Cognitive Psychology 69 (2014) 1–24

Contents lists available at ScienceDirect

Cognitive Psychology

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

Costs and benefits of automatization in category learning of ill-defined rules

Maartje E.J. Raijmakers [a,b,*], Verena D. Schmittmann [c], Ingmar Visser [a,b]

[a] Department of Psychology, University of Amsterdam, The Netherlands
[b] Amsterdam Brain and Cognition (ABC), University of Amsterdam, The Netherlands
[c] Department of Methodology and Statistics, Tilburg School of Social and Behavioral Sciences, The Netherlands

A R T I C L E I N F O

Article history:
Accepted 11 December 2013

Keywords:
Category learning
Latent Markov analysis
Representational shifts
Strategies
Automaticity
Individual differences
Exemplar-based learning
Rule-based learning
Ill-defined categories

A B S T R A C T

Learning ill-defined categories (such as the structure of Medin & Schaffer, 1978) involves multiple learning systems and different corresponding category representations, which are difficult to detect. Application of latent Markov analysis allows detection and investigation of such multiple latent category representations in a statistically robust way, isolating low performers and quantifying shifts between latent strategies. We reanalyzed data from three experiments presented in Johansen and Palmeri (2002), which comprised prolonged training of ill-defined categories, with the aim of studying the changing interactions between underlying learning systems. Our results broadly confirm the original conclusion that, in most participants, learning involved a shift from a rule-based to an exemplar-based strategy. Separate analyses of latent strategies revealed that (a) shifts from a rule-based to an exemplar-based strategy resulted in an initial *decrease* of speed and an *increase* of accuracy; (b) exemplar-based strategies followed a power law of learning, indicating automatization once an exemplar-based strategy was used; (c) rule-based strategies changed from using pure rules to rules-plus-exceptions, which appeared as a dual processes as indicated by the accuracy and response-time profiles. Results suggest an additional pathway of learning ill-defined categories, namely involving a shift from a simple rule to a complex rule after which this complex rule is automatized as an exemplar-based strategy.

* Corresponding author. Address: Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, The Netherlands. Fax: +31 206390279.

E-mail address: m.e.j.raijmakers@uva.nl (M.E.J. Raijmakers).

0010-0285/$ - see front matter © 2014 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.cogpsych.2013.12.002

1. Introduction

Human category learning is a highly debated subject in cognitive science with increasingly complex conclusions about underlying learning mechanisms being rule based or similarity based (Ashby & Maddox, 2010; Goldstone & Kersten, 2003; Palmeri & Gauthier, 2004). Over many decades of category learning research, the focus has shifted from rule-based learning of well-defined categories (e.g., Bruner, Goodnow, & Austin, 1956) to learning ill-defined categories, i.e., category structures that are only partially describable by simple rules (e.g., Medin & Schaffer, 1978; Rosch & Mervis, 1975) and/or that are only well-defined by integrating information from multiple dimensions (Ashby & Ell, 2001). To understand the results of these and related studies, many single-component models of category learning have been proposed in the literature. These include prototype models (e.g., Posner & Keele, 1968; Rosch & Mervis, 1975), exemplar models (e.g., Kruschke, 1992), connectionist models (e.g., Gluck & Bower, 1988), Bayesian models (e.g., Anderson, 1991), and decision-boundary models (e.g., Ashby & Gott, 1988). Exemplar-based models in particular have been very successful in explaining many empirical results in category learning research (e.g., Kruschke, 1992; Nosofsky, 1986, 1988; Nosofsky & Palmeri, 1997).

Notwithstanding their success, there are important empirical results that the single-component models fail to account for. Those results reveal that different types of representations are formed within and between experiments (Ashby & Ell, 2001; Erickson & Kruschke, 1998). For example, analysis of the performance after learning an ill-defined category structure showed that individuals differed in the types of generalizations they made, suggesting the involvement of different learning systems (e.g., Erickson & Kruschke, 1998; Johansen & Palmeri, 2002; Nosofsky, Clark, & Shin, 1989). Dissociation studies and cognitive neuroscience studies revealed additional evidence for the existence of multiple systems of category learning (e.g., Davis, Love, & Preston, 2012; Maddox & Ing, 2005; Nomura et al., 2007), which is now a more commonly accepted hypothesis (Ashby & Maddox, 2010; Hélie, Waldschmidt, & Ashby, 2010). Hence, several hybrid models that combine multiple learning systems were introduced (Anderson & Betz, 2001; Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Nosofsky & Palmeri, 1998; Vandierendonck, 1995; for an overview see Palmeri, Wong, & Gauthier, 2004). Assuming the existence of multiple modes of categorization learning, an important question in current research is to identify the interaction between those learning systems (Ashby & Crossley, 2010; Ashby & Maddox, 2010).

In this article, we study the interaction between two category-learning systems by means of detailed analyses of participants' process of learning an ill-defined category structure using data partially presented in Johansen and Palmeri (2002). As they found representational shifts during learning of an ill-defined categorization structure, representational formats are latent, that is, not directly observable. In the current article, we use a statistical approach that allows for the identification of (subgroups of) participants that use different representations. This approach then extends the Johansen and Palmeri (2002) results by analyzing these subgroups of participants separately, allowing for a more detailed characterization of the process of changing representations. Following Rickard (2004, p. 65), to denote a type of categorization-learning process (either rule-based or exemplar-based) we will use henceforth the term categorization strategy, which is defined as "a unique series of mental steps toward a solution" and which "does not necessarily have direct implications regarding intention or awareness".

Before presenting our statistical approach we discuss the occurrence of representational shifts in relation to theories of automatization. This results in several hypotheses about the (latent) categorization strategies during learning.

1.1. Representational shifts and automatization

Johansen and Palmeri (2002) showed that there exist important inter-individual differences in learning ill-defined categories: during and at the end of their experiment, some participants had formed exemplar-based representations and others had formed rule-based representations. They also observed intra-individual differences: individual participants seemed to change their representations from rule-based to exemplar-based during the course of learning.