Acta Psychologica 155 (2015) 8–18

Contents lists available at ScienceDirect

Acta Psychologica

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

Monitoring of learning at the category level when learning a natural concept: Will task experience improve its resolution?☆

Sarah K. Tauber [a,*], John Dunlosky [b]

[a] Department of Psychology, Texas Christian University, USA
[b] Department of Psychology, Kent State University, USA

ARTICLE INFO

Article history:
Received 29 May 2014
Received in revised form 24 November 2014
Accepted 27 November 2014
Available online xxxx

PsycINFO classification:
2343 Learning & Memory
2340 Cognitive Processes
2260 Research Methods & Experimental Design

Keywords:
Monitoring
Category learning judgment
Concept formation
Metacognition

ABSTRACT

Researchers have recently begun to investigate people's ability to monitor their learning of natural categories. For concept learning tasks, a learner seeks to accurately monitor learning at the *category level* — i.e., to accurately judge whether exemplars will be correctly classified into the appropriate category on an upcoming test. Our interest was in whether monitoring resolution at the category level would improve as participants gain task experience across multiple study-test blocks, as well as within each block. In four experiments, exemplar birds (e.g., American Goldfinch, Cassin's Finch) paired with each family name (e.g., Finch) were studied, and participants made a judgment of learning (JOL) for each exemplar. Of most interest, before and after studying the exemplars, participants made category learning judgments (CLJs), which involved predicting the likelihood of correctly classifying novel birds into each family. Tests included exemplars that had been studied or exemplars that had not been studied (novel). This procedure was repeated for either one or two additional blocks. The relative accuracy of CLJs did not improve across blocks even when explicit feedback was provided, whereas item-by-item JOL accuracy improved across blocks. Category level resolution did improve from pre-study to post-study on an initial block, but it did not consistently increase within later blocks. The stable accuracy of CLJs across blocks poses a theoretical and empirical challenge for identifying techniques to improve people's ability to judge their learning of natural categories.

Imagine people trying to learn to identify new birds by their appropriate family, such as deciding whether a bird is a Sparrow or a Finch. To do so, they may study exemplars with the goal of being able to correctly classify new birds into the same family. New bird watchers may study images of exemplars in a field guide book (e.g., an image of a Chipping Sparrow or a White-throated Sparrow) so as to classify never-before-seen birds from these families in the wild (e.g., identify a Song Sparrow as a Sparrow). They may also try to monitor their learning by judging how well they can classify exemplars in each category. For example, people who are learning to classify birds may judge how well they can identify different bird families, such as judging how well they have learned to identify Finches, Sparrows, Grosbeaks, and so on. As another example, medical students may be learning to make different electrocardiogram diagnoses and judge how well they can identify myocardial infarctions, ventricular hypertrophy, ischemia, and so on. In these and other applied contexts, accurate monitoring at the category level involves discriminating between concepts that have been well learned versus those that have been less well learned.

Moreover, accurately monitoring how well one has learned categories and can classify new instances has the potential of improving people's study decisions (e.g., to focus future study on the least well-learned categories), which in turn can contribute to better learning and classification (cf. Dunlosky & Rawson, 2012; Metcalfe & Finn, 2008; Tauber & Rhodes, 2010; Thiede, 1999; Thiede et al., 2003). Given the importance of understanding how people monitor their learning at the category level, researchers have recently begun to investigate people's category level judgments for learning concepts (Jacoby et al., 2010; Rawson, Thomas & Jacoby, 2014; Wahlheim et al., 2012). This research has not focused on CLJs and has largely just demonstrated that people's CLJs show above-chance accuracy (but see Wahlheim et al., 2012). By contrast, the present research systematically explores one factor — task experience across study-test blocks — that promises to improve monitoring resolution at the category level and that will provide foundations for developing theory of CLJs and CLJ accuracy. Concerning the latter, we evaluated a number of hypotheses about factors that were expected to boost CLJ resolution, such as increasing the amount of task experience and providing test feedback. We consider these factors prior to the relevant experiments and now turn to describing our approach to explore whether task experience improves CLJ resolution.

☆ This research was supported by the James S. McDonnell Foundation (220020166) 21st Century Science Initiative in Bridging Brain, Mind, and Behavior Collaborative Award.

* Corresponding author at: Department of Psychology, Texas Christian University, TCU Box 298920, 2800 S. University Dr., Fort Worth, TX 76129, USA.
E-mail address: uma.tauber@tcu.edu (S.K. Tauber).

http://dx.doi.org/10.1016/j.actpsy.2014.11.011
0001-6918/

To explore the impact of task experience, we adapted methods used by Wahlheim et al. (2011) in which participants studied exemplars paired with the bird family name (cf. phases 2–4 of Fig. 1). Participants were presented with images of exemplar birds (e.g., an image of a Chipping Sparrow) along with the family name (e.g., Sparrow). During study, exemplars were either grouped together by family (i.e., blocked condition) or not (i.e., interleaved condition). Participants studied 10 individual exemplars from 12 bird families after which they made 12 category learning judgments (CLJs) — one for each family of birds (cf. post-study CLJs, Fig. 1, phase 3). For CLJs, participants indicated the likelihood of correctly classifying novel (unstudied) birds into each respective family (e.g., Sparrow, Thrasher). Following the CLJs, participants took tests in which they were shown images of birds along with the 12 family names. The participants' task was to correctly classify each exemplar (cf. classification tests, Fig. 1, phase 4). On one test, participants classified the birds they had studied (i.e., studied exemplar classification test), and on the other test participants classified novel (unstudied) birds. CLJ resolution was assessed by computing within-participant correlations between CLJs and the mean classification of novel birds for each family (see also, Rawson et al., 2014; Wahlheim et al., 2012). These correlations provide an index of the degree to which participants can discriminate the bird families for which classification will be high from bird families for which classification will be poor. Thus, resolution (also referred to as *relative accuracy*) will increase to the degree to which participants provide higher CLJs for families that subsequently have higher rates of correct classification and lower CLJs for families that have lower rates of correct classification.

