Acta Psychologica 154 (2015) 77–85

Contents lists available at ScienceDirect

Acta Psychologica

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

How does aging affect recognition-based inference? A hierarchical Bayesian modeling approach☆

Sebastian S. Horn [a,*,1], Thorsten Pachur [a,1], Rui Mata [a,b,1]

[a] Max Planck Institute for Human Development, Berlin, Germany
[b] University of Basel, Switzerland

ARTICLE INFO

Article history:
Received 16 June 2014
Received in revised form 16 October 2014
Accepted 3 November 2014
Available online 17 December 2014

PsycINFO classification:
2340
2860

Keywords:
Aging
Recognition memory
Heuristics
Strategy selection
Hierarchical Bayesian modeling

ABSTRACT

The recognition heuristic (RH) is a simple strategy for probabilistic inference according to which recognized objects are judged to score higher on a criterion than unrecognized objects. In this article, a hierarchical Bayesian extension of the multinomial *r*-model is applied to measure use of the RH on the individual participant level and to re-evaluate differences between younger and older adults' strategy reliance across environments. Further, it is explored how individual *r*-model parameters relate to alternative measures of the use of recognition and other knowledge, such as adherence rates and indices from signal-detection theory (SDT). Both younger and older adults used the RH substantially more often in an environment with high than low recognition validity, reflecting adaptivity in strategy use across environments. In extension of previous analyses (based on adherence rates), hierarchical modeling revealed that in an environment with low recognition validity, (a) older adults had a stronger tendency than younger adults to rely on the RH and (b) variability in RH use between individuals was larger than in an environment with high recognition validity; variability did not differ between age groups. Further, the *r*-model parameters correlated moderately with an SDT measure expressing how well people can discriminate cases where the RH leads to a correct vs. incorrect inference; this suggests that the *r*-model and the SDT measures may offer complementary insights into the use of recognition in decision making. In conclusion, younger and older adults are largely adaptive in their application of the RH, but cognitive aging may be associated with an increased tendency to rely on this strategy.

1. Introduction

A central tenet of Herbert Simon's (1956) time-honored concept of bounded rationality is that the use of simple mental tools—if attuned to the environmental structure—can often lead to surprisingly good decisions. Consider the following question: Which city is has more residents, Nashville or Tulsa? As a prime example of a frugal inference strategy, the recognition heuristic (RH; Goldstein & Gigerenzer, 2002) assumes that individuals base their judgments in this sort of task solely on whether or not the options are recognized (and ignore any further knowledge). That is, the RH predicts that a recognized object (e.g., Nashville) has a higher value on the criterion (city population) than an unrecognized one (e.g., Tulsa). The success of this simple strategy depends on its ecological rationality: it exploits the phenomenon that known objects differ from unknown ones in systematic ways in many natural environments (e.g., larger cities, more successful athletes, and higher mountains tend to be recognized more often; Pachur, Todd, Gigerenzer, Schooler, & Goldstein, 2011, 2012).

One key issue surrounding research on the RH is to what extent it is used *adaptively*: When do people use the RH and how do they adjust their reliance across different situations (Gigerenzer & Goldstein, 2011; Pachur et al., 2011)? Research indicates that *younger* adults are largely sensitive to characteristics of the environment, such as the relation between recognition of an object and the criterion (the recognition validity),[2] time pressure, or available cognitive resources (e.g., Pachur & Hertwig, 2006; Pohl, Erdfelder, Hilbig, Liebke, & Stahlberg, 2013). However, there is also considerable diversity across people in reliance on the RH, suggesting that individual-level variables may moderate its use

☆ This research was supported by a fellowship from the Max Planck Society to Sebastian Horn.

* Corresponding author at: Center for Adaptive Rationality (ARC), Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Tel.: +49 30 82406 202.
E-mail address: sebastian.horn@mpib-berlin.mpg.de (S.S. Horn).

[1] Sebastian S. Horn, Thorsten Pachur, and Rui Mata, Center for Adaptive Rationality (ARC), Max Planck Institute for Human Development, Berlin, Germany. Rui Mata is now at the University of Basel, Institute for Psychology, Department for Cognitive and Decision Sciences, Basel, Switzerland.

