

Table 6
Correlations between Perfect Recall Percentage and Mean Shift of Recalled Judgments of Incorrectly Recalled Pairs, both High Depletion Group and Low Depletion Group (the latter ones in brackets) (Exp.3)

Variables	Δz_pre_ctrl.	Δz_pre_exp.	Δz_post_ctrl.	Δz_post_exp.
Perf.recall_pre_ctrl.	.194 (-.170)	.243 (.351)	.228 (.019)	.110 (.196)
Perf.recall_pre_exp.	.125 (-.155)	.351 (.368)	.271 (-.066)	.059 (.202)
Perf.recall_post_ctrl.	-.133 (.025)	-.015 (.264)	.389*(-.011)	-.152 (.107)
Perf.recall_post_exp.	.083 (-.244)	.085 (.200)	.161 (-.134)	.031 (.083)

$N = 26$, * $p < .05$ ($N = 27$)

Results showed that the correlations between the two measures of HB were mostly not significant, with the exception being the control condition after depletion ($r = .389$, $p < .05$). Moreover, I conducted the correlation analyses with the average across experimental and control conditions and with the difference between the two item type conditions, across two depletion intensity groups. No single correlation was significant.

Moreover, since the HB items in this experiment were only partly counterbalanced across time (pre vs. post-depletion), I further tested the item set effect. Results showed that the difference in participants' original estimates and the correct judgments between two item sets (pre-set 1-44 vs. post-set 45-88) was not significant, t (51) = 2.33, $p < .01$. Thus, there was no significant difference in material difficulty between two item sets across both groups. Consequently, the difference between pre and post-depletion for both groups could not be due to differences in material difficulty, albeit without full counterbalancing.

Multinomial Model-Based Analyses

For all model tests and parameter estimations according to HB 13 (Erdfelder & Buchner, 1998, discussed in introduction session) and its sub-models in this study, I excluded the 1.99% pairs in which the original estimations equaled the correct answers. The raw frequencies of all 10 rank orders of OJ, CJ and ROJ for experimental and control conditions across all participants within each depletion group are provided in Appendix A. Based on these frequencies, I conducted the multinomial modeling analyses using multiTree, a java-based computer program by Moshagen (2010). First, the goodness of fit of the model was evaluated for the whole data set (i.e., 2 depletion intensity groups x 2 item type conditions x 2 depletion time conditions, thus resulting in 8 trees), using the likelihood ratio chi-square statistic G^2 (with $df = 20$). The G^2 (20) value was 18.89, with $p = 0.53$, indicating a satisfactory model fit.

Table 7 shows the results derived from testing the presence of recollection bias, reconstruction bias, and CJ adoption, as well as the null hypotheses regarding these tests for each depletion group both before and after depletion. To test the presence of recollection bias, I compared a model variant in which the r_C and r_E parameters were free to vary with a restricted model in which the two parameters were equal. Should the significance of the comparison be present ($p < .01$), I would in turn conclude that the restricted model assuming the two parameters were equal did not fit the data, and therefore the two parameters would not differ significantly, and vice versa. Likewise, to test the presence of reconstruction bias and CJ adoption, I compared a model variant in which b and y were set to free with a restricted model in which the two parameters were constant at zero. If the comparison was found to be significant, I would then conclude that the parameters were different from zero, and vice versa.

Table 7

Results of Testing for the Presence of Recollection Bias, Reconstruction Bias, and Correct-Judgment Adoption in Each Depletion Group, Pre and Post-depletion Respectively (Exp.3)

		Pre-depletion		Post-depletion	
Type of bias	Depletion group	Null hypothesis	$\Delta G^2(df = 1)$	Null hypothesis	$\Delta G^2(df = 1)$
Recollection bias	high	$r_{C1} = r_{E1}$	2.06	$r_{C2} = r_{E2}$	0.84
	low	$r_{C3} = r_{E3}$	0.04	$r_{C4} = r_{E4}$	4.20
Reconstruction bias	high	$b_1 = 0$	34.94*	$b_2 = 0$	22.45*
	low	$b_3 = 0$	18.93*	$b_4 = 0$	19.80*
CJ adoption	high	$c_1 = 0$	9.61*	$c_2 = 0$	0.17
	low	$c_3 = 0$	0.89	$c_4 = 0$	3.87

Note. ΔG^2 = decrement in the likelihood ratio chi-square goodness-of-fit statistic compared with the unrestricted model; r_C = probability of participants' correctly recalling the original judgment (control items); r_E = probability of participants' correctly recalling the original judgment (experimental items); b = probability of a biased reconstruction given a failure to recall the original judgment; c = probability of adopting the correct judgment, with 1, 2, 3 and 4 denoting high depletion group pre and post-depletion; low depletion group pre and post-depletion time condition respectively.

* $p < .001$

Table 8 presents the pre and post-depletion comparisons of each depletion intensity group for the parameters of interest. To test whether the parameter differences between the depletion time conditions were significant for each depletion intensity group, I compared a model for two depletion time conditions in which all parameters were allowed to differ between these two conditions with a restricted model in which the parameters between the two depletion time conditions were set as equal. If the comparison was significant, I would then conclude that the restricted model assuming two equal parameters of the two depletion time conditions (pre and post) did not fit the data anymore, and hence the two parameters would be significantly different, and vice versa.

Table 8

ΔG^2 values of Testing Pre vs. Post-depletion Differences in Multinomial-Model Parameters (Exp.3)

Comparison conditions	r_C	r_E	b	c
High depl. pre vs. post-	2.25	1.06	0.83	0.65
Low depl. pre vs. post-	1.36	1.15	0.01	2.23

Note. ΔG^2 = decrement in the likelihood ratio chi-square goodness-of-fit statistic ($df = 1$) compared with the unrestricted model; r_C = probability of participants' correctly recalling the original judgment (control items); r_E = probability of participants' correctly recalling the original judgment (experimental items); b = probability of a biased reconstruction given a failure to recall the original judgment; c = probability of adopting the correct judgment.
* $p < .001$

Figure 11 shows the estimates of the parameters of interest, i.e., recollection bias parameters, reconstruction bias parameters and the CJ adoption parameters for both depletion intensity groups and both depletion time conditions.

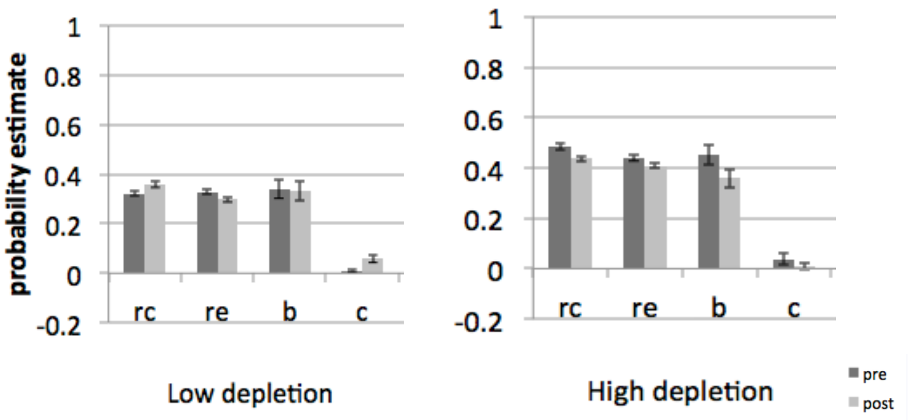


Figure 11. Multinomial model parameter estimates for depletion group of Exp.3. Recollection bias is the difference between correct recollection in the control condition (r_C) and correct recollection in the experimental condition (r_E). Error bars indicate 95% confidence intervals.

Recollection parameters

As shown in Table 7, overall there were no recollection biases either before or after depletion, irrespective of the depletion group (high vs. low). The exact estimates of parameter r_C (with their 95% confidence intervals) were 0.48 (±0.03), 0.44 (±0.02), 0.32 (±0.02) and 0.36 (±0.02) for the high depletion group pre and post-depletion and for the low depletion group pre and post-depletion condition respectively; and the corresponding r_E values were 0.44 (±0.02), 0.41 (±0.02), 0.33 (±0.02) and 0.30 (±0.02). Moreover, for both depletion groups there was no significant difference in recollection bias parameters before and after depletion across control and experimental items, as indicated in Table 8.

Reconstruction bias

As indicated by the reconstruction parameter *b*, results from Table 7 show that reconstruction bias exists in all conditions. The exact *b* values (with their 95% confidence intervals) were 0.45 (±0.07), 0.36 (±0.08), 0.34 (±0.07) and 0.33 (±0.08) for the high depletion group in pre and post-depletion conditions, and for the low depletion group in the respective depletion time conditions. Unfortunately however, the difference in reconstruction bias between before and after depletion for both groups was not significant (as shown in Table 8 and Figure 11).

CJ adoption

The *c* parameters denoting the correct judgment adoption were found to exist in only one condition, this being the high depletion group in the pre-depletion condition (see Table 7).

This CJ adoption process disappeared after depletion with the high depletion group. The exact *c* values were 0.04 (±0.04), 0.01 (±0.03), 0.01 (±0.01) and 0.06 (±0.03) for the high depletion group in the pre and post-depletion conditions, and for the low depletion group in these conditions respectively. In regard to differences in the CJ adoption parameter between before and after depletion, these were not significant for either group (see Table 8 and Figure 11).

Discussion

To test whether the depletion procedure worked as effectively as in Exp.2, I included one trial AB-AD task before and after depletion in the current experiment. Indicated by the ANOVAS, the significant lower response accuracy of the high interference condition compared to that of the low interference condition affirmed the validity of the AB-AD task (i.e., the AB-AD task proved suitable as an interference task). However, the response accuracy was then no significant difference when comparing between pre- and post-depletion conditions. This unfortunately revealed that the depletion procedure did not work well in Exp.3, which could however be due to the fact that only one trial AB-AD task was presented to participants.

As to the quality of OJs, I compared the percentage of correct responses during the OJ phase across two groups (the high depletion group and the low depletion group) before depletion. The non-significant difference of correct OJ estimates confirmed that the knowledge about the HB items between two groups did not differ. Thus, randomization worked well in the current experiment.

Regarding the quality of recall of OJs, results showed that participants in the high depletion group remembered their OJs significantly better than their counterparts in the low depletion group, regardless of whether or not they had gone through the depletion procedure. Overall, the pattern of the perfect recollection percentage between pre- and post-depletion was rather consistent with the pattern of the interference task measure (response accuracy). While no significant depletion effect was revealed by either measure, the descriptive pattern was similarly decreasing.

Despite that the two groups (high and low depletion) were significantly different in terms of the perfect recall of OJs, the shift toward CJs was statistically equal. This rather intriguing pattern showed that participants' biased reconstruction (indicated by the delta-*z* measure) was not due to their memory deficit (indicated by the perfect recollection measure). Further correlational analyses showed in Table 6 further provided good empirical support for the assertion that perfect recall and bias of imperfectly recalled judgments should be kept apart as two distinct measures (Dehn & Erdfelder, 1998; Pohl & Haracic, 2005).

The most stimulating result observed here was a decreasing trend both in the perfect recall percentage and in the secondary shift toward the correct judgments of incorrectly recalled items after depletion. As discussed above, participants' interference resistance ability was expected to be fatigued through the depletion procedure, which was also demonstrated in Exp.2. Furthermore, the theoretical conjecture that I formed was that the memory type of HB could in fact be nothing more than a retroactive interference paradigm. It then follows that participants' HB magnitude would increase after specific depletion, since their ability to resist interference had been exhausted. However, the decreasing trends seen in this experiment ran counter to what I had anticipated, while not reaching the level of statistical significance.

Several factors lead us to view these results as rather meaningful. First, although ANOVAs revealed no significant differences in both perfect recall percentage and the shift toward correct answers for either group (high and low depletion), the differences in mean of delta-*z* between pre and post-depletion for both groups were already approaching significant, $p_1 = .06$, $p_2 = .10$. Second, the pattern of differences in HB magnitude was quite consistent across the two groups (low and high depletion), and could not be dismissed as an accidental result. Third, power analyses computed with the G*POWER program (Faul, Erdfelder, Lang, & Buchner, 2007) showed that the non-significant results could be due to low power (0.327 and 0.237 respectively for low and high depletion group). Taken as a whole, attempts to interpret the results were still necessary.