The resolution of CLJs was significantly above zero (which indicates above chance accuracy), with correlations of .47 and .48 per condition (Wahlheim et al., 2011, Experiment 1). The magnitude of CLJ resolution was similar in a second experiment (overall correlation of .49), and was somewhat higher than reported in Jacoby et al. (2010; Experiment 2, $r = .28$; Experiment 3, $r = .29$). Given that this monitoring resolution is well below the maximum of 1.0, we decided to explore a technique that promises to improve it for learning concepts. One technique that can improve resolution in *episodic memory tasks* is to have people experience performing the to-be-judged task (e.g., Ariel & Dunlosky, 2011; Koriat, 1997; Koriat et al., 2002; Scheck & Nelson, 2005; Serra & Ariel, 2014; Serra & Dunlosky, 2005; Tauber & Rhodes, 2012). For example, Tauber and Rhodes (2012) had two groups of college students study pairs of words, make item-by-item judgments of learning (JOLs) predicting the likelihood of correctly recalling the target when given the cue, and complete a cued-recall test. They studied, judged, and were tested on the same word pairs across three study-test blocks. Of most interest, JOL resolution (i.e., within-participant correlations between JOLs and recall) increased with task experience. In particular, it increased from .23 and .55 on block 1 to .64 and .65 on block 2, and to .82 and .80 on block 3 (per respective group; Tauber & Rhodes, 2012). Such increases in resolution reflect improvements in monitoring at the item level that arise from excellent

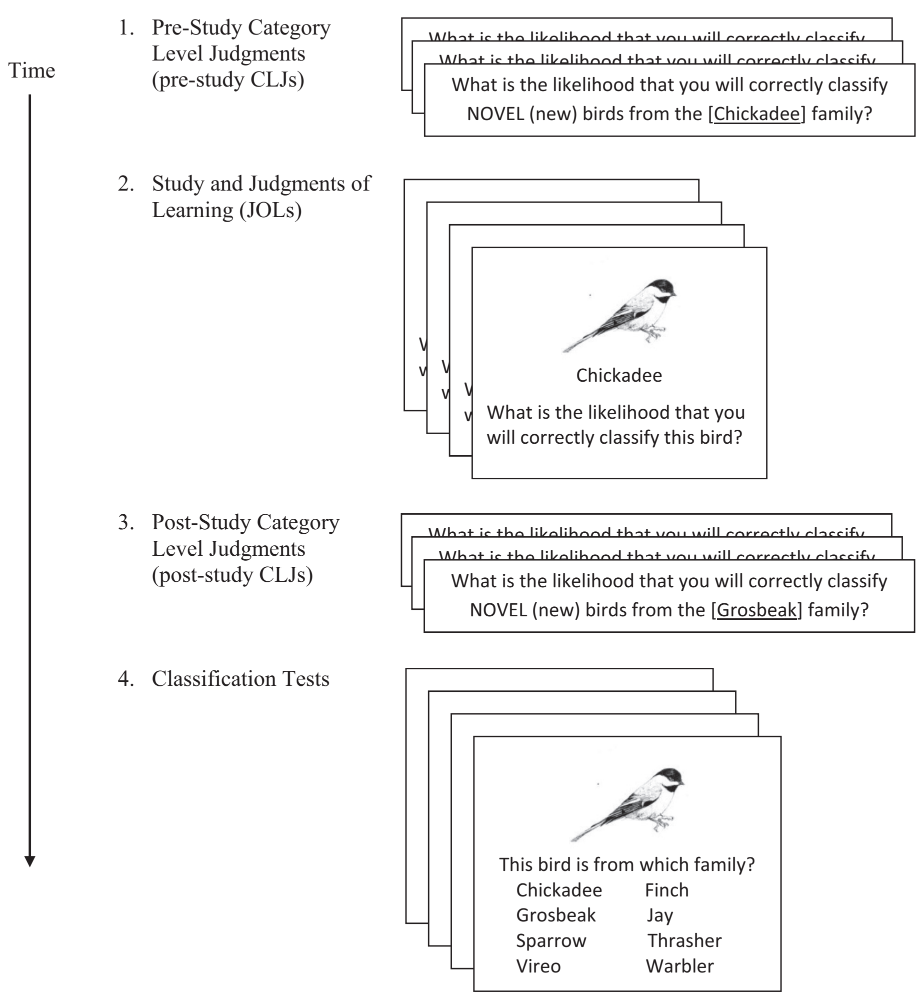

Fig. 1. Overview of the methodology employed in the current experiments. Phases 1–4 represent one complete study-test block. In Experiment 1 and Experiment 2, participants completed 2 blocks. That is, participants completed phases 1–4 twice. In Experiment 3 and Experiment 4, participants completed 3 blocks.