Journal of Mathematical Psychology 55 (2011) 8–24

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

A Bayesian hierarchical model for the measurement of working memory capacity

Richard D. Morey

DMPG, Grote Kruisstraat 2/1, 9712NS Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 31 January 2010
Received in revised form
13 May 2010
Available online 1 October 2010

Keywords:
Hierarchical modeling
Bayesian methods
Hybrid Monte Carlo
Working memory
Working memory capacity
Multinomial models

ABSTRACT

Working memory is the memory system that allows for conscious storage and manipulation of information. The capacity of working memory is extremely limited. Measurements of this limit, and what affects it, are critical to understanding working memory. Cowan (2001) and Pashler (1988) suggested applying multinomial tree models to data from change detection paradigms in order to estimate working memory capacity. Both Pashler and Cowan suggested simple formulas for estimating capacity with these models. However, in many cases, these simple formulas are inadequate, and may lead to inefficient or biased estimation of working memory capacity. I propose a Bayesian hierarchical alternative to the Pashler and Cowan formulas, and show that the hierarchical model outperforms the traditional formulas. The models are easy to use and appropriate for a wide range of experimental designs. An easy-to-use graphical user interface for fitting the hierarchical model to data is available.

Human memory is often described as being composed of separate systems (Atkinson & Shiffrin, 1968; James, 1890). One of these systems, called working memory (WM), is the memory system which allows for conscious storage and recall of information, within the span of several seconds (Baddeley & Hitch, 1974; Cowan, 2001). The use of working memory is fundamental to many everyday tasks, including reading, driving, having conversations, and reasoning. As an example, consider that when you read you only read a few words at a time. As you read a paragraph, you must temporarily store words, word order, and ideas, and piece the words and ideas together in a coherent manner. Without the ability to temporarily store and manipulate information, reading would be impossible.

The storage component of working memory is often thought of as a buffer that allows for the rehearsal and manipulation of information from a variety of sources, including sensory stores or long-term memory. Working memory capacity is known to be extremely limited. Miller (1956) estimated the average capacity of human working memory as seven items; more recently Cowan (2001) has estimated average capacity at about four items. The limit of working memory capacity is a fundamental aspect of our conscious experience, which is governed by the fact that we apprehend only a small proportion of the information gathered by our senses. The ability to measure participants' individual WM capacity is at the center of several important areas of research in psychology. For instance, WM capacity is correlated with reading comprehension (Daneman & Carpenter, 1980) and measures of aptitude (Cowan et al., 2005). WM capacity estimates are central to the research claiming that WM resources are modality general rather than separate (Morey & Cowan, 2005). Measurement models of WM capacity have also been used to argue for discrete, all-or-none representation of items in WM (Rouder et al., 2008b). For each of these areas of research, the measurement of WM capacity is a critical part of the endeavor. In this paper, I will review current methods for estimating working memory capacity, and some difficulties with these methods. I will then present a hierarchical model which does not suffer from these difficulties, and show that the hierarchical model recovers parameters well in simulation. Following the simulations, I will apply the hierarchical models to the data set of Rouder et al. (2008b), showing that in a visual array change detection task, attention to the task is highly correlated with capacity, and different colors vary in the guessing and encoding biases they instantiate.

1. Estimating working memory capacity

In the working memory literature, there are a number of conceptions of working memory. They largely fall into two classes: resource-based models (Alvarez & Cavanagh, 2004; Bays & Husain, 2008) and item-based models (Cowan, 2001; Luck & Vogel, 1997). The difference between the two viewpoints is whether they treat the item as the atomic unit of working memory capacity. In item-based models, each item takes up the same amount of space regardless of the number of features the item may have. Items can be thought of like physical objects that take up discrete slots in memory, and storage is either all-or-none; partial storage does not occur. In a resource-based model, properties of items, such as the number of features, are primary. Complex items may take up more

E-mail address: r.d.morey@rug.nl.

0022-2496/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2010.08.008

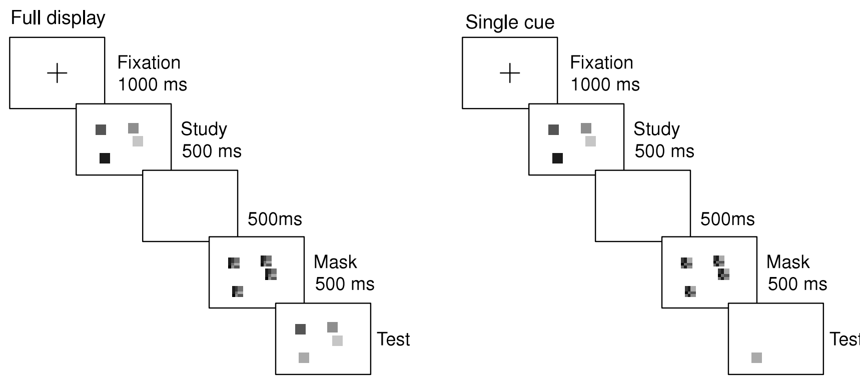

Fig. 1. The two change detection paradigms to be modeled.

storage in working memory than simple items, and partial storage (or a change in resolution of storage) may occur. In this paper, I will address the item-based models of Cowan (2001) and Rouder et al. (2008b). The item-based view is assumed throughout this manuscript.

The most common method of estimating WM capacity is by means of the change detection task. In a typical change detection task, a number of stimuli are presented for study. After a brief interval, the stimuli are presented again, either exactly the same as at study, or with one item different. The participant's task is to detect the change and respond with either "same" or "different". The visual array paradigm shown in Fig. 1(left) is one such popular change detection task (Luck & Vogel, 1997; Phillips, 1974; Wheeler & Treisman, 2002). The stimuli are colored squares, one of which may change color from the study to the test array. Because the entire array is displayed at test, we may call this design the *full-display* design.

Pashler (1988) formalized the simple item-based model described above in order to estimate WM capacity using the full-display paradigm. Pashler assumed that participants had K spaces in memory, each of which could store a single item. If a participant is presented with N items to remember, only K of these items can be encoded in working memory; the remaining items are discarded for lack of space, and not encoded in working memory. Encoding occurs for K items, in an all-or-none fashion. Although the model is quite simple, Pashler (1988) and Cowan (2001) advocated its use in measuring working memory capacity. The formulas that Pashler, and later Cowan, suggested for estimating working memory capacity have become standard tools for working memory researchers. In the next two sections, I will present Pashler's and Cowan's approaches to estimating working memory capacity, which serve as the foundation for the hierarchical model developed herein.

1.1. Pashler's formula

Pashler (1988) suggested a simple formula to estimate working memory capacity in the full-display design. Let N be the total number of items in the study array, and K be the participant's capacity. Under the item-based view, an item either occupies a slot in WM, or it is not encoded. If the display contains one changed item, the probability that it occupies an available memory slot is K/N if the number of items exceeds the set size. If capacity exceeds the number of items, all items are stored and the probability that the changed item is stored is 1.0. Thus, the participant detects the change with probability $D = \min(K/N, 1)$. If the participant does not detect the change, they enter a state of uncertainty, guessing that a change occurred with probability G. This leads to a hit rate H of

$$H = D + (1 - D)G.$$

If the display did not change and the number of items exceeds the participant's capacity, the participant is unsure of whether a change occurred. A change may have occurred in one of the items not stored, or there may have been no change. Again the participant enters a guessing state, guessing "change" with probability G. However, if the participant has a capacity greater than the array set size, the participant is able to store the entire array, and the false alarm rate is 0. The probability of a false alarm F is

$$F = G'$$
$$G' = \begin{cases} G & K < N \\ 0 & K \geq N. \end{cases}$$

The parameter G' represents the "effective" guessing rate. G' equals G in the case that capacity is less than the set size, leading to uncertainty about whether there was a change. If there is no uncertainty about a change because entire array can be stored, G' is 0.[1] The trees resulting from formalizing Pashler's model for the full-display are depicted in the first row of Fig. 2.

Assuming that $K < N$, it is possible to solve for K and G to obtain estimates:

$$\hat{K}_p = N\left(\frac{\hat{H} - \hat{F}}{1 - \hat{F}}\right) \quad K < N, \tag{1}$$
$$\hat{G}_p = \hat{F} \quad K < N,$$

where $\hat{H}$ and $\hat{F}$ are the observed hit and false alarm rates, respectively. The formula for $\hat{K}_p$ may be called the "Pashler formula". Note that the formula only yields good capacity estimates when $K < N$; if $K \geq N$, performance is perfect and no information can be gained about WM capacity. In this case, the formula simply yields the set size.

1.2. Cowan's formula

Cowan (2001) noted that in some experimental designs, Pashler's formula for estimating working capacity is inappropriate.

[1] Pashler did not mention the case when $K \geq N$, perhaps because the experiment he considered had a set size of 16, which is well above any measured human WM capacity. Nevertheless, recent experiments using the full-display methodology have used lower set sizes (see, for example, Lin and Luck (2008), with set sizes of 3 and 4). The formula makes little sense without restriction when $K \geq N$.