The multinomial model-based analyses showed that no significant difference in recollection bias for both depletion groups, across all conditions (pre- vs. post-depletion and control vs. experimental condition). So was it the case with reconstruction bias. However, CJ adoption process disappeared after depletion with high depletion group, implying that participants were less convinced that the present correct answers exactly matched their original estimates after depletion, as the *c* parameter only existed in pre-depletion condition of high depletion group.

In summary, results of the conventional ANOVAs showed that after depletion there were no significant differences in perfect recall percentages or the proximity index for both depletion groups (see Figure 8 and Figure 10). These results for the two HB magnitude measures suggest that the depletion manipulation intended to fatigue participants' interference resistance process did not work as expected. That is to say, participants did not become more biased by the later presented correct judgments after they were depleted by their interference resistance ability. This is in line with what the multinomial modeling analyses revealed. Specifically, the reconstruction bias parameter *b* did not differ significantly after depletion for either group. Nevertheless, it should be noted that the trend direction of critical parameters (delta-*z* mean in ANOVAs and *b* in multinomial modeling analyses) was rather congruous, i.e., decreasing rather than increasing. This decreasing trajectory was contrary to the expectations before the experiment was conducted, which prompted the current search for an appropriate interpretation; in turn, I proposed a new experiment with a refined manipulation to further test the proposition.

As demonstrated in the results, it could not be the case that the specific interference resistance ability was depleted as there was no difference in decrement between the high and low depletion groups. Clearly a specific "other" feature common to the two groups may have been diminished through the depletion procedure. The next step then was to detect and identify this underlying feature.

Experiment 4

Previous research has suggested that typical memory HB paradigm may be no more than a retroactive interference process. However, results of Exp.3, both from the ANOVAs using the conventional HB measure and from more sophisticated multinomial modeling analyses, indicated that there must be MORE than just interference processes at work in HB. Furthermore, equipped with the commonly predisposed interference resistance mechanism, people attempt to counteract HB in normal cases. That is to say, if one devotes sufficient effort to interference resistance activity, there should be a decrease in HB magnitude. Linked to the more specific results of Exp.2, that is decreasing rather than increasing HB magnitude after depletion, I can expect to find a discrete process involved that stimulates the occurrence of HB in normal cases. Hence, HB ought to decrease after the cognitive resources demanded by such process deteriorate. Review of the theoretical discussion in previous studies suggests the concept of knowledge-updating may fit with the current conjecture here. The knowledge-updating notion was embedded in more than one theory or model explaining HB, examples being the immediate assimilation account by Fischhoff (1975), the SARA model by Pohl et al. (2003) and the RAFT model by Hoffrage et al. (2000). However, the immediate association account argues for an automatic process right after the correct judgment is presented. Consequently, HB could not become depleted due to this automaticity, assuming that depletion acts only on controlled activities, but not on automatic ones in which minimal cognitive resources are required (Schmeichel et al., 2003). Both the SARA and RAFT models also take the knowledge-updating process into consideration, however no specific clarification is made as to whether this process is more automatic or controlled. Furthermore, if it is the latter case, how could it be controlled with greater precision? In relation to the plausibility effect (for discussion see introduction), I hypothesize a dual process account of HB in which both a certain knowledge-updating mechanism and interference resistance activity are fundamental to the HB memory paradigm. Both components are controlled, rather than automatic. The former is adaptive in nature, in the sense that people endorse more knowledge-updating activity only when the later presented information is correct.

In Exp.4, I test this hypothesis using the depletion procedure in combination with the plausibility manipulation of subsequently presented information (i.e., CJ). More specifically, in the HB task, information presented later would be labeled as either the “correct answer” or the “wrong answer” of a question. As a result, there are 2 experimental conditions and 1 control condition in the HB task. Methodically, in order to get comparable results across experiments, I should in theory have the same number of HB items within each item type condition, meaning 3 x 44 = 132 items per condition. However, empirically this quota itself would lead to confusion and fatigue for participants. As a result, I used the same number of HB items as in Exp.3, separating them into 3 conditions. What is more, Exp.4 compelled us to recruit additional participants. Experiment 3 had comprised 88 items x 53 participants / 2 conditions (experimental vs. control) = 2332 cases per condition. In order to keep the overall frequency of cases within each condition constant, I would need at least 2332 cases x 3 conditions (2 experimental vs. control) ÷ 88 items = 80 participants for Exp.4.

With this labeling manipulation, expectations in terms HB magnitude would be as follows:

5. There should be a plausibility effect in which HB is significantly larger when the later presented information is labeled “correct answer” as opposed to when it is labeled “wrong answer”.

6. After depletion, HB should decrease in the “correct answer” condition, as the adaptive knowledge-updating behavior leading to increased HB in the normal condition would be exhausted.

7. In contrast, after depletion, HB should increase in the condition labeled “wrong answer”, as the interference resistance ability that counters HB in the normal situation will have been eroded.

8. As a consequence, I anticipate the pre-depletion plausibility effect to disappear after depletion.

Method

Participants

A total of 118 students participated in Exp.4, of which forty-eight were male. The participants ranged in age from 19 to 29 years (M = 23.01, SD = 2.37). All but one were native German speakers. They were paid either 12 Euro or a 1.5-hour credit toward their Bachelor in Psychology course.

Design

The dependent variables for ANOVAs in Exp.4 were HB task measures: the perfect recall percentage of ROJ and the proximity index of HB. As with Exp.3, I conducted a manipulation test with the dependent variable being the response accuracy of the AB-AD task. The design of ANOVAs was a 3 (item type) x 2 (group or depletion intensity) x 2 (depletion time) mixed factorial design with item type (whether the correct information was presented as feedback or not, and if yes, how was it labeled, i.e., the control condition and 2 experimental conditions: one labeled “correct answer” and the other labeled “wrong answer”) and time (pre vs. post-depletion condition) as two within-subjects variables, and depletion intensity (high vs.

low depletion group) as a between-subjects variable. Finally, I performed multinomial modeling analyses for data generated in Exp.4.

Materials

As in Exp.3, three types of task were applied in this experiment: a transfer task, a training task and a filling task. All materials for this experiment were the same as in Exp.3, except for one more experimental manipulation in the HB task – labeling of the later presented information. More specifically, the same 88 almanac questions were used for the OJ phase. With the ROJ phase on the second day, one more experimental condition was employed. Control items were identical to those in Exp.3. An example of an item labeled in the“correct answer” condition would be:

Wie viele Kapitel haben die drei "Herr der Ringe" Bände (ohne Anhänge) insgesmt?

*Die **richtige** Antwort ist 62 Kapital.*

Deine Antwort von Gestern war

_______ *Kapital*

By comparison, this same example labeled in the “wrong answer” condition would be:

Wie viele Kapitel haben die drei "Herr der Ringe" Bände (ohne Anhänge) insgesmt?

*Die **falsche** Antwort ist 62 Kapital.*

Deine Antwort von Gestern war

_______ *Kapital*

As illustrated above, the same numbers (answers) were presented, only with different labeling, these being emphasized in bold font and underline. Although the same HB questions were employed in Exp.4 as in Exp.3, the items remain valid. The new experiment was conducted one semester after the third experiment, meaning that it is improbably participants could recall data. What is more, as most participants took part in the experiments in order to gain one-off credits for their bachelor courses, there was no incentive to take part a second time; as such, it is unlikely *any* participants were involved in both trials.

Procedure

The detailed experimental procedure for Exp.4 (see Figure 12) was the same as Exp.3, with the following 2 exceptions.

1. In the HB ROJ phase on the second day, the 44 HB items applied pre and post-depletion were further split into 3 conditions, rather than 2 conditions in Exp.3. More specifically, 14 questions were presented as control items in which no feedback was provided at all during ROJ. The remaining 30 questions were divided into two groups, with half (15) being presented with the correct answers to the questions, which were labeled "correct answer"; while the other half were also accompanied with the correct answers, they were misleadingly labeled "wrong answer". Once more, participants were informed at the beginning of the ROJ phase that their next task was to recall as accurately as possible their answers from the preceding day. They were told that for some questions, the correct answers would be provided, while for other questions, wrong answers would be presented.
2. To further strengthen the intensity of the depletion procedure, I added 5 more trials of AB-AD task as the training task. Participants in the high depletion group performed the task only in the high interference condition, i.e., the AB-AD condition. Participants in the low depletion group performed the AB-only condition.

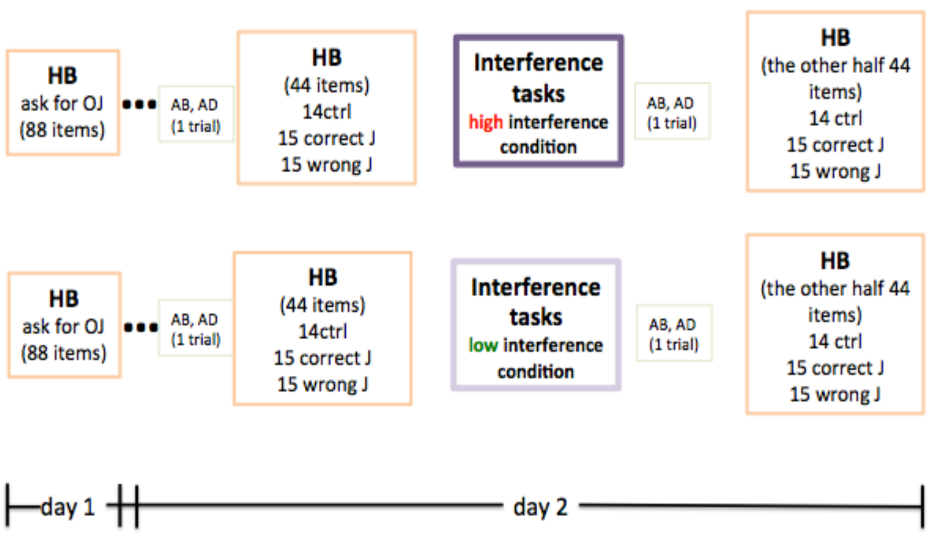


Figure 12. Schematic depiction of the procedure of Exp.4. From left to right, task applied pre-depletion (both in session 1 and in session 2), during depletion, and post-depletion. The first and second rows illustrate the high depletion group and low depletion group, respectively.

Results

All participants returned on the second day for the depletion procedure and the HB ROJ task. Again data from 8 participants was excluded due to too much missing data for final analyses. As with Exp.3, results from traditional ANOVAs of the perfect recall percentage and HB magnitude using the proximity index (Pohl, 1992) and results obtained using multinomial modeling (Erdfelder & Buchner, 1998) with this experiment are reported below. Comparisons across these two methods are also made and discussed. Data analyses were conducted in the same manner as in Exp.3.

Before conducting detailed analyses, data (both OJs and ROJs) were sorted as follows:

1. All missing data was excluded (616 cases out of 9680 = 110 participants × 88 questions).
2. Units that were confused by the participants were corrected before data analyses as in Exp.3.
3. Extreme data was excluded. I excluded 825 data cases out of all 18128 cases, this being a 4.25% exclusion rate.
4. The data was then standardized so that comparisons could be conducted across questions that were scaled differently (i.e., with different units).

With the *F* tests (alpha level of 0.01), the current sample size (110) was sufficient to yield a power of .99 to detect the depletion effects of a medium size (*f* = .25; see Cohen, 1988), computed with the G*POWER program (Faul et al., 2007).

ANOVAs

I conducted ANOVAs with four variables of Exp.4, namely, response accuracy of the interference task (AB-AD task) and 3 measures of HB task: the quality of original judgments, the quality of recall of the original judgments, and the shift of the recalled judgments.

Response Accuracy of the Interference Task (AB-AD task)

To test whether the depletion procedure worked as effectively as in Exp.2 and 3, I included one trial AB-AD task before and after depletion in the current experiment. The response accuracy, i.e., percentage of correct recall, was analyzed using ANOVAs. Results showed that the response accuracy 1) was significantly lower in the high interference condition compared with the low interference condition across two depletion groups, F (1, 108) = 18.06, p <. 01, partial η^2 =. 21, and 2) decreased significantly in the post-depletion condition compared to the pre-depletion condition, F (1, 108) = 12.20, p < .01, partial η^2 =. 33.