[2] The predictive power of recognition is usually quantified in terms of the *recognition validity* α (Goldstein & Gigerenzer, 2002). For a given environment, it is calculated as $\alpha = C_{RU} / (C_{RU} + I_{RU})$, where C_{RU} and I_{RU} are frequencies of correct and incorrect inferences, respectively, that the recognition heuristic would predict across all trials in which *one* of the objects is recognized (*RU* cases). *Knowledge validity* β is calculated as $\beta = C_{RR} / (C_{RR} + I_{RR})$, where C_{RR} and I_{RR} are correct and incorrect inferences in cases in which *both* objects are recognized (*RR* cases).

http://dx.doi.org/10.1016/j.actpsy.2014.11.001
0001-6918/

(Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2010; Pachur, Bröder, & Marewski, 2008; cf. Hilbig & Pohl, 2008).

Here, we investigate one potential source of individual differences in the use of RH: cognitive aging. Aging is associated with significant changes on various psychological dimensions that may impact the use of the RH, including decrements in fluid cognitive abilities and increments in knowledge and experience (e.g., Baltes, Staudinger, & Lindenberger, 1999). Pachur, Mata, and Schooler (2009) examined this issue by asking younger and older adults to make inferences regarding pairs of cities ("Which city has more inhabitants?") and infectious diseases ("Which disease has a higher incidence rate?"). Referring to *environment* or *task adaptivity* as people's sensitivity to differences in recognition validity between domains (i.e., cities vs. diseases), they found that participants chose recognized objects more frequently over unrecognized ones in a domain with high recognition validity (cities) than in a domain with low recognition validity (diseases). The proportion of choices of the recognized object—the *adherence rate*—did not differ between younger and older adults: Both age groups showed a similarly lower adherence rate in the environment with low recognition validity than in the environment with high recognition validity. This result may seem surprising: Other research has found that older adults generally tend to rely more on simple strategies (e.g., the take-the-best heuristic; Gigerenzer & Goldstein, 1996) than younger adults, even in environments in which another strategy may be more appropriate (e.g., Mata, Schooler, & Rieskamp, 2007). Importantly, however, Hilbig, Erdfelder, and Pohl (2010) pointed out that adherence rates might be an inappropriate measure of people's reliance on the RH and proposed a multinomial processing tree (MPT) model—the *r*-model—as a more valid measurement approach to recognition-based inference.

Our goal in this article is to re-evaluate possible age differences as well as individual variability in the use of the RH in the Pachur et al. (2009) study using the *r*-model. We extend on previous applications of the *r*-model by implementing a hierarchical approach, which is particularly suitable for studying individual differences. A secondary goal is to explore how parameters of the *r*-model relate to measures of people's use of recognition and further knowledge derived from signal-detection theory (SDT; e.g., Macmillan & Creelman, 2005): Pachur et al. proposed a discriminability index to measure how well people can distinguish between cases where recognition leads to correct versus incorrect inference. On this SDT index, older adults showed lower discriminability than younger adults when inferring which of two diseases was more frequent in Germany.[3] It is currently unclear to what extent SDT measures of recognition-based inference and the parameters of the *r*-model offer alternative or complementary perspectives on how younger and older adults use recognition to make inferences about the world. Next, we describe in greater detail the multinomial approach to measuring use of the RH, followed by the hierarchical extension of the *r*-model.

1.1. The *r*-model: a multinomial processing tree approach to measuring RH use

Many investigations of the RH—including Pachur et al.'s (2009) aging study—have relied on adherence rates as a measure of people's use of this strategy (e.g., Goldstein & Gigerenzer, 2002; McCloy, Beaman, Frosch, & Goddard, 2010). Yet, as Hilbig, Erdfelder, et al. (2010) have emphasized, adherence rates may lead to erroneous conclusions regarding the use of the RH: In many natural environments, not only recognition but also other knowledge is correlated with the criterion. The choice of a recognized object may therefore also be due to reliance on other knowledge, which according to the RH is ignored. Consequently, "the adherence or accordance rate … is not a valid measure of use of the RH versus incorporation of further knowledge, because recognition and knowledge are necessarily confounded" (p. 123). To address this issue, Hilbig and colleagues introduced the *r*-model, an MPT model, to disentangle pure reliance on the RH from the use of further information (for a critical discussion of the use of this model, see Pachur, 2011). The strength of the MPT modeling framework is that it provides a foundation for improved measurement of cognitive components underlying a task (cf. Jacoby, 1991) and a well-developed statistical machinery for model comparison and goodness-of-fit tests (for overviews, see Batchelder & Riefer, 1999; Erdfelder et al., 2009).

As is generally the case in MPT models (Batchelder & Riefer, 1999), the *r*-model accounts for the frequency of responses in different outcome categories through combinations of a set of latent parameters, assuming that the data follow a multinomial distribution. The parameters are estimated simultaneously in such a way that a loss function (e.g., G^2) between the observed and predicted categorical frequencies is minimized. The *r*-model considers three possible cases in a comparative judgment task, represented by the $J = 3$ separate trees in Fig. 1: In the upper tree, a decision maker recognizes both objects (*RR* case) and therefore has to recruit further information beyond recognition, leading to a correct inference with probability b and to an incorrect inference with complementary probability $1 - b$. Parameter b thus indexes the validity of the decision maker's further knowledge (comparable to *knowledge validity* β in the original RH formulation, see footnote 2; a visualization of the relationship between the recognition and knowledge validities α and β and the estimates for the a and b parameters for the present data are provided in the online supplemental materials). The second tree represents the situation in which one of the two objects is recognized (*RU* case) and the RH can thus be applied. With probability r, the decision maker uses the RH and chooses the recognized item. This leads to a correct inference with probability a and to an incorrect inference with probability $1 - a$. Parameter a is thus conceptually equivalent to the *recognition validity* α (footnote 2) and reflects the strength of association between recognition and the criterion variable (e.g., city size). Importantly, with complementary probability $1 - r$, the RH is not applied and the inference is based on further information beyond recognition (or any other strategy). This leads to a correct inference with probability b. In this case, the recognized object is chosen with probability a and the unrecognized object is chosen with probability $1 - a$. With probability $1 - b$, the inference is incorrect. In this case, the *unrecognized* item is chosen with probability a and the recognized item is chosen with probability $1 - a$. The model thus acknowledges that the observed choice of a recognized object (outcome categories C_{21} and C_{22}) may result from the use of the RH (upper two branches of the *RU* tree) or, alternatively, from the use of further knowledge or another strategy (lower branches in the *RU* tree). In the bottom tree, neither of the objects is recognized (*UU* case) and the decision maker has to guess, leading to a correct inference with probability g and to an incorrect inference with probability $1 - g$.[4]

Several studies have shown that estimated reliance on the RH is lower when using the *r*-model than when using adherence rates (Hilbig, Erdfelder, et al., 2010; Hilbig & Richter, 2011). In fact, the two measures sometimes even lead to different conclusions: Using the *r*-model, Pohl et al. (2013) found that cognitive depletion led to greater reliance on the RH. In contrast, they found no reliable difference in the adherence rates of depleted versus nondepleted participants. Such results call into question the use of adherence rates as a measure of RH use. For this

[3] Note that discriminability between cases where the RH leads to a correct vs. incorrect inference (including the ability to *suspend* the RH on specific trials) implies the use of further information, beyond mere recognition. Discriminability has been assumed to correlate with fluid abilities (Pachur et al., 2009); moreover, RH-inconsistent decisions take longer than RH-consistent decisions (Pachur & Hertwig, 2006) and are associated with evaluative frontal brain activation (Volz et al., 2006). These findings suggest that discriminability may incur cognitive costs.

[4] In principle, unique values for a, b, and r could be obtained with a single-tree model for *RU* trials only; however, this approach would not provide any degrees of freedom for testing goodness of fit. By considering the other possible cases in the comparative judgment task and by constraining b to be equal across the *RR* and *RU* cases (i.e., by assuming that the probability of valid knowledge is the same across these situations, $b_{RR} = b_{RU}$; for further discussion, see Castela, Kellen, Erdfelder, & Hilbig, 2014), the model is testable with a χ^2 statistic with $df = 1$ and comprises a set of four free parameters $\boldsymbol{\theta} = (a, b, g, r)$ and eight outcome categories, five of which are independent.