Cognition 131 (2014) 300–310

Contents lists available at ScienceDirect

Cognition

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

Building theory-based concepts: Four-year-olds preferentially seek explanations for features of kinds ☆

Andrei Cimpian *, Gina Petro

Department of Psychology, University of Illinois at Urbana–Champaign, 603 E. Daniel St., Champaign, IL 61820, United States

ARTICLE INFO

Article history:
Received 10 June 2013
Revised 4 December 2013
Accepted 28 January 2014

Keywords:
Concepts
Explanations
Development

ABSTRACT

Is the structure of human concepts continuous across development, or does it undergo qualitative transformations? Extensive evidence with adults has demonstrated that they are motivated to understand why categories have the features they do. To investigate whether young children display a similar motivation—an issue that bears on the question of continuity vs. transformation in conceptual structure—we conducted three studies involving 4-year-olds ($N = 90$) and adults ($N = 124$). Experiments 1 and 2 suggested that 4-year-olds indeed display a strong motivation to explain why categories have the features they do. Specifically, when provided with the option of asking "why?" about features of novel categories vs. features of individuals from other novel categories, children preferred to ask "why?" about the category features. Moreover, children's explanatory preference was specific to facts about categories per se and did not extend to facts that were merely presented in the context of multiple category instances. Experiment 3 also ruled out the possibility that the category facts were preferred because these facts were more surprising. In sum, these three studies reveal an early-emerging motivation to make sense of the categories encountered in the world and, more generally, speak to the richness of children's conceptual representations.

1. Introduction

A fundamental characteristic of human cognition is its ability to group distinct objects in the world into equivalence classes. Although no two apples are exactly the same, for example, on many occasions we think of them as equivalent tokens of the same *category* or *kind*. The ability to represent unique objects as interchangeable members of these broader classes is essential for much of human activity, from the mundane (e.g., referring to distinct objects with the same count noun) to the esoteric (e.g., diagnosing different patients with the same illness). Despite the centrality of categorization to our thinking and behavior, and despite many decades of research, key questions about its operation are still subjects of debate (for a review, see Murphy, 2004): What are the processes that underlie our ability to categorize, and how are the resulting categories represented? Many of the early answers to these questions emphasized the role of similarity, usually defined as feature overlap (e.g., Medin & Schaffer, 1978; Nosofsky, 1986; Rosch & Mervis, 1975). For example, if an object shares a sufficient number of features with remembered exemplars of the apple category (or, on other

☆ This research was supported by research funds from the University of Illinois to Cimpian. The Cognitive Development Lab team provided assistance in collecting and coding the data. We also thank Luke Butler, Vikram Jaswal, Joe Robinson, and the members of the Cognitive Development Lab for helpful discussion.

* Corresponding author. Address: Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820, United States. Tel.: +1 217 333 0852; fax: +1 217 244 5876.

E-mail address: acimpian@psych.illinois.edu (A. Cimpian).

http://dx.doi.org/10.1016/j.cognition.2014.01.008
0010-0277/

accounts, with this category's prototype), then it would be categorized as an apple. Although the extent of feature overlap is no doubt important, more recent arguments have increasingly recognized that categorization is also deeply *theoretical*, in the sense that it is bound up with people's causal beliefs and explanations rather than relying solely on statistical facts about feature frequencies (e.g., Ahn & Luhmann, 2004; Carey, 1985; Gelman, 2003; Heit, 1994; Keil, Smith, Simons, & Levin, 1998; Medin, 1989; Murphy & Medin, 1985; Prasada & Dillingham, 2006, 2009; Rehder, 2003). For example, even if an object shared very *few* features with previously encountered exemplars of the apple category, it may nevertheless be categorized as an apple if it was picked from an apple tree or if an apple sapling sprouted from its seeds—information that connects with our intuitive theories about biological reproduction.