Hindsight Bias measure: Quality of Original Judgments

The percentage of correct responses during the OJ phase before depletion was compared for the low depletion and high depletion groups. The difference in the percentage of correct OJs between the two groups was not significant, F (1, 108) = 13.88, p <.01, partial η^2 =. 24. In total, 1.53% of the original judgments equaled the correct answers across two groups. These correct original estimations were excluded from further ANOVAs.

Hindsight Bias measure: Quality of Recall of the Original Judgments

The perfect recall percentage was first analyzed; that is, cases in which participants gave exactly the same ROJs as their OJs. Unlike Exp.3, results of the current experiment showed there was no significant main effect of group, F (1, 108) = 1.21, p = .22, partial η^2 = .014 (Figure 13).

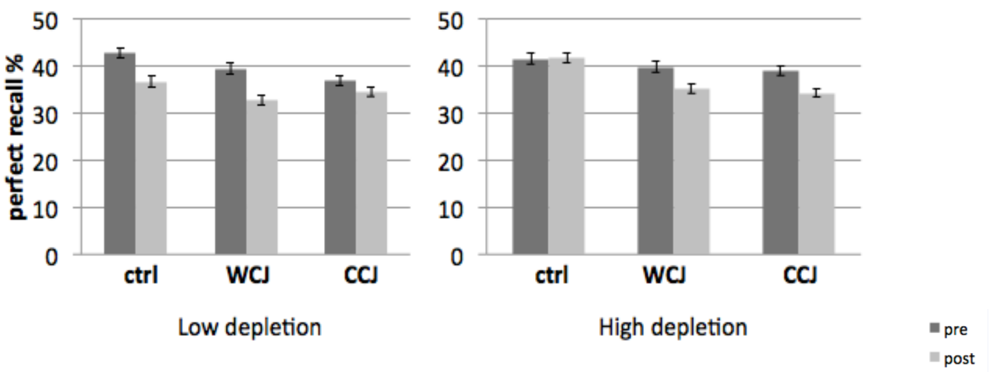


Figure 13. Response accuracy (percentage of ROJ = OJ) of the HB task, as a function of group (high vs. low depletion), time (pre and post-depletion) and item type (ctrl: no feedback, WCJ: labeling as "wrong answer" vs. CCJ: labeling as "correct answer") of Exp.4.

As shown in Figure 13, there was a significant main effect of item type (no feedback, labeling as "correct answer" and labeling as "wrong answer"), F (2, 108) = 8.82, p < .01, partial η^2 = .075 across two groups. Further paired sample t-tests demonstrated a significant decrease in perfect recall percentage in both the "correct answer" and "wrong answer" conditions as compared to the control conditions, t (54) = 1.33, $p < .05$ and t (54) = 1.46, $p < .05$; however, this was only the case in the high depletion group after the depletion procedure. No such significant decrease was found in the low depletion group either before or after depletion.

Another significant main effect was time (pre vs. post-depletion), F (2, 108) = 17.01, $p < .01$, partial $\eta^2 = .136$. Further paired sample t-tests showed that a significant decrease existed only with items labeled "correct answer" after depletion compared with before depletion, and then only in the low depletion group, t (54) = 2.71, $p < .05$. No significant interaction was found with the perfect recall percentage measure.

Hindsight Bias measure: Mean Shift of the Recalled Judgments

As in Exp.3, I excluded all the perfect ROJ and OJ pairs. The remaining pairs provided an appropriate means of determining the degree to which the ROJs were biased away from the OJs toward the CJs.

As with Exp.3, the 3-way ANOVA also yielded an overall main effect of item type, F (2, 108) = 15.73, $p < .01$. The mean shift values were 0.42(±0.02) for experimental items, and only 0.03(±0.03) for control items, indicating an overall robust HB effect. Moreover, the item type by depletion time interaction that I am principally interested in was shown to be significant, F (2, 108) = 8.13, $p < .01$, partial η^2 =.07 (Figure 14).

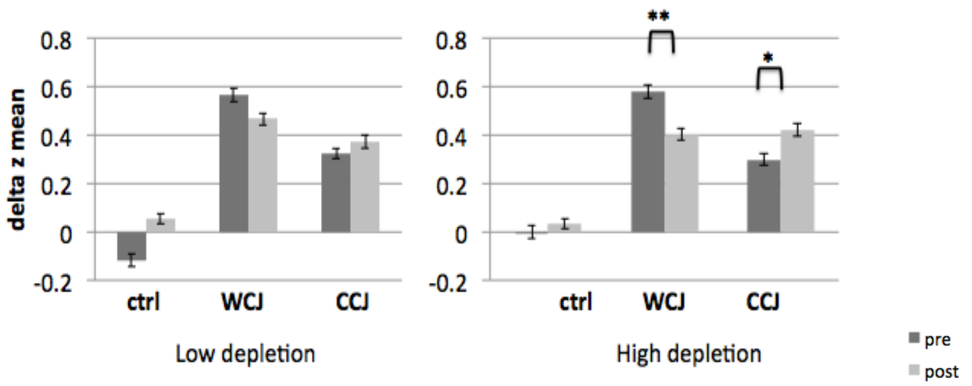


Figure 14. The mean shift of the recall of the original judgments toward the correct answers from the incorrectly recalled pairs of HB task, as a function of group (low vs. high depletion), time (pre and post-depletion) and item type (ctrl: no feedback, WCJ: labeling as "wrong answer" vs. CCJ: labeling as "correct answer") of Exp.4.

Figure 14 shows the results of further paired sample t-tests which reveal that after depletion the delta-z mean significantly decreased in the condition labeled "wrong answer" and significantly increased in the condition labeled "correct answer", but only in the high depletion group, *t* (54) = 2.50, $p < .01$, d = .61, 1-β = .27, *t* (54) = -1.67, $p < .01$, d = .55, 1-β = .24, respectively.

Moreover, the response time was also checked (more discussion about the reason for this step see discussion). The means of response time within each condition is shown in Table 9.

Table 9
Response Times (mms) of the Incorrectly Recalled Pairs for the HB Task of Exp.4 (ctrl: no feedback condition, CCJ: condition labeled correct answer, WCJ: condition labeled wrong answer)

Depletion condition	**Low depletion group**		
	Ctrl.	CCJ	WCJ
Pre-	3678.92 (239.77)	3888.32 (245.45)	3721.53 (213.44)
Post-	3421.98 (256.21)	3698.91 (258.41)	3622.19 (229.10)
	High depletion group		
	Ctrl.	CCJ	WCJ
Pre-	3559.21(199.87)	2991.69 (317.69)	3315.00 (290.88)
Post-	3319.80 (231.98)	3017.22 (222.15)	3169.54 (229.44)

Multinomial Model-Based Analyses

For all model tests and parameter estimations according to HB 13 (Erdfelder & Buchner, 1998, discussed in introduction session) and its sub-models in this study, I excluded the 1.53% pairs in which the original estimations equaled the correct answers. The raw frequencies of all 10 rank orders of OJ, CJ and ROJ for experimental and control conditions across all participants within each depletion group were provided in Appendix B. Based on these frequencies, I conducted multinomial modeling analyses using the multiTree (Moshagen, 2010). The goodness of fit of the model was first evaluated for the whole data set (i.e., 2 depletion intensity groups x 3 item type conditions x 2 depletion time conditions, thus resulting in 12 trees), using the likelihood ratio chi-square statistic G^2 (with $df = 32$). The G^2

(32) value was 46.02, with $p < .01$, indicating a non-model fit. Moreover, results showed that there were several negative variance estimates and standard errors could not be computed. Thus I checked the identifiability of the model using repeated analyses 30 times. Results showed that the model was not identifiable. Outcomes demonstrated on the parameters panel of multiTree revealed that the CJ adoption parameters *c* were the items with potential problems. Moreover, the theoretical implication of the parameter, that is, that participants may be convinced the presented feedback was identical to their first estimation, could make little sense in the context of the current experiment. To elaborate here, if the presented feedback was constantly and with certainty the correct answer to given questions, there is a probability that participants might adopt the feedback as their own previous estimation, as in Exp.3. However, in the current experiment, participants were informed explicitly that simultaneously presented feedback might also be the "wrong answer" to a question. Thus, there is minimal probability that, being informed of this inherent risk, they would be inclined to adopt the feedback to the same extent as when only the correct answer was presented as feedback. Therefore, both statistical indication and theoretical implication led us to constrict the *c* parameters (i.e., setting them all constant at zero) for further analyses. Results then showed an identifiable model; however, not a satisfactory model fit. Analyses were then conducted with discrete groups. Results showed a satisfactory model fit with the data of the high depletion group, $G^2 = 35.44$, $df = 20$, $p > .01$, but not for the low depletion group $G^2 = 43.28$, $df = 20$, $p < .01$. Inspired by Bayen et al (2006) and Pohl et al (2010), the reason of this non-fit for low depletion group might be due to the violation of the symmetry assumption. Considering that the HB item set used in the current study was never used in previous study, there might be some items that violate the symmetry assumption of the HB13 model. Four such items were then identified and excluded from the data set, i.e., items that had significantly high number of response frequencies in the ROJ = OJ < CJ category compared with the ROJ = OJ > CJ category. Participants in low depletion group typically gave underestimated OJs for these items and during later ROJ phase gave perfect ROJs. Excluding these 4 items resulted in a satisfactory model fit of the data for low depletion group $G^2 = 39.25$, $df = 20$, $p > .01$. The parameter estimates of the model changed only slightly compared with the set including all 88 items (smaller than 0.01). Thus, the exclusion of the 4 items did not change the main results. More importantly, the high depletion group was more of the interest of the current study, but not the low depletion group. Reporting results of low depletion group below was only for the integrality of the dissertation.

Parameters denoting probability of correct ROJs and of a biased reconstruction if direct recall of OJs failed were then obtained for each condition before and after depletion (see Table 10).[2]

Table 10

Results of Testing the Presence of Recollection Bias and Reconstruction Bias in the High Depletion group (upper part of the table) and in the Low Depletion group (lower part of the table), Pre and Post-depletion Respectively (Exp.4)

(high depl. Group)		Pre-depletion		Post-depletion	
Type of bias	HB condition	Null hypothesis	$\Delta G^2(df = 1)$	Null hypothesis	$\Delta G^2(df = 1)$
Recollection bias	CCJ	$r_{C1} = r_{EC1}$	0.35	$r_{C2} = r_{EC2}$	10.74*
	WCJ	$r_{C1} = r_{EW1}$	1.80	$r_{C2} = r_{EW2}$	9.39*
Reconstruction bias	CCJ	$b_{C1} = 0$	88.21**	$b_{C2} = 0$	60.90**
	WCJ	$b_{W1} = 0$	30.99**	$b_{W2} = 0$	63.03**

(low depl. Group)		Pre-depletion		Post-depletion	
Type of bias	HB condition	Null hypothesis	$\Delta G^2(df = 1)$	Null hypothesis	$\Delta G^2(df = 1)$
Recollection bias	CCJ	$r_{C1} = r_{EC1}$	0.01	$r_{C2} = r_{EC2}$	5.89*
	WCJ	$r_{C1} = r_{EW1}$	0.02	$r_{C2} = r_{EW2}$	3.30*
Reconstruction bias	CCJ	$b_{C1} = 0$	53.29*	$b_{C2} = 0$	31.07*
	WCJ	$b_{W1} = 0$	42.38**	$b_{W2} = 0$	57.26**

[2] For the high depletion group, I partly restricted the guessing parameter (*h*) for the following analyses due to their large standard errors and in order to obtain more salient comparison results. The restriction was to set guessing parameters equal across control and 2 experimental HB conditions, but different across depletion group and depletion time. Moreover, a complete restriction of the *h* parameter, i.e., set *h* all equal across all conditions, yielded the same pattern of results as those with a partly *h* restriction.

Note. ΔG^2 = decrement in the likelihood ratio chi-square goodness-of-fit statistic compared with the unrestricted model; r_C = probability of participants' correctly recalling the original judgment (control items); r_E = probability of participants' correctly recalling the original judgment (experimental items); r_{EC} = probability of participants' correctly recalling the original judgment (experimental items when feedback was labeled as "correct answer"); r_{EW} = probability of participants' correctly recalling the original judgment (experimental items when feedback was labeled "wrong answer"); *b* = probability of a biased reconstruction given a failure to recall the original judgment; with 1, 2 denoting the high depletion group pre and post-depletion, respectively; and with $_C$, $_W$ denoting the condition labeled "correct answer" and that labeled "wrong answer", respectively). CCJ = condition labeled "correct answer" condition, WCJ = condition labeled "wrong answer".
** $p < .01$, * $p < .05$

The Subsequent Table 11 presents the pre and post-depletion comparisons of each depletion group for the parameters of interest. The same principle of testing such comparisons was adopted as in Exp.3.

