Cognition 168 (2017) 46–64

Contents lists available at ScienceDirect

Cognition

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

Original articles

People learn other people's preferences through inverse decision-making

Alan Jern [a,*], Christopher G. Lucas [b], Charles Kemp [c]

[a] Department of Humanities and Social Sciences, Rose-Hulman Institute of Technology, United States
[b] School of Informatics, The University of Edinburgh, United Kingdom
[c] Department of Psychology, Carnegie Mellon University, United States

ARTICLE INFO

Article history:
Received 15 June 2016
Revised 14 June 2017
Accepted 15 June 2017

Keywords:
Preference learning
Trait inference
Social cognition
Inverse decision-making

ABSTRACT

People are capable of learning other people's preferences by observing the choices they make. We propose that this learning relies on inverse decision-making—inverting a decision-making model to infer the preferences that led to an observed choice. In Experiment 1, participants observed 47 choices made by others and ranked them by how strongly each choice suggested that the decision maker had a preference for a specific item. An inverse decision-making model generated predictions that were in accordance with participants' inferences. Experiment 2 replicated and extended a previous study by Newtson (1974) in which participants observed pairs of choices and made judgments about which choice provided stronger evidence for a preference. Inverse decision-making again predicted the results, including a result that previous accounts could not explain. Experiment 3 used the same method as Experiment 2 and found that participants did not expect decision makers to be perfect utility-maximizers.

1. Introduction

One way to learn what other people like is by observing the choices they make. For example, suppose that Alice orders a boxed lunch that includes an eggplant sandwich and you want to know how much Alice likes eggplant sandwiches. If Alice ordered the only box with an eggplant sandwich, you might infer that Alice has a strong preference for eggplant sandwiches. If the eggplant sandwich is part of the only box that contains a cookie, you might instead infer that Alice has no particular preference for eggplant sandwiches and she really wanted the cookie. Although people commonly make these sorts of inferences, this example illustrates that someone's choice could have many different explanations, and deciding which of these explanations is best can be a challenging inductive problem.

Inferences like these have been studied in the literature on interpersonal attribution (Gilbert, 1998; Hamilton, 1998), and have been the target of developmental work with children (Diesendruck, Salzer, Kushnir, & Xu, 2015; Hu, Lucas, Griffiths, & Xu, 2015; Kushnir, Xu, & Wellman, 2010; Lucas et al., 2014; Luo, Hennefield, Mou, vanMarle, & Markson, in press; Ma & Xu, 2011; Repacholi & Gopnik, 1997). Most of this literature, however, does not emphasize computational approaches (for some exceptions, see Kunda, 1998; Lucas et al., 2014; Medcof, 1990). Research in economics and marketing has produced multiple computational methods for inferring consumers' preferences from their choices (Green & Srinivasan, 1990; Varian, 2006), but these methods have not been explored as psychological models. By contrast, there are multiple psychological models of how people make choices (Busemeyer & Johnson, 2008; Schneider, Oppenheimer, & Detre, 2007; Shenoy & Yu, 2013; Train, 2009), