Cognition 129 (2013) 615–636

Contents lists available at ScienceDirect

Cognition

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

The difficulties of executing simple algorithms: Why brains make mistakes computers don't

Gary Lupyan *

University of Wisconsin, Madison, United States

A R T I C L E   I N F O

Article history:
Received 31 December 2012
Revised 9 August 2013
Accepted 12 August 2013

Keywords:
Concepts
Categorization
Human algorithms
Prototypes
Distributed representations
Inference

A B S T R A C T

It is shown that educated adults routinely make errors in placing stimuli into familiar, well-defined categories such as TRIANGLE and ODD NUMBER. Scalene triangles are often rejected as instances of triangles and 798 is categorized by some as an odd number. These patterns are observed both in timed and untimed tasks, hold for people who can fully express the necessary and sufficient conditions for category membership, and for individuals with varying levels of education. A sizeable minority of people believe that 400 is more even than 798 and that an equilateral triangle is the most "trianglest" of triangles. Such beliefs predict how people instantiate other categories with necessary and sufficient conditions, e.g., GRANDMOTHER. I argue that the distributed and graded nature of mental representations means that human algorithms, unlike conventional computer algorithms, only approximate rule-based classification and never fully abstract from the specifics of the input. This input-sensitivity is critical to obtaining the kind of cognitive flexibility at which humans excel, but comes at the cost of generally poor abilities to perform context-free computations. If human algorithms cannot be trusted to produce unfuzzy representations of odd numbers, triangles, and grandmothers, the idea that they can be trusted to do the heavy lifting of moment-to-moment cognition that is inherent in the metaphor of mind as digital computer still common in cognitive science, needs to be seriously reconsidered.

1. Introduction

In October, 2012 *Slate* magazine reported on a court case concerning a disputed election of a juvenile judge in Hamilton County, OH (Hasen, 2012). At issue were split-precinct polling places that required poll workers to hand out the appropriate ballots based on a rule such as whether the voter's address was even or odd. A poll worker testified to sending a voter with the address "798" to vote in the precinct for voters with odd-numbered addresses. Court testimony reveals that when asked whether the house number 798 was even or odd, the poll worker responded: "Odd." (*Tracie Hunter v. Hamilton County Board of Elections*, 2012). The remaining testimony follows:

Q. . .So on Election Day, if somebody came in with an address 798 and you had two ranges to choose from, you would choose the odd for them?
A. Yes.
Q. Okay. And is that how you did it for all the ballots that you looked up on Election Day?
A. To determine if they were even – yes.
Q. To determine if they were even or odd, you looked at the first digit of the address?
A. No. I looked at the whole address.
Q. And [if] there were more odds than even numbers, it would be an odd address?
A. Yes.

Although we can all agree with Hasen's conclusion that "no one should lose the right to vote because a poll worker can't tell an odd from an even number," it is worth considering whether such mistakes reveal something deeper

* Address: 1202W. Johnson St., University of Wisconsin, Madison, Madison, WI 53706, United States. Tel.: +1 917 843 4868.
E-mail address: lupyan@wisc.edu

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.08.015

about human cognition than an individual's confusion about the definition of numerical parity. In a series of experiments, I show that such classification errors are endemic, even when individuals' explicit definitions for determining category membership appear entirely correct. I argue that the reason people err in classifying items into categories with clear boundaries and known membership criteria is that human categorization algorithms are inherently sensitive to the particulars of the input. Thus, although the proposition *N* IS EVEN is *either* true or false, the mental representations—the *psychological* concept of parity[1]—may display the kind of graded, probabilistic structure that is characteristic of other concepts with fuzzier boundaries.

The question of how concepts are represented by the mind is at the very core of cognitive science (Fodor, 2001; Murphy, 2002; Prinz, 2004). The past 50 years has seen classical theories of concepts stressing necessary and sufficient conditions give way to theories stressing vagueness and context-dependence (Barsalou, 1987; Hampton, 2006; Lakoff, 1990; Medin & Smith, 1984; Prinz, 2004; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Rosch, 1973). In large part, these theories were created to account for the ease with which people adapt their knowledge to novel contexts (e.g., Clark, 1983; Fauconnier & Turner, 2003). Much of the evidence used to support these probabilistic and prototype-based theories of concepts came from studies in which one measures how people identify category members under various circumstances. In a now classic paper, Armstrong, Gleitman, and Gleitman (1983) challenged the idea that such tests can tell us much about the nature of conceptual representations by showing that both categories like FRUIT and ODD NUMBER showed graded structure as revealed by typicality ratings and longer classification times of "atypical" members (cf. Larochelle, Richard, & Soulières, 2000; cf. Sandberg, Sebastian, & Kiran, 2012). Armstrong et al. (1983) argued that because it is inconceivable that someone who knows the definition of numerical parity would truly believe that some numbers are odder than others, the finding that categorizing members from such categories as ODD NUMBERS—palpably different, according to the authors from fuzzier categories like PET and FRUIT—meant that the results from rapid classifications tasks must reflect functioning of peripheral identification procedures rather than tapping into "core" conceptual content (see Geeraerts, 1989 for discussion; and Gleitman, Armstrong, & Connolly, 2012 for a restatement of this position). Thus, although the difference in the time to classify an apple and a coconut as fruits may stem from a more "central" position of apple within the feature-space of FRUIT, the finding that it takes longer to identify 18 than 22 as EVEN cannot, according to the standard view, reflect such a difference. I present a series of studies showing that people, in fact, represent some numbers as odder than others, some triangles as more triangular, and argue that these effects stem from a failure to fully abstract from the details of the input making human algorithms qualitatively different from context-free computer algorithms that have inspired classical cognitive science.

Understanding the computations that underlie classification is relevant not only for understanding explicit categorization, but also informs theories of cognition more broadly. For example, many language parsers require words to be classified into abstract categories on which further computations are performed (Chomsky, 1995; cf. Anderson, 2006; Sleator & Temperley, 1995). Such assumption have led some to argue that e.g., infants' sensitivity to the similarity structure of the syllable sequences ABA and CDC arises from algebraic computations that treat syllables as context-independent variables (Marcus, 1999; cf. Seidenberg, 1999). On some theories, such symbolic manipulation is not limited to any special domain, but characterizes the entirety of mental processes (e.g., Gallistel & King, 2009). Given the relative simplicity of e.g., the algorithm for computing numerical parity, any symbolic device worth its salt should be able to abstract from the "surface" properties of the input in computing parity. The 13 experiments below test this basic hypothesis. For convenience, a summary of the basic manipulations and results is listed in Table 2.

2. Experiment 1. Speeded parity judgments

In the first experiment participants completed a standard classification task requiring judgments of numerical parity. Of interest was whether people who could all articulate the correct definition of parity would nevertheless make errors in classifying numbers having opposite-parity digits, such as 798.

2.1. Participants and procedure

Ten undergraduate students participated for credit. Each trial began with a fixation cross (0.9–1.1 s) followed by a 1–4 digit numeral displayed for 1.0 s or until response. The numerals appeared in a random position within an invisible horizontally-oriented rectangle ($\sim 15° \times 5°$). Each digit subtended $\sim 0.6° \times 1°$ of visual angle. On half of the trials, the numerals were shown obliquely (±45° or ±60°). This oblique presentation helped to measure the contribution of perceptual-selection errors, as described below. Each participant completed 16 practice trials during which incorrect answers or timeouts were indicated by buzzes, followed by 243 experimental trials (Table 1) with timeout

Table 1
Distribution of trials in Exp. 1.

Number of digits	Number of opposite-parity digits			
	Zero	One	Two	Three
One	27[a]	–	–	–
Two	24	24	–	
Three	24	24	24	–
Four	24	24	24	24

[a] Zero was omitted; (1–9) × 3 repetitions.

[1] It is necessary to distinguish between concepts in the philosophical sense, concerned with the actual state of the world, and concepts in the psychological sense, concerned with mental content—how people actually represent the world. It is this psychological definition that is used here (see also Hampton, 2012).