Table 11
ΔG^2 values of Testing Pre vs. Post-depletion Differences in Multinomial-Model Parameters (Exp.4)

Comparison conditions	r_C	r_{EC}	r_{EW}	b_C	b_W
High depl. pre vs. post-	.14	5.43*	1.70	2.10	5.09**
low depl. pre vs. post-	.09	5.32*	1.15	0.18	1.99

Note. See Table 10 Note.

Figure 15 shows the estimates of the parameters of interest, i.e., recollection bias parameters and reconstruction bias parameters for both depletion intensity groups and both depletion time conditions.

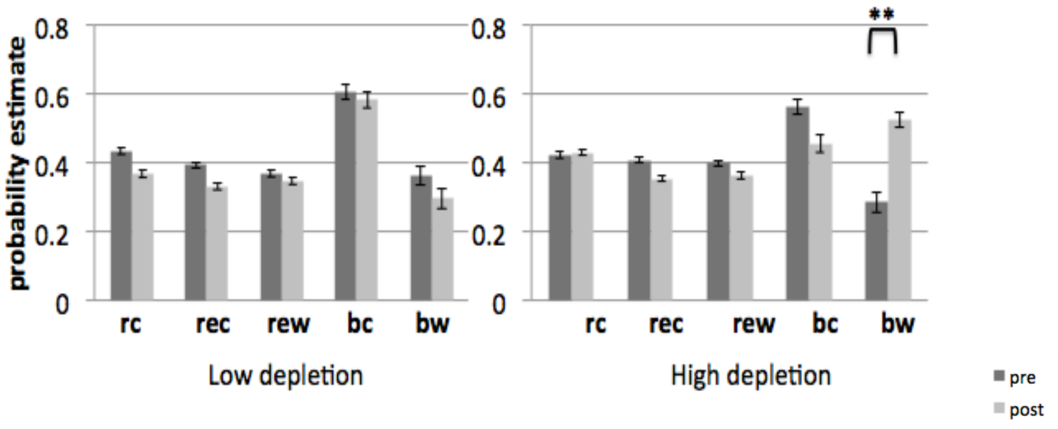


Figure 15. Multinomial model parameter estimates for the depletion groups of Exp.4. Recollection bias is the difference between the correct recollection in the control condition (r_C) and the correct recollection in the experimental condition (r_E). Error bars indicate 95% confidence intervals.

Recollection parameters

As shown in Table 10, while recollection bias was present in both experimental conditions after depletion in the high depletion group, it was only present in the "wrong answer" condition before depletion in the low depletion group. In the case of the high depletion group (also being the group I was most interested in), the exact estimates of parameter r_C (with their 95% confidence intervals) were 0.42 (±0.02) and 0.43 (±0.02) for pre and post-depletion respectively; the corresponding r_E values were 0.41 (±0.02), 0.40 (±0.01), 0.36 (±0.02) and 0.37 (±0.01) for the "correct answer" condition pre and post-depletion, and for the "wrong answer" condition pre and post-depletion. Moreover, as shown in Figure 15, there was an overall tendency toward a decrease in recollection bias parameters across all conditions after depletion. Further tests (Table 11) showed that this decrease was statistically significant in the control condition of the low depletion group and the "correct answer" condition for both depletion groups. Meanwhile, the decrease was not significant for the "wrong answer" condition of either depletion group.

Reconstruction bias

As indicated by the reconstruction parameter b, results from Table 10 show reconstruction bias is present to a significant degree in all conditions. The exact b values (with their 95% confidence intervals) for the high depletion group were 0.56 (±0.05), 0.45 (±0.07), 0.29 (±0.06) and 0.53 (±0.05) for the condition labeled as "correct answer" both pre and post-depletion and for the "wrong answer" condition pre and post-depletion respectively. As shown in Figure 15, the reconstruction bias parameters decreased for both experimental groups after depletion in the low depletion group. However, neither decrease was significant as demonstrated in Table 11. As for the high depletion group, a more intriguing pattern for the reconstruction parameter was found indicating a decrease in the “correct answer” condition, but an increase in the “wrong answer” condition. Table 11 further shows that only the increase was significant through all conditions. In regard to the labeling effect indicated by the reconstruction parameter, the multinomial modeling analyses revealed the same results pattern as the conventional ANOVAs (thus, I will not discuss this point further).

Discussion

Results on the response accuracy of the AB-AD task acted as manipulation checks. Specifically, the significant decrease in response accuracy of the high interference condition compared with that of the low interference condition across two depletion groups indicated the validity of the AB-AD task. Moreover, the significant increase in response accuracy after depletion procedure showed that the depletion procedure did work in Exp. 4.

As with the percentage of correct OJs, the non-significance of the difference between two depletion intensity groups before depletion ensured that there was no difference in knowledge about the HB items for the two groups of participants.

Regarding the perfect recall percentage during ROJ, results showed that participants assigned to the high depletion group did not remember their OJs significantly better or worse than their counterparts in the low depletion group. Furthermore, as the labeling effect in perfect recall percentage only existed after high depletion, we may surmise that participants only remembered their initial estimations worse after highly intensive cognitive tasks. This deterioration in performance was not dependent on the nature of later presented feedback, i.e., whether participants were provided with correct or incorrect answers to the questions. The main effect of depletion time (pre- vs. post-depletion) and results from further paired sample t-tests showed that while overall participants were less able to remember their OJs after depletion than prior to depletion, only those in the low depletion group performed significantly worse, and then only when feedback was labeled as “correct answer”. While we might suspect this general decline in performance after depletion is due to lower motivation, the same column height of pre and post-depletion in the control condition in the high depletion group resists this interpretation (see Figure 13). Also, for the paired sample t-tests, I set the error probability to .05 instead of .01 for all other tests in the current study. When the significant value cut-off was set to .01, none of the above given t-tests for perfect recall percentage was at all significant. Results of Figure 14 revealed that the depletion procedure did work well on the delta-z measure of HB. More specifically, after intensively performing continuous tasks involving interference resistance ability, participants were significantly less biased when the feedback was labeled as “correct answer” as compared to their pre-depletion state. In contrast, they were significantly more biased toward the feedback if it was labeled “wrong answer” after depletion. Note that this double-dissociation was only demonstrated in the high depletion group, and not in the low depletion group.

When we switch the columns of Figure 13, the labeling effect[3] is observed. Namely, compared to the control condition (i.e., no feedback), the magnitude of bias toward feedback is significantly larger in the condition labeled “correct answer”. However, the HB delta-z measure dropped significantly when feedback was labeled “wrong answer”, but was still significantly larger than in the control condition. The robustness of this labeling effect of the current experiment could be due to the within-subjects manipulation, instead of the between-subjects manipulation typically used in previous studies in which a much smaller labeling effect was found. Interestingly, this labeling effect was present in both groups, however, only before depletion. After depletion, the labeling effect disappeared, regardless of the depletion intensity. More specifically, while participants were still significantly more biased when feedback was presented, the bias did not differ whether this feedback was labeled as “correct answer” or “wrong answer”.

At first glance, the disappearance of labeling effect after depletion perhaps invites the supposition that after depletion participants have lower motivation to focus on the task, but higher motivation to finish the task, thus ignoring the detailed nature of the feedback provided, i.e., how it is labeled. However, upon further deliberation of the results, several factors counter this initial response. First, if it were due to a problem of motivation, participants’ perfect recall percentage should also be influenced. However, as illustrated in

[3] I prefer the terminology of “labeling effect” than “plausibility effect” for the Exp.4, as no subjective plausibility was examined as what some previous studies did (Hardt & Pohl, 2003).

Figure 13, the perfect recall rate did not differ significantly when we compare the item type conditions before and after depletion in almost all cases across both groups. Second, given the relative measure of HB, if it were due to a motivation problem participants should become more biased after depletion to an equal degree, regardless of the labeling condition. In fact, the dissociation pattern revealed that participants were significantly less biased after depletion when the feedback was labeled "correct answer". Third, to further rule out the possibility of a motivation problem, the response time was thus checked. As shown in Table 9, participants did not differ significantly in their response time when completing the HB task after depletion as compared to before depletion, as indicated by the absence of the main effect of time (pre vs. post-depletion) across both groups and non-significant paired sample t values before and after depletion within each experimental condition.

Taken together, the evidence suggests the depletion procedure could not have depleted participants' motivation. Rather, some other fundamental mechanisms could have been fatigued, leading to the post-depletion dissociation pattern, i.e., significantly less bias when the given feedback was labeled as "correct answer" as in a typical HB situation; likewise, significantly more bias was present when the feedback offered was labeled "wrong answer". As I hypothesized prior to the experiment, in the "correct answer" situation, the adaptive knowledge-updating process could be more involved yielding a substantial bias toward the feedback. As such, after depletion participants would allocate fewer cognitive resources to this process, and therefore less bias should be produced by the feedback. Similarly, in the situation in which items are labeled as "wrong answer", the interference resistance process could be more involved resulting in relatively less bias. After depletion, fewer cognitive resources would be employed for this process, consequently leading to more bias toward the feedback. It should however be noted that the fundamental anchoring effect in both experimental conditions still existed, irrespective of how feedback was labeled.

In summary, the hypotheses were supported by the results. The conventional ANOVAs showed that after depletion there were no significant differences in perfect recall percentage (see Figure 13). However, HB magnitude indicated by the proximity index was significantly altered after depletion (Figure 14). Moreover, comparison between pre and post-depletion conditions in high depletion group suggests that: 1) before depletion, participants' HB were robust, regardless of how feedback was labeled, although less bias was present in the condition labeled "wrong answer", and 2) after depletion, participants were less biased toward the feedback when it was labeled “correct answer”, but more biased toward the feedback when it was labeled “wrong answer” (i.e., the single dissociation).

When examined using multinomial modeling analyses, as shown in Figure 15, the recollection parameters were not significantly different between the two experimental conditions (i.e., no labeling effect with recollection bias parameters). This was inconsistent with previous studies in which discrediting manipulation had been applied; such work had found estimates of recollection bias parameter differed significantly between different discrediting groups (Hasher et al., 1981, Erdfelder & Buchner, 1998, Exp.3). This discrepancy could be due to different procedures being employed to those used in the current experiment, namely a 1-day retention interval instead of a 1-week gap in the other two studies, and the simultaneous presence of feedback and ROJs here rather than the separate presentation used in previous studies. Moreover, in the current experiment, feedback was presented as being the “correct answer” or “wrong answer” to a given question. In Erdfelder & Buchner (1998)’s Exp.3, however, participants in the feedback discrediting group were told that *all* the supposedly “correct answers” they received were wrong. Such manipulation may have led to a more powerful discrediting effect as participants could gain better access to their memory of OJs than those in the group without discrediting.

Both methods of analysis showed a similar result pattern. Namely, the most heuristic response strategy of the HB memory task (i.e., direct recall) was not affected by depletion, suggesting that this strategy required only minimal cognitive resources. Thus, both the perfect recall percentage showed by ANOVAs and the recollection bias parameter estimations obtained from multinomial modeling analyses were no different before and after the depletion procedure. Instead, only the shift toward feedback measure and reconstruction bias parameter estimations differed significantly after depletion. Moreover, this difference could not be due to the longer retention interval between OJs and ROJs after depletion, because if this were so there would also be a significant difference in perfect recall percentage and recollection bias parameter estimations, which was not the case as discussed above.