To elaborate, the main thrust of the claim that categorization is theoretical is that human concepts comprise not just information about *what* features are characteristic of each category but also information about *why* each category has the features it does. For example, beyond simply learning that apples have seeds, stems, and other such features, people might also invoke—often implicitly—their intuitive biological knowledge to arrive at some understanding of why apples have these features (how they came about, what functions they may serve, etc.). For our purposes here, it is important to note that these two conceptual components (*what* and *why*) differ not just in their content but also in how they are typically acquired. Although the features associated with a category can be learned more or less passively by exposure to exemplars of the category, the *reasons* for these features are never on display in the world and must instead be generated by means of additional, often self-initiated, processing (e.g., retrieving relevant knowledge from memory, searching through the retrieved information for plausible reasons). In light of the extra steps required to infer why categories have the features they do, it is apparent that humans must at some level be *motivated* to understand the categories they identify in the world (Kaplan & Murphy, 2000; Murphy, 2000; Gopnik, 1998); if they were not, it is unclear why they would routinely, and without prompting, attempt to find the reasons behind the features of categories.

The present research explored the developmental origins of this motivation to understand why categories have the features they do. We pursued the question of origins for two reasons. First, from a descriptive viewpoint, it is important to know whether this motivated aspect of human concepts is a sophisticated late addition or a basic, early-developing component. Does the motivation to ask "why?" about features of categories arise only after the bulk of cognitive development has already occurred—perhaps once people have accumulated a certain amount of general world knowledge and the neurocognitive processes involved in generating explanations have matured sufficiently? Or is this motivation present even in young children, despite their sparse knowledge and limited cognitive resources? A second reason to pursue the origins question is that the timing of the emergence of this motivation has strong implications for theories of conceptual development. Specifically, its timing would bear on current debates about whether early concepts consist entirely of perceptual associations or incorporate more abstract components such as explanations. Before we spell out the predictions of these two theoretical perspectives with respect to whether the motivation to explain category features would be present in childhood, we review some of the evidence that this motivation is present in adults.

The claim that adults actively seek reasons for category features finds support in studies in which undergraduate participants were asked to learn novel categories. If adults are motivated to make sense of the categories they are learning, rather than learning them by rote, then category learning should be faster when the circumstances facilitate explanation (and thus sense-making). Consistent with this prediction, categories whose features fit together in explainable ways (e.g., novel vehicle categories with thematically-related features such as "made in Norway," "drives on glaciers," and "heavily insulated") were learned faster than control categories whose features were equally predictive of category membership but lacked clear explanatory connections (e.g., "white," "automatic," and "cloth seat covers"; Murphy & Allopenna, 1994;Pazzani, 1991; Wisniewski, 1995; Wisniewski & Medin, 1994). Not only did subjects learn faster when they could generate explanations, but their learning was also more detailed and precise. For example, subjects' estimates of feature prevalence among the members of these theme-based categories were often more accurate than their analogous estimates for control categories (Spalding & Murphy, 1999).

Subsequent research expanded on this initial work, providing additional support for the argument that people are motivated to make sense of the categories they identify in the world. For example, this motivation seems to operate even when the theme-related features (e.g., "drives on glaciers") are relatively infrequent, making up only about 15% of the novel categories' features (Kaplan & Murphy, 1999, 2000) rather than the majority of them, as in some of the original studies (e.g., Murphy & Allopenna, 1994; Spalding & Murphy, 1996). The fact that category learning benefits from the theme information even in a context in which extra cognitive effort is required in order to identify and use the themes for this purpose is consistent with the claim of a background motivation to understand why categories fit together as they do.

In fact, participants often seem motivated to integrate *all* of a category's features into a sensible whole—even ones that do not have a straightforward explanatory fit with the themes (Kaplan & Murphy, 2000). For example, it is not obvious whether Norwegian-made vehicles used for glacier driving should have cloth seats or vinyl seats. Nevertheless, when subjects were randomly assigned to learn that these vehicles had one type of seat or the other, they often generated explanations that allowed them to make sense of that feature and link it up to the broader theme (e.g., a subject who learned that these vehicles had cloth seats reasoned that "cloth seat covers would be better in the arctic because vinyl would get too cold"; p. 842). Here again we see evidence for a strong motivation to fit all aspects of a category into a sensible, explainable whole.