Journal of Applied Research in Memory and Cognition 3 (2014) 72–76

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

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

Improved classification of mammograms following idealized training

Adam N. Hornsby, Bradley C. Love *

Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

ARTICLE INFO

Article history:
Received 2 February 2014
Received in revised form 22 April 2014
Accepted 24 April 2014
Available online 2 May 2014

Keywords:
Categorization
Memory retrieval
Decision making
Mammograms
Idealization
Medical diagnosis

ABSTRACT

People often make decisions by stochastically retrieving a small set of relevant memories. This limited retrieval implies that human performance can be improved by training on idealized category distributions (Giguère & Love, 2013). Here, we evaluate whether the benefits of idealized training extend to categorization of real-world stimuli, namely classifying mammograms as normal or tumorous. Participants in the idealized condition were trained exclusively on items that, according to a norming study, were relatively unambiguous. Participants in the actual condition were trained on a representative range of items. Despite being exclusively trained on easy items, idealized-condition participants were more accurate than those in the actual condition when tested on a range of item types. However, idealized participants experienced difficulties when test items were very dissimilar from training cases. The benefits of idealization, attributable to reducing noise arising from cognitive limitations in memory retrieval, suggest ways to improve real-world decision making.

1. Introduction

Classifying mammograms as tumorous vs. non-tumorous is a complex, probabilistic, and error-prone task. When performing such tasks, one possibility is that people selectively and stochastically retrieve relevant memories to guide their decisions (Giguère & Love, 2013; Nosofsky & Palmeri, 1997). Unfortunately, selectively sampling memory introduces noise at the time of decision that results in suboptimal performance (Giguère & Love, 2013).

Recently, there has been interest in tailoring training conditions to promote better test performance (Pashler & Mozer, 2013). For instance, Giguère and Love (2013) find that the harmful effects caused by limited memory retrieval at the time of decision can be reduced by training people on idealized distributions of category information. This idealization, which deemphasizes ambiguous cases, reduces the likelihood that misleading memories will be retrieved at the time of test. For example, in one study, two groups were trained on a random set of baseball games and asked to predict at test the outcomes of the remaining games for that season. The group trained on the actual outcomes of the games did not perform as well at test as the group trained on idealized game outcomes based on the total wins by teams in the training set.

Idealization may be complementary to other techniques that aim to improve human performance given limits in cognitive abilities. For example, manipulating the presentation order of training examples is another method to make the underlying structure of information more salient. Presentation orders that make the category structure more salient lead to better learning (Avrahami et al., 1997; Clapper & Bower, 1994; McClelland, Fiez, & McCandliss, 2002; Medin & Bettger, 1994). For example, people are better able to separate contrasting structures when a number of items from one category are presented together followed by a number of contrasting items from the other category (Clapper & Bower, 1994). These ordering effects, which isolate the categories (cf. Goldstone, 1996), make the contrasting structure evident. Other order manipulations emphasize presenting unambiguous cases (from either category) early in learning and only presenting the ambiguous cases later in learning after the learner is properly anchored (Avrahami et al., 1997; McClelland et al., 2002). These ordering manipulations that promote structure discovery share a kindred spirit with idealization. Whereas an advantageous item ordering makes the underlying category structure more apparent by strengthening contrast, idealization of category structures increases contrast by removing or altering ambiguous cases.

One key question is whether the idealization advantage extends to classification tasks that involve complex real-world stimuli. Extending to complex real-world stimuli would bring the idealization manipulation one step closer to useful application. Previous results in the literature suggest that idealization and the psychological theory underlying it should extend to real-world settings. Even

* Corresponding author. Tel.: +44 0207 679 1515.
E-mail addresses: adam.hornsby.10@alumni.ucl.ac.uk (A.N. Hornsby), b.love@ucl.ac.uk (B.C. Love).

http://dx.doi.org/10.1016/j.jarmac.2014.04.009
2211-3681/

when provided with explicit instruction, dermatological diagnoses are guided by similarity to experienced examples (Chan, Brooks, & Norman, 2001). In a simulated psychiatric diagnosis task, participants relied on easily accessible instances and their decisions were guided by the idiosyncratic properties of the training items (Young, Brooks, & Norman, 2011). These results align closely with Giguère and Love (2013) characterization of memory retrieval at the time of decision. To the extent that memory retrieval is limited to available (i.e., recent, familiar, and similar) instances, idealized training should improve test performance.

Here, we examine whether idealized training improves people's ability to classify novel mammograms as tumorous or non-tumorous (i.e. normal). In Experiment 1, we normed mammograms to determine the a priori ambiguity or difficulty level (easy, medium, or hard) of the images. In the main study, Experiment 2, a second group of participants were trained to classify mammograms using trial-and-error learning (i.e. stimulus → response → feedback). We correctly predicted that participants trained on an idealized distribution of mammograms (i.e. only including unambiguous easy cases) would be more accurate in classifying novel mammograms (across difficulty levels) at test than participants trained on a representative distribution of mammograms that included easy, medium, and hard items as in the test set. However, the results were nuanced in that the idealization advantage was strongest for images that were somewhat similar to those experienced during training.

2. Experiment 1

Unlike the simple stimuli typically used in category learning studies, mammograms are subtle, complex, and high-dimensional. These stimuli are not easily described in terms of basic stimulus dimensions (e.g., size, shape, and color) that are psychologically meaningful to participants. Rather than attempt to discover the dimensional structure of mammograms, which may be an intractable task and is likely not agreed upon across individuals, the goal of Experiment 1 is to norm mammograms to determine prior to training how likely people are to view an image as containing a tumor. These stimulus ratings are used in the main study, Experiment 2.

2.1. Method

2.1.1. Participants

One hundred participants were recruited using Amazon Mechanical Turk (mturk). Mturk (www.mturk.com) has been used for a wide variety of psychological studies and has been shown to be an inexpensive, fast, and reliable source of human data (Buhrmester, Kwang, & Gosling, 2011; Crump, McDonnell, & Gureckis, 2013). All participants were required to have had 90% or more of their previous mturk assignments approved. Eighteen participants were removed from the final analyses because they failed two or more of the catch trials (described below). The mean age of the final sample was 33.0 (SD = 28.9). Participants were from 12 different countries with 85.37% either from the USA or India. Participants were paid $.50 for participating, and were awarded an additional $.50 bonus for correctly responding in all five catch trials. This pay level is typical for mturk (Horton & Chilton, 2010).

2.1.2. Apparatus and stimuli

The study was designed using Adobe's ActionScript language and was accessed using Adobe Flash Player in a web browser. The task was presented in a black window, which was 600 × 450 pixels. Participants responded by clicking on a green 'Normal' button on the left of the window or a red 'Tumour' button on the right with their computer mouse. All mammograms were at a mediolateral oblique (MLO) angle, left-facing, taken with a 43.5 μm HOWTEK scanner, and presented in the lossless Portable Network Graphics (PNG) format. Example images are shown in Fig. 1.

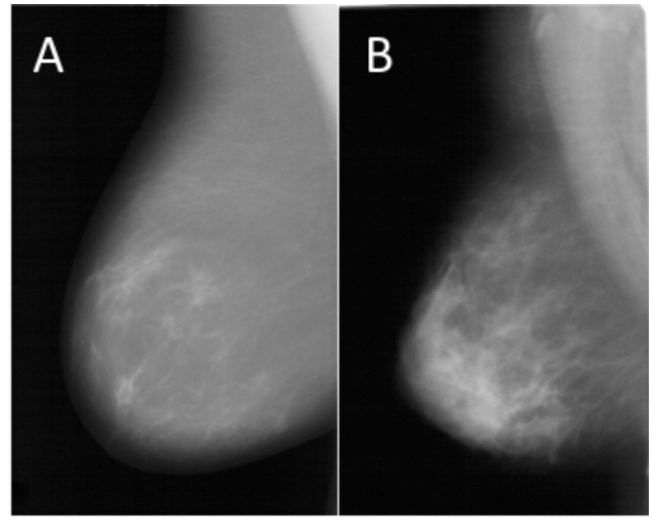

Fig. 1. A normal (A) and tumorous (B) mammogram.

2.1.3. Procedure

Participants were asked to enter their age, sex and location, and to confirm that they had no prior medical training or professional experience with classifying mammograms. On each of the 200 rating trials, 3 images were randomly selected from the bank of 358 possible images, subject to the constraint that the left image was normal, the right image was tumorous, and all 3 images were unique. The participant's task was to decide whether the central image was normal or contained a tumor. The three images were presented for 2000 ms before the response buttons appeared. Images were presented until a response was made and then a blank screen was shown for 1000 ms (no corrective feedback was provided). In addition to the 200 rating trials, there were five randomly interspersed catch-trials in which the image in the center was identical to one of the flanking images, which should make the correct response clear. A progress bar was displayed at the top of the screen and participants were fully debriefed at the end the study.

2.2. Results and discussion

Fig. 2 summarizes the rating data. The results confirm that images vary greatly in their a priori difficulty with some images being very misleading and hard to classify. The percentage of correct responses made to each image in the norming study (see Fig. 2)

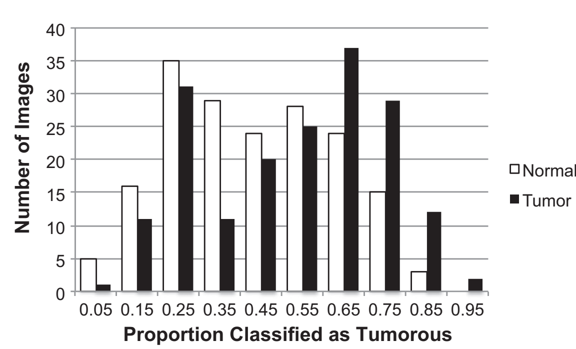

Fig. 2. Participants' distribution of tumor judgments in the norming task for normal (white) and tumorous (black) mammograms. Each bin is centered on the value shown beneath it.