When we compare results obtained from the conventional ANOVAs with results from the multinomial modeling analyses, we see a divergent pattern in the high depletion group. More specifically, ANOVAs showed the difference in delta-*z* mean between pre and post-depletion was more significant in the condition labeled “correct answer” than in the condition labeled as "wrong answer" (Figure 14). However, the multinomial analyses suggested that the difference in reconstruction bias parameter estimations between pre and post-depletion was only significant in the "wrong answer" condition (Figure 15). This inconsistent result pattern derived from two methods of analysis points to the advantage of multinomial modeling when applied to the HB paradigm. That is to say, the depletion procedure included tasks involving the interference resistance process. Thus, a specific interference resolution ability was supposedly more depleted, as opposed to general cognitive resources. Consequently, the condition labeled "wrong answer", in which increased interference process was most relevant, could have been more depleted, leading to a more significant difference in the reconstruction bias parameter estimation. Quite apart from the superiority of multinomial analyses, the inconsistent results pattern could also be due to the different dependent variables adopted by the two analyzing methods, with the *amount of shift or difference* toward feedback (delta-*z*) in ANOVAs, and *probability* of reconstruction bias in multinomial modeling analyses.

3. General Discussion

The current study was a theory-driven research in HB memory design, rather than a confirmatory project. The central goal was to explore the fundamental mechanisms underlying this robust cognitive phenomenon. I applied the ego-depletion philosophy on a proposed potential cognitive process related to HB, this being interference resistance. The pattern of depletion effect on HB was then obtained. The main aim was accomplished through a robust, step-by-step methodology. Prior to the experiment used to test the main idea, I conducted 2 experiments for task selection and to check manipulation.

Experiment 1 confirmed the appropriate task type and detailed task structure for further application. The results of Exp.1 verified the interference effect using the classic retroactive interference task, i.e., AB-AD task by Müller & Pilzecker, 1900 (cited in Anderson & Neely, 1996). More specifically, participants performed worse in recollection of the numbers from the first study list (word-number pairs) when they were also presented with a second study list, consisting of the same words paired with different numbers, as compared to when no such second study list was presented. Furthermore, although this typical retroactive interference effect existed in all conditions, a statistically significant effect was only demonstrated when 5 word-number pairs were included within one study list. Thus, the most appropriate length of word-number pairs for the interference task was confirmed. The significant increase in response time provided further support for the decreased recall performance when there was a similar second list to study, confirming that including 5 word-number pairs within one study list would yield the greatest interference effect.
Exp.2 checked the depletion manipulation based on Baumeister et al. (1998)'s ego depletion philosophy and Persson et al (2007)'s depletion technique. I used the verified AB-AD task as the transfer task before and after the depletion procedure, and three tasks requiring interference resistance from Persson et al. (2007) and Persson & Reuter-Lorenz (2008), these being a letter recognition task, an *N*-back task and a verb generation task as the training task during the depletion procedure. Consistent with my expectations, the results showed there was no difference in performance of the AB-AD task after depletion if participants went through the low depletion procedure, in which only conditions requiring minimal interference resistance ability were included in the training tasks. However, after participants went through the high depletion procedure, in which only conditions demanding high interference resistance ability were included, their performance in the AB-AD task was significantly

deteriorated. Furthermore, this performance deterioration only occurred in the condition when there was a second list to study. These results first proved the effect of the depletion procedure. They also confirmed that only a specific cognitive process, i.e., the interference resistance ability in the current study, was depleted, rather than general cognitive resources. A specific depletion nature of this kind was crucial for further experiment, as the intention was to determine whether the HB memory paradigm involves a specific interference resistance process. Therefore, the depletion procedure must work specifically on this process for us to apply it to a HB task. The results of Exp.2 provided the requisite affirmation of this.

In the principal experiments of this study (Exp.3 and 4) I adopted the depletion procedure on the HB memory paradigm. Participants in Exp.3 were given a typical HB task including a control condition in which no feedback was presented and an experimental condition wherein the correct answer was present as feedback. I then compared HB magnitude in the pre and post-depletion phase, with participants going through either the low depletion procedure or the high depletion procedure. I was more interested in the difference in HB task performance in the high depletion group, while the low depletion group worked as a control baseline. Results of the ANOVAs on the perfect recall percentage and the shift measure of HB both indicated a decreasing trend in the high depletion group after depletion. More specifically, participants were less able to remember their previous estimations of answers after having performed intensive interference tasks demanding a high interference resistance process than they were beforehand. In situations where participants did not successfully recall their original estimations, there was a less biased trend in shift magnitude toward the feedback after depletion. Unfortunately, such results were neither statistically significant nor in line with the prior expectations. Prior to Exp.3, I offered the hypothesis that the interference resistance process was most probably involved in HB memory paradigm. In other words, during the recollection of previous judgments, the robust bias toward later presented feedback information could be due to the failure of resisting the interference induced by feedback information. Thus, after depletion of such interference resistance ability, people should be more biased toward the feedback; this ran counter to the results. I inferred then that the typical HB memory paradigm was more than just interference. There must be a certain process involved in HB that was crucially different to what we would observe in typical retroactive interference paradigm (such as the AB-AD task). Considering the difference between HB and the classic retroactive interference paradigm, the nature of later presented information in HB could play a fundamental role. That is, feedback in typical HB memory tasks are *correct* answers to the questions. Thus, people are more adept at learning the feedback in an implicit and embedded way, resulting in the biased judgment toward the correct information when they are asked to recall their own initial judgment and when they fail to directly access the memory trace of their initial estimation. This bias could therefore be produced by updating their knowledge-base and be adaptive in the evolutionary aspect (Dekker, 2004). In turn, it was not the interference resistance process that was most fatigued in Exp.3. Rather, this knowledge-updating process was exhausted. Consequently, people were less influenced by feedback during their recollection of the original answers. As with the non-significance in the decreasing trend, a small sample size in Exp.3 may account for the low degree of significance. Thus, in the next experiment, the sample size was raised to increase the potential to attain a significant level of difference.

In order to test post-hoc interpretation of Exp.3, I added one more experimental condition to the HB task in Exp.4. Namely, participants were presented with no feedback, or with feedback either labeled as “correct answer” (as in Exp.3) or “wrong answer” (new experimental condition in Exp.4). Combining this labeling manipulation with the depletion procedure, a clearer picture of what plays a leading role in HB memory paradigm is displayed. Results highly supported the hypotheses. The ANOVAs revealed no significant difference in participants’ perfect recall percentage after depletion in almost all conditions. Analyzing the shift measure, HB was still robust, indicated by the significant increase of delta-*z* mean in both experimental conditions (with feedback) compared with the control condition (no feedback), regardless of pre or post-depletion phases. Meanwhile, the shift measure was shown to be significantly smaller with the traditional “correct answer” labeled condition and larger in the “wrong answer” condition after high depletion. Thus, although the same feedback values were presented, both leading to strong HB, different mechanisms could be underlying these outcomes when the feedback is labeled differently. As argued prior to conducting Exp.4, while a particular form of knowledge updating could play the primary and adaptive role in the typical HB situation (feedback labeled as “correct answer”), the interference resistance process may exert more influence when feedback is labeled as “wrong answer”. As a result, after depletion of different processes, the HB shift measure was influenced in different directions. The multinomial modeling analyses further supported the results pattern of the ANOVAs; this being, that after depletion, reconstruction bias parameter estimation decreased (showing less HB) in the condition labeled “correct answer” condition but increased (showing more HB) in the condition labeled “wrong answer”. In addition, the multinomial modeling analyses were more sensitive and accurate in their ability to detect nuances in the results. For example, while the shift measure ANOVAs showed that the differences with both experimental conditions were significant after depletion, the multinomial modeling analyses indicated that the reconstruction bias parameter estimation was only significantly different in the condition labeled “wrong answer”. The decrease in the condition labeled “correct answer” was less and no longer significant after depletion. This is consistent with the fact that the depletion procedure targeted the interference resistance process rather than other more general cognitive resources. In the condition labelled "wrong answer", the interference resistance process dominated. This process was subsequently depleted more than other cognitive processes via the current depletion procedure, leading to a significant increase in the bias toward the feedback.

When comparing pre and post-depletion in Exp.3 (the traditional HB situation) with the condition labeled “correct answer” in Exp.4, results are in line with knowledge-updating theories in previous studies (for a review, see Hawkins & Hastie, 1990). This strand of theory asserts that outcome information updates the knowledge-base from which the initial judgment is drawn. Subsequent reconstruction is therefore based on the freshly updated knowledge-base if direct recall strategy fails. As a result, the reconstructed judgment is biased toward the outcome information, i.e., hindsight bias. In Exp.3, the depletion procedure produced fewer cognitive resources demanded by such knowledge-updating behavior, causing a decreasing trend in the bias. The knowledge-updating notion is the only line of existing theory that could explain such a decrease. The condition labeled “correct answer” in Exp.4 provided further confirmation of this interpretation.

As with the comparison between different labeling of the two experimental conditions (either correct or wrong answer of the question) prior to the depletion procedure in Exp.4, results supported the anchoring theory and plausibility effect of HB (Strack & Mussweiler, 1997; Pohl, 1998a; Erdfelder & Buchner, 1998 and Hardt & Pohl, 2003). Namely, compared with the control condition (no feedback), participants still biased hindsight to a robust degree even though feedback was explicitly labeled as “wrong answer”. It suggests that people may still anchor to presented information while they attempt to reconstruct their initial answer to a question. However, as compared to the condition labelled “correct answer”, the magnitude of HB was significantly less in the “wrong answer” condition. In the former situation, participants could engage in more *integrating* behavior (e.g., knowledge-updating or anchoring), whilst in the latter case, more *gating* behavior is involved. It has not however been uniformly clarified in previous research when and how such gating behavior occurs. That is, it could be that participants employ inhibition process early in the presentation of outcome information due to the "wrong" labeling of it. Also, previous research does not rule out the possibility that suppression only occurs after the outcome information has entered working memory or episodic memory (varying according to the retention interval between CJ and ROJ), and has consequently updated the initial knowledge-base. During reconstruction of the initial judgment, people would then have to resist the interference induced by the feedback, i.e., less anchoring. Taken together however, different labeling conditions make an observable difference in the extent to which people are biased toward feedback.

The plausibility operation varied the nature of feedback information at the two ends of the spectrum, either right or wrong. This was a more powerful manipulation than that used in some previous studies, e.g., Pohl (1998)'s labeling of the feedback as "another person's answer". In previous studies that explicitly told participants feedback was wrong (Hasher et al., 1981, Erdfelder & Buchner, 1998), this information was only presented *after* participants had initially encoded feedback as the correct answer. The timing of plausibility manipulation could be critical in that the inhibition process might have occurred early during the presentation of outcome information. Therefore, a more salient plausibility effect was demonstrated in Exp.4.

Aside from plausibility effect, the comparison of the results of pre and post-depletion in the condition labeled "wrong answer" in Exp.4 provided further evidence of inhibition theory explaining HB. This line of theory is typically suggested by developmental studies in HB (Bayen et al., 2006, Pohl & Haracic, 2005, Pohl et al., 2010). Through a lifespan, both HB and inhibition-related abilities exhibit the same form of U-shaped function, only in a reversed pattern, with young children and older adults being at the two ends of the curve and younger adults in the lowest middle point with HB. As discussed, the condition labeled "wrong answer" could enlarge the extent to which inhibition process is involved in HB. In addition, the depletion procedure was specifically tailored to fatigue the interference resistance process. After such depletion, people should have less cognitive resources to engage in the inhibition process, resulting in more HB, which was exactly what the results of Exp.4 illustrated.

However, the current results clearly did not support any of these single process theories, especially indicated by the dissociation with the two HB experimental conditions in Exp.4. It could be argued that the strong plausibility manipulation has changed the typical nature of HB phenomenon in the “wrong answer” condition. If so, this condition would be nothing more than a normal anchoring paradigm, rather than HB. However, the results of Exp.4 showed that there was still robust HB in the condition labeled “wrong answer” across both groups and phases. That is, although feedback information was strongly discredited, a considerable amount of HB could still be elicited. Thus, both experimental conditions in Exp.4 could be considered as variations of the HB paradigm. Consequently, different magnitudes of HB between the two conditions before depletion and different trends in direction after depletion suggest the inadequacy of a single process theory explaining HB. There should be more than one process involved in the HB memory paradigm that is mitigated by the depletion procedure, eventually leading to the adverse direction of the difference, albeit via the same depletion technique.

Another feature of the current results that cannot be overlooked is the control aspect. As discussed previously in regard to the principle of ego depletion (Baumeister et al., 1998), fatigue after intensively performing certain tasks acts only on *controlled* activities that require active guidance of self-control, but not on simpler and *automatic* activities in which no engagement of cognitive resource is demanded. Both Exp.3 and Exp.4 revealed effective depletion effect suggesting that the underlying processes of the HB paradigm are controlled behaviors demanding cognitive resource to function. This notion is inconsistent with a number of previous theories. Some of these theories overtly assert that an automatic assimilation is involved during presentation of outcome information inducing the biased post-outcome judgments of pre-outcome estimations (Fischhoff, 1975), while other theories[4] do not specify the controlled nature of the underlying processes of HB, e.g., SARA model (Pohl et al., 2003).

Taken together, I would propose a dual-process framework for HB memory paradigm, including two parallel processes – interference resistance and adaptive knowledge updating – that are both controlled (Figure 16). As shown in Figure 16, the magnitude of HB (the vertical coordinate) is modulated by the amount of the involvement of these two processes. The

---

[4] Again, as I applied the HB memory task using almanac questions for the current study, only theories relating to this specific HB paradigm are included in discussion of current study.

horizontal coordinate denotes the dimension of influencing elements, generating the relative engagement extent of the two underlying mechanisms. Taking the current experiments as an example, the horizontal dimension would be feedback plausibility. More specifically, with conditions in which feedback is highly plausible (e.g., being the correct answer of one question), more knowledge updating behavior would be invested, leading to a large HB magnitude. In contrast, when the feedback is highly implausible (e.g., being the wrong answer of the question), the horizontal dimension is pulled to the right, i.e., more interference resistance action would be elicited, resulting in smaller HB below the baseline amount. However, it is hard to fully eliminate HB. With other manipulations or other situations, the horizontal dimension could be some other factors, e.g., the accessibility of the initial estimation, etc., which was not empirically demonstrated in the current experiments. Furthermore, both processes are controlled rather than automatic, demanding a certain amount of cognitive resources to accomplish.

Results of the Exp.3 and 4 provide empirical support for the theoretical dual process framework. With Exp.3, the typical HB paradigm was employed, where CJ is presented as “correct answer” of the question for the experimental items. According to the dual process account, knowledge updating behavior was taking the dominant role during ROJ, thus resulting in robust HB. The depletion procedure correspondingly fatigued the knowledge updating behavior and yielded a decreased HB (descriptively). Further with Exp.4, for experimental items of HB task, CJ is either presented as “correct answer” (as in Exp.3) or labeled as “wrong answer” of the question. Thus, although there was an “answer” appeared on the screen when participants were required to give their ROJ, this “answer” was supposed to be “wrong answer”. Consequently, participants were more likely to engage in inhibitory behavior with “wrong answer” HB items, so that they would not influence their direct recall of OJ from memory, and/or would not interfere with their reconstruction process of the OJ. Less cognitive resource would be assigned for any knowledge updating behavior resulting in the labeling effect of HB before depletion. After depletion, the cognitive resource that was supposed to be allocated for interference resistance process was cut short. HB magnitude was then increased after depletion.

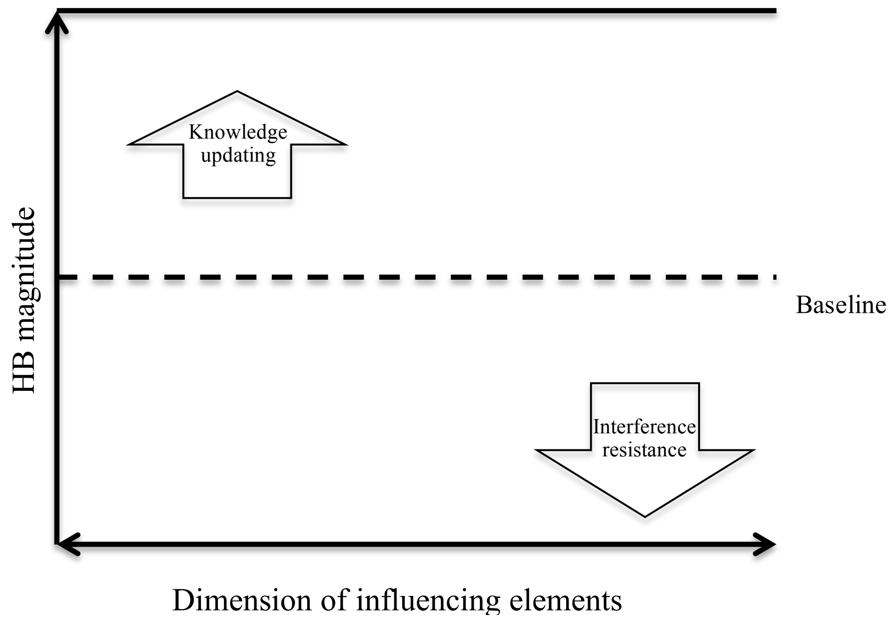


Figure 16. A refined framework of the dual-process account of HB memory paradigm.

Aside from the controlled nature of the two processes, I would also argue that with the current framework "interference resistance" and "knowledge updating" cannot be seen as merely two sides of the same coin. It is more likely that they are two parallel independent processes, with one taking the dominant role under most conditions. There are extreme cases in which the feedback is surprising and the accessibility of the initial estimation is high (e.g., short retention interval between OJ and ROJ); in such instances people could still engage in knowledge updating while rather effectively resisting the interference of the outcome, resulting in a less biased ROJ and reaching a near baseline level of HB. These extreme cases suggest that interference resistance and knowledge updating are more than just different means of interpreting HB. Instead, there are two possibilities of how these two processes function in this dynamic framework. First, they could function independently as long as there are sufficient cognitive resources, a contention that the depletion results lend weight to. Specifically, if they merely represented the same fundamental idea from varying perspectives, HB under two experimental conditions (labeled "correct answer" or "wrong answer" respectively) should be influenced by the depletion procedure in the same direction (both increasing or both decreasing). Consequently, a middle level of HB magnitude could be reached with both conditions after depletion. However, the dissociation I found counters such an argument. As such, there must be two processes that are separately depleted with respect to different situations, either one of which may take the dominant role. The second possibility of how these two processes function takes the form of a lever that adjusts the final magnitude of HB (see Figure 17). The right hand of the lever could be regulated by interference resistance ability while the left hand is accommodated by knowledge updating behavior. In this way, the two underlying processes may influence each other. Higher interference resistance ability could lead to less knowledge updating behavior, and consequently less HB. This second solution is a more dynamic possibility of how the two processes interact. Nevertheless, the current study was unable to rule out either possibility, and I therefore could not determine the precise mode of dynamic function operating between these two underlying processes.

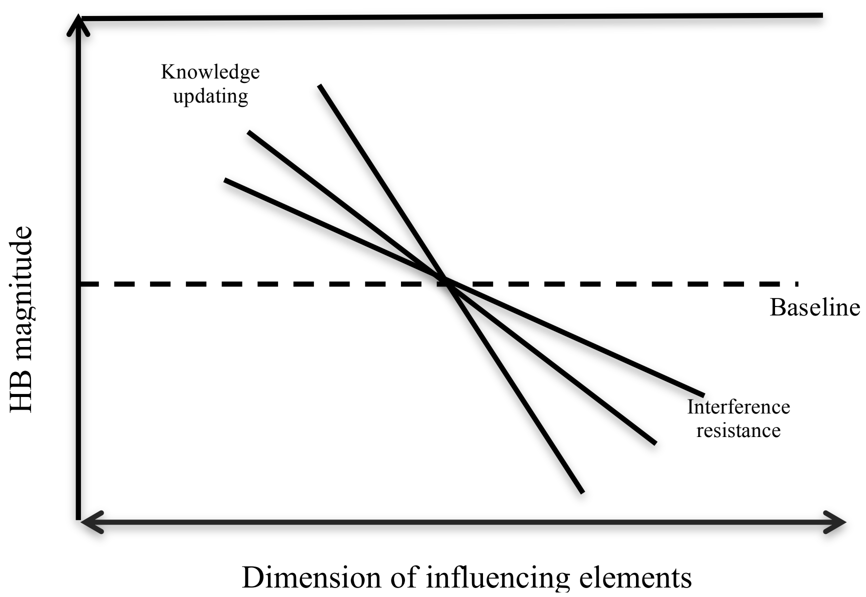


Figure 17. A refined framework of the dual-process account of HB memory paradigm (alternative dynamic).

For an interim short summary, the dual-process framework is superior to previous theories explaining HB memory paradigm in the following three respects. First, it acknowledges and adopts two processes as jointly producing HB rather than attributing the phenomenon to a single theory. In doing so it renders a compelling explanation for some of the critical findings about HB that could not be explained by previous theories. For example, the assimilation and knowledge updating theory alone cannot explain the difference in HB between children, younger adults and older adults in developmental findings. Likewise, inhibition theory by itself cannot explain the depletion effect with the typical HB condition, i.e., feedback labeled as “correct answer”. Furthermore, the more sophisticated SARA model does not offer a conscientious elucidation of plausibility effect in previous studies and the quite robust labeling effect seen in the current experiment. Thus, by not overlooking the influence of either of the two crucial processes, the current account offers a dynamic framework that explains HB from a more flexible, nuanced perspective.

Secondly, the controlled nature that is attributed to the framework, which is not proposed in any previous theory, is found in two respects. Both processes can be *controlled* to engage at varying levels under different circumstances. After situations are evaluated meta-cognitively, the two processes would then be distributed with a different weight of engagement, resulting in a certain amount of HB. Both processes are *controlled*, as opposed to automatic, and demand cognitive resources to function. They can be fatigued through intensive performance of certain related tasks, and could likely be improved through regular training in such tasks; the former claim here was empirically supported by the current experiments, while future research can examine the latter proposition using training procedures. Thus, thirdly, by specifying these two underlying processes, the current theory provides exclusive direction for any future training study, i.e., either structuring training to improve people's interference resistance ability or constructing situations to prevent them updating their original knowledge-base if the intention is to reduce HB.

Despite the advantages of the current framework vis-à-vis previous theories discussed above, I acknowledge there are certain limitations and flaws in the current study. Firstly, I used general knowledge questions requiring numerical judgments as materials; thus, the theoretical conclusions drawn from the current results are limited in their application to this specific type of HB. In both research literature and real life contexts however, various types of HB exist. The current theory could not provide insight into paradigms other than the memory distortion mode.

Secondly, the labeling effect applied in the current study was situated on the two extremities of the plausibility dimension, i.e., correct answer being extremely plausible and wrong answer being highly implausible. Thus, we obtained a rather robust labeling effect implying the obvious difference in the involvement of each process under corresponding conditions, i.e., dominant knowledge updating engagement when feedback is labeled as "correct answer" of the question on the one hand, and the overriding role of interference resistance in the condition labelled "wrong answer" on the other. However, the relative flowing dynamic in the middle was not examined by the current study. Future research could vary the feedback plausibility to allow for divergence in intensity, e.g., labeling the feedback as “another person’s estimation” as in Pohl (1998), or “random number”. When combined with the depletion procedure, we could test whether the engagement of underlying processes of HB follows the dynamic entailed by the current framework. Moreover, although I adopted a rather strong manipulation of labeling effect, the genuine plausibility effect bearing on participants was not explicitly measured. Thus, in future research, subjective plausibility would be better tested using either an opening questionnaire or a rating assessment (Pohl & Gawlik, 1995; Werth, 2002).

Thirdly, according to Ash (2009), HB is not a single time-point psychological phenomenon, but rather a hybrid of different processes; that is, people making a predictive estimation, followed by outcome information influencing their retrospective judgment about their initial prediction. Thus, a complete HB theory should explain all three processes. The current theory is unfortunately not complete in that sense. I am unable to draw any conclusions about when the two fundamental processes occur and their weight of involvement due to the experimental set-up, i.e., presenting CJ, labeling of CJ and ROJ requirement on the same screen. I would assume that both knowledge updating and interference resistance are principally active during the feedback presentation phase. The subsequent post-feedback recall would then be based on the updated knowledge-base. However, people may still try to resist the interference brought about by drawing judgment from the new knowledge-base, especially under situations with high accessibility to the original estimation, or with a high meta-cognitive consciousness of the HB concept, i.e., knowledge of the bias. The results and the theory of current study could not shed any light on this aspect of the problem. Thus, I am unable to distinguish when *either* of the two processes or *both* of them come to play their roles and how engaged people are. Future research could try to deconstruct and discern the interplay of these two processes during different phases by exerting corresponding manipulations with separate feedback presentation, feedback discrediting and recollection requirement.

Moreover, little weight has been given to the involvement of interference-related function, aside from in developmental studies on HB. While the current theory addresses the value of such function, it is unable to determine the precise occurrence timing of the interference resistance behavior. More specifically, when feedback is evaluated as being implausible, it could be inhibited either by gating before encoding during the feedback presentation phase, or by suppression after it is encoded and has entered the working memory during the re-judgment phase. This uncertainty remains unresolved. Future work could seek to illuminate this issue through a recognition test on the feedback. There should be some difference in the performance of feedback recognition between different labeling conditions (correct or wrong answer of the question) if the inhibition occurs early in the feedback presentation phase. In contrast, if it occurs mainly in the subsequent re-judgment phase, there should be no such difference.

I also suggest future research could apply manipulations to test other potential factors of influence in the horizontal dimensions of the current framework. For example, high accessibility of the original judgment could lead to considerable engagement of interference resistance behavior, in spite of substantial knowledge updating activity taking place at the same time because of our ability to adapt and absorb new information. However, the latter process could not be very influential in this situation, yielding a small HB, i.e., re-judgment from the blend of the previous knowledge-base and the updated one due to the partly effective participation of the interference resistance process.

To conclude, the current study was a theory-driven research on HB memory paradigm. I explored the underlying mechanisms of HB step by step through the combination of a labeling manipulation and a depletion procedure. Stemming from the current results, I propose a dual-process theory to explain HB. The refining knowledge updating process results from the enormous volume of everyday knowledge acquisition we experience in daily life; as a consequence, swift and ongoing update of our knowledge-base is ecologically necessary. Notwithstanding this, when a situation requires individuals to recollect an initial estimation that differs markedly from the correct answer, they then attempt to exert the corresponding executive functions (i.e., interference resistance) to overcome the influence of the adaptive updating inclination. The current theory offers an adaptable dynamic framework of HB making it more progressive than previous theories. More specifically, the underlying processes are not fixed, but flexible in terms of the timing and weight of their occurrence. Also, they can be subjectively controlled and regulated according to the objective and perceived state. This is crucial in the sense that hindsight bias is not a bias of lengthy evolutionary terms, but a result of an adaptive ecological function.

References

Anderson, M. C., & Neely, J. H. (1996). Interference and Inhibition in Memory Retrieval. *In Memory. Handbook of Perception and Cognition* (2nd edn) (Bjork, E. L. and Bjork, R. A., eds), pp. 237 - 313, Academic Press.

Arkes, H. R. (2013). The Consequences of the Hindsight Bias in Medical Decision Making. *Current Directions in Psychological Science*, *22*(5), 356–360.

Arkes, H.R., Wortmann, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology, 66*(2), 252-254.

Ash, I. K. (2009). Surprise, memory, and retrospective judgment making: testing cognitive reconstruction theories of the hindsight bias effect. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(4), 916–33.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*(5), 1252–65.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The Strength Model of, *16*(6), 351–356.

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*(5), 1003–18.

Bernstein, D. M., Atance, C., Loftus, G. R., & Meltzoff, A. (2004). We saw it all along: visual hindsight bias in children and adults. *Psychological Science*, *15*(4), 264–7.

Bernstein, D. M., Atance, C., Meltzoff, A. N., & Loftus, G. R. (2007). Hindsight bias and developing theories of mind. *Child Development*, *78*(4), 1374–94.

Blank, H., Musch, J., & Pohl, R. F. (2007). Hindsight Bias: On Being Wise After the Event. *Social Cognition*, *25*(1), 1–9.

Blank, H. & Nestler, S. (2006). Perceiving events as both inevitable and unforeseeable in hindsight: The leipzig candidacy for the olympics. *British Journal of Social Psychology, 45* (1), 149 - 160.

Blank, H., & Nestler, S. (2007). Cognitive Process Models of Hindsight Bias. *Social Cognition*, *25*(1), 132–146.

Blank, H., Nestler, S., von Collani, G., & Fischer, V. (2008). How many hindsight biases are there? *Cognition*, *106*(3), 1408–40.

Carlson, S. M., & Moses, L. J. (2001). Individual Differences in Inhibitory Control and Children's Theory of Mind. *Child Development*, *72*(4), 1032–1053.

Choi, D., & Choi, I. (2010). A comparison of hindsight bias in groups and individuals: The moderating role of plausibility. *Journal of Applied Social Psychology*, 325–343.

Christensen-Szalanski, J. J. ., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, *48*(1), 147–168.

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. http://dx.doi.org/ 10.1080/0361073X.2014.896667

Coolin, A., Erdfelder, E., Bernstein, D. M., Thornton, A. E., & Thornton, W. L. (2014). Explaining individual differences in cognitive processes underlying hindsight bias. Psychonomic Bulletin & Review. Advance online publication. http://dx.doi.org/10.3758/s13423-014-0691-5

Dehn, D. M. & Erdfelder, E. (1998). What kind of bias is hindsight bias? *Psychological Research, 61* (2), 135 - 146.

Dekker, S. W. A. (2004). The hindsight bias is not a bias and not about history Sidney W . A . Dekker Technical Report 2004-01 Lund University School of Aviation.

Eisenhauer, M (2000). *Selektive Aktivierung als Erklärung des Rückschaufehlers.* Unpublished Dissertation, University of Trier.

Erdfelder, E. (2000). *Multinomiale Modelle in der kognitiven Psychologie.*

Habilitationsschrift zur Erlangung der Lehrbefugnis im Fach Psychologie an der Philosophischen Fakultät der Rheinischen Freidrich-Wilhelms-Universität in Bonn. Bonn.

Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial Processing Tree Models. *Zeitschrift Für Psychologie / Journal of Psychology*, *217*(3), 108–124.

Erdfelder, E., Brandt, M. & Broder, A. (2007). Recollection biases in hindsight judgments. *Social Cognition, 25*(1), 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*(2), 387–414.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–91.

Fessel, F., Epstude, K., & Roese, N. J. (2009). Hindsight bias redefined: It's about time. *Organizational Behavior and Human Decision Processes*, *110*(1), 56–64.

Fischer, I. & Budescu, D. V. (1995). Desirability and hindsight biases in predicting results of a multi-party election. In J. P. Caverni, M. Bar-Hillel, F. H. Barron & H. Jungermann (Eds.), *Contributions to decision making* (Bd. 1, pp. 193–211). Amsterdam: Elsevier.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1* (3), 288 - 299.

Fischhoff, B. (1977). Perceived informativeness of facts. J*ournal of Experimental Psychology: Human Perception and Performance, 3* (2), 349 - 358.

Fischhoff, B., & Beyth, R. (1975). " I Knew It Would Happen ": Remembered probabilities of once-future things. *Organizational Behaviour and Human Performance, 13* (1), 1 - 16.

Gailliot, M. T., Plant, E. A., Butz, D. a, & Baumeister, R. F. (2007). Increasing self-regulatory strength can reduce the depleting effect of suppressing stereotypes. *Personality & Social Psychology Bulletin*, *33*(2), 281–94.

Groß, J., & Bayen, U. J. (2015). Hindsight bias in younger and older adults: The role of access control. *Aging, Neuropsychology, & Cognition, online*. doi: 10.1080/13825585.2014.901289

Guilbault, R., Bryant, F., Brockway, J. H., & Posavac, E. (2004). A Meta-Analysis of Research on Hindsight Bias. *Basic and Applied Social Psychology*, *26*(2), 103–117.

Hardt, O., & Pohl, R. (2003). Hindsight bias as a function of anchor distance and anchor plausibility. *Memory*, *11*(4-5), 379–394.

Harley, E. M. (2007). Hindsight Bias In Legal Decision Making. *Social Cognition*, *25*(1), 48–63.

Harley, E. M., Carlsen, K. a, & Loftus, G. R. (2004). The "saw-it-all-along" effect: demonstrations of visual hindsight bias. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*(5), 960–8.

Hasher, L., Attig, M., & Alba, J. (1981). I knew it all along: or, did I? *Journal of Verbal Learning and Verbal Behavior*, *20*(1), 86–96.

Hawkins, S. a., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311–327. doi:10.1037//0033-2909.107.3.311

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: an interaction of automatic and motivational factors? *Memory & Cognition*, *16*(6), 533–8.

Henriksen, K., & Kaplan, H. (2003). Hindsight bias, outcome knowledge and adaptive learning.

Hertwig, R., Fanselow, C., & Hoffrage, U. (2003). Hindsight bias: How knowledge and heuristics affect our reconstruction of the past. *Memory*, *11*(4-5), 357–377.

Hertwig, R., Gigerenzer, G., & Hoffrage, U. (1997). The reiteration effect in hindsight bias. *Psychological Review*, *104*(1), 194–202.

Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight Bias : A By-Product of Knowledge Updating ? *Cognition*, *26*(3), 566–581.

Hölzl, E., & Kirchler, E. (2005). Causal attribution and hindsight bias for economic developments. *The Journal of Applied Psychology*, *90*(1), 167–74.

Kane, M. J., Core, T. J., & Hunt, R. R. (2010). Bias versus bias: harnessing hindsight to reveal paranormal belief change beyond demand characteristics. *Psychonomic Bulletin & Review*, *17*(2), 206–12.

Louie, T. a., Rajan, M. N., & Sibley, R. E. (2007). Tackling the Monday-Morning Quarterback: Applications of Hindsight Bias in Decision-Making Settings. *Social Cognition*, *25*(1), 32–47.

Moshagen, M. (2010). multiTree: a computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, *42*(1), 42–54.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, *126*(2), 247–259.

Musch, J. (2003). Personality differences in hindsight bias. *Memory (Hove, England)*, *11*(4-5), 473–89.

Musch, J., & Wagner, T. (2007). Did Everybody Know It All Along? A Review of Individual Differences in Hindsight Bias. *Social Cognition*, *25*(1), 64–82.

Nestler, S., Blank, H. & Collani, G. (2008). Hindsight bias and causal attribution: A causal model theory of creeping determinism. *Social Psychology, 39* (3), 182 - 188.

Persson, J., & Reuter-Lorenz, P. A. (2008). Gaining Control: Training executive function and far transfer of the ability to Resolve Interference. *Psychological Sceince, 19*(9), 881–888.

Persson, J., Welsh, K. M., Jonides, J., & Reuter-Lorenz, P. a. (2007). Cognitive fatigue of executive processes: interaction between interference resolution tasks. *Neuropsychologia*, *45*(7), 1571–9.

Pezzo, M. V. (2003). Surprise, defense, or making sense: What removes hindsight bias? *Memory, 11* (4-5), 421 - 441.

Pezzo, M. V, & Beckstead, J. W. (2008). The Effects of Disappointment on Hindsight Bias for Real-World Outcomes. *Applied Cognitive Psychology*, *506*(June 2007), 491–506.

Pezzo, M. V., & Pezzo, S. P. (2007). Making Sense of Failure: A Motivated Model of Hindsight Bias. *Social Cognition*, *25*(1), 147–164.

Pieters, R., Baumgartner, H., & Bagozzi, R. (2006). Biased memory for prior decision making: Evidence from a longitudinal field study. *Organizational Behavior and Human Decision Processes*, *99*(1), 34–48.

Pohl, R. F. (1992). Der Rückschaufehler: Systematische Verfälschung der Erinnerung bei Experten und Novizen. *Kognitionswissenschaft, 3* (1), 38 - 44.

Pohl, R. F. (1995). Disenchanting hindsight bias. In J. P. Caverni, M. Bar-Hillel, F. H. Barron & H. Jungermann (Eds.), *Contributions to decition making* (Bd. 1, pp. 323 - 334). Amsterdam: Elsevier.

Pohl, R. F. (1998). The effects of feedback source and plausibility of hindsight bias. *European Journal of Cognitive Psychology, 10* (2), 191 - 212.

Pohl, R. F. (2004). Hindsight bias. In R. F. Pohl (Ed.), *Cognitive illusions: A handbook on fallacies and biases in thinking, judgment and memory* (p. 363 - 378). Hove, UK: Psycholgoy Press.

Pohl, R. F. (2007). Ways to Assess Hindsight Bias. *Social Cognition*, *25*(1), 14–31.

Pohl, R. F., Bayen, U. J., & Martin, C. (2010). A multiprocess account of hindsight bias in children. *Developmental Psychology*, *46*(5), 1268–82.

Pohl, R. F., Bender, M., & Lachmann, G. (2002). Hindsight Bias Around the World. *Experimental Psychology*.

Pohl, R. F., Eisenhauer, M., & Hardt, O. (2003). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory*, *11*(4-5), 337–356.

Pohl, R. F., & Gawlik, B. (1995). Hindsight bias and the misinformation effect: separating blended recollections from other recollection types. *Memory (Hove, England)*, *3*(1), 21–55.

Pohl, R. F. & Haracic, I. (2005). Der Rückschaufehler bei Kindern und Erwachsenen. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie*, *37*(1), 46–55.

Pohl, R. F. & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior and Human Decision Processes, 67* (1), 49 - 58.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psycholgoy of learning and motivation* (Bd. 14, p. 207 - 262). San Diego, CA: Academic Press.

Renner, B. (2003). Hindsight bias after receiving self-relevant health risk information: A motivational perspective. *Memory*, *11*(4), 455–472.

Roberts, K. P., & Powell, M. B. (2005). The relation between inhibitory control and children’s eyewitness memory. *Applied Cognitive Psychology*, *19*(8), 1003–1018.

Salthouse, T. a., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763–776.

Sanna, L. J., & Schwarz, N. (2007). Metacognitive Experiences and Hindsight Bias: It’s Not Just the Thought (Content) That Counts! *Social Cognition*, *25*(1), 185–202.

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology*, *85*(1), 33–46.

Schwarz, S., & Stahlberg, D. (2003). Strength of hindsight bias as a consequence of meta-cognitions. *Memory (Hove, England)*, *11*(4-5), 395–410.

Smith, H. D. (1999). Use of the anchoring and adjustment heuristic by children. *Current Psychology*, *18*(3), 294–300.

Stahlberg, D. (1994). *Der Knew-it-all-Along-Effekt: Eine urteilstheoretische Erklärung.* Unpublished Habilitationsschrift. University of Kiel.

Stahlberg, D. & Maass, A. (1997). Hindsight bias: Impaired memory or biased reconstruction? *European Review of Social Psychology, 8*, 105 - 132.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social …*, *73*(3), 437–446.

Tykocinski, O. E., Pick, D., & Kedmi, D. (2002). Retroactive pessimism : a different kind of hindsight bias. *European Journal of Social Psychology, 32*, 577–588.

Wasserman, D., Lempert, R. O. & Hastie, R. (1991). Hindsight and causality. *Personality and Social Psycholgoy Bulletin, 17* (1), 30 - 35.

Werth, L. (2002). Certainty and Uncertainty: The Two Faces of the Hindsight Bias. *Organizational Behavior and Human Decision Processes*, *87*(2), 323–341.

Winman, A., Juslin, P. & Bjorkman, 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* (2), 415 - 431.

Wood, G. (1978). The knew-it-all-along effect. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(2), 345–353.

Appendix A

Raw Response Category Frequencies for Exp. 3

Response Category	Low depletion				High depletion			
	Ctrl		Exp		Ctrl		Exp	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
ROJ<OJ<CJ	90	67	61	59	78	82	53	54
ROJ=OJ<CJ	108	136	111	107	161	153	142	147
OJ<ROJ<CJ	85	95	98	133	69	69	101	110
OJ<ROJ=CJ	1	3	0	7	0	4	3	3
OJ<CJ<ROJ	36	40	51	50	25	30	20	30
CJ<OJ<ROJ	48	53	29	30	41	36	16	26
CJ<ROJ=OJ	68	62	67	60	97	87	94	76
CJ<ROJ<OJ	62	33	86	59	36	39	72	62
ROJ=CJ<OJ	0	3	2	5	0	3	2	3
ROJ<CJ<OJ	45	46	36	37	27	31	34	24

Appendix B

Raw Response Category Frequencies for Exp. 4

Response Category	Low depletion						High depletion					
	Ctrl.		CCJ		WCJ		Ctrl.		CCJ		WCJ	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
ROJ<OJ<CJ	110	103	45	56	72	75	90	99	55	70	75	65
ROJ=OJ<CJ	193	178	209	164	185	162	202	217	194	167	197	171
OJ<ROJ<CJ	86	109	210	215	153	151	103	105	166	163	145	162
OJ<ROJ=CJ	2	4	5	7	13	21	3	2	2	6	7	3
OJ<CJ<ROJ	28	42	26	38	48	67	35	26	27	38	43	34
CJ<OJ<ROJ	72	50	27	22	31	28	66	51	22	29	44	35
CJ<ROJ=OJ	132	99	105	99	108	113	108	97	120	107	120	107
CJ<ROJ<OJ	39	56	80	92	80	78	56	63	88	114	93	99
ROJ=CJ<OJ	2	2	6	5	8	6	4	1	2	1	8	6
ROJ<CJ<OJ	39	56	60	65	56	45	40	48	75	62	38	70

Appendix C

HB bias items used in Experiment 3 and 4

1. Wie alt wurde Goethe? 82 Jahre
2. Wie lang war die laengste nachgewiesene menschliche Hand? 32m
3. Wie gross ist das Volument des Vierwaldstaettersees? 12km^3
4. Wie lange dauert die Schwanerschaft von Fuechsen? 52 Tage
5. Wie hoch ist die Freiheitsstatue (mit Sockel)? 93m
6. Wie viele Kapitel haben die drei “Herr der Ringe” Baende (ohne Anhaenge) insgesamt? 62
7. Wie viele Meter ueber dem Meeresspiegel liegt die Ortsmitte von Duesseldorf? 36m
8. Wie lang ist die kuerzere Seite eines Zehn-Euro Scheins? 67mm
9. Wie alt wurde Mozart? 35 Jahre
10. Wie viel kg Gefluegel verzehrt ein Bundesbuerger pro Jahr? 18kg
11. Wie lang war das laengste nachgewiesene menschliche Beinhaar? 12cm
12. Wie hoch war der hoechste Sprung eines Schweins? 70cm
13. Wie gross ist das Volumen des Bodensees? 48 km^3
14. Wie lange dauert die Schwangerschaft von Koalas? 35 Tage
15. Auf welchem Laengengrad liegt Budapest? 19
16. Wie viele Felder hat ein Mensch-aergere-dich-nicht-Spiel fuer 4 Personen (inclusive Start- und Zielfeldern)? 72
17. Wie lang kann der Gemeine Regenwurm warden? 30cm
18. Bis zu wie viele Fluegelschlaege macht ein Kolibri pro Sekunde? 50
19. Wie hoch war der Pkw- Bestand in Deutschland in Januar 2004? 45 Millionen
20. In welchem Jahr wurde Nelson Mandela geboren? 1918
21. Wie viele kg Zucker verbrauchte jeder Deutsche durchschnittlich in 2000? 33
22. Wie lang ist das laengste Fahhrad der Welt? 28m
23. Wie viele Junge wirft ein Wildkaninchen durchschnittlich pro Jahr? 30
24. Wie viele Zaehne hat das bleibende Gebiss eienr Hauskatze? 30
25. In Welchem Jahr starb Mahatma Gandhi? 1948
26. Wie hoch kann eine Fichte warden? 60m
27. Was ist die hoechste Temperatur die in den USA gemessen
28. Wie lang kann eine Bachforelle am höchstens werden? 40 cm
29. Eine normale Spitzmaus wiegt *30 Gramm*.

30. Ein Elefant kann *75 Jahre* alt werden.
31. Ein normales Skateboard ist *65 Zentimeter* lang.
32. Ein Liter Meerwasser enthält *40 Gramm* Salz.
33. Der Schwanz eines Eichhörnchens ist *20 Zentimeter* lang.
34. Der älteste Esel aller Zeiten wurde *54 Jahre* alt.
35. Ein Tiger nimmt *18 Kilogramm* Nahrung zu sich.
36. Ein Elefant ist *21 Monate* schwanger.
37. Die Grenze zwischen Deutschland und Dänemark ist *67 Kilometer* lang.
38. Ein Klavier hat *88 Tasten*.
39. Die längste Zeit, die eine Königin je regiert hat, ist *63 Jahre*.
40. Die größte Kaugummiblase war im Durchmesser *58 Zentimeter* groß.
41. Die Körpertemperatur von Vögeln beträgt *42 Grad Celsius*.
42. Eine Biene kann eine Strecke von *18 Kilometer* in der Stunde fliegen.
43. Der längste Eisenbahntunnel der Welt ist *54 Kilometer* lang.
44. Die menschliche Hand besteht aus *18 Knochen*.
45. Die Erde legt in 3 Sekunden eine Strecke von *91 Kilometern* zurück.
46. Es gibt *54 verschiedene* Känguruarten.
47. Ein normaler Marsriegel wiegt *54 Gramm*.
48. Ein Bobbycar ist *60 Zentimeter* lang.
49. In einer 1-Liter-Flasche Coca Cola sind *36 Stück* Würfelzucker.
50. Das Herz des Vogels Strauß schlägt *65 mal* pro Minute.
51. Eine Buche kann *40 Meter* hoch werden.
52. Die schnellste Weltumsegelung dauerte *73 Tage*.
53. Seepferdchen können in einer Stunde *15 Meter* zurücklegen.
54. Ein Feuerwehrlöschzug besteht aus 22 Feuerwehrleuten.
55. Die längste Zeit, die ein Mensch ohne Nahrung und Wasser überlebte, war *18 Tage*.
56. Der gesehene Kopfsalat hatte *56 Blätter*.
57. Man darf ab dem *35 Lebensjahr*, Präsident der USA werden.
58. Ein Pottwal wird *20 Meter* lang.
59. Ein erwachsener Schäferhund hat *42 Zähne*.
60. Ein Bambus kann *38 Meter* hoch werden.
61. Nach *14 Wochen* ist aus einer Kaulquappe ein Frosch geworden.
62. Ein Murmeltier wird höchstens *40 Zentimeter* lang.
63. Es sind bisher *12 Menschen* auf dem Mond gelandet.

64. Dinosaurier sind seit *65 Millionen Jahren* ausgestorben.
65. Es gibt *53 Länder* in Afrika.
66. Ein Skorpion wird höchstens *18 Zentimeter* lang.
67. Alle drei Pippi Langstrumpf Bücher haben insgesamt *32 Kapitel*.
68. Die Drehleiter eines Feuerwehrwagens kann bis zu *30 Meter* ausgefahren werden.
69. Man kann ab *12 Jahren*, Mitglied bei der Jugendfeuerwehr werden.
70. Eine Kreuzotter kann *80 Zentimeter* lang werden.
71. In einer normalen Tube UHU befinden sich *31 Gramm* Klebstoff.
72. Der erste bemannte Flug eines Heißluftballons dauerte *25 Minuten*.
73. Wie groß ist der Durchmesser einer normalen CD? 12cm
74. Wie lang ist die Strecke zwischen Düsseldorf und Duisburg? 34km
75. Wie lang sind die Füße von Menschen mit einer Schuhgröße von 45? ca. 28cm
76. Wie lang ist der Darm eines erwachsenen Menschen? Ca. 8m
77. Wie viel Volumen hat der Aralsee zwischen 1960 und 2008 verloren? 90%
78. Wie viele Meter über dem Meeresspiegel liegt die Ortsmitte von Hannover? 55m
79. Bei welcher Temperatur schmilzt Phosphor? 44°C
80. Wie hoch ist der Sauerstoffanteil in der Erdatmosphäre? Ca. 21%
81. Wie lange hat Steve Fossett 2005 für eine Non-Stop Weltumrundung mit einem Flugzeug benötigt? 67 Stunden
82. Um wie viele Prozentpunkte ist der Amerikanische Aktienindex am 28.10.1929 („schwarzer Montag") gefallen? ca. 13%
83. Wie lange war die russische Raumstation Mir in Betrieb? 15 Jahre
84. Wir groß ist der Durchmesser eines 2€ Stücks? 26mm
85. Welchen Ranplatz hat der international Flughaften von San Francisco auf der Liste der verkehrsreichsten Flughaefen der USA? 21
86. Was ist die Rekordlaenge einer Westlichen Diamant Klapperschlange? 84 Inch
87. Welche Flaeche bedeck Daenemark? 42 tausend km^2
88. Wie viel Saeulen hat das Parthenon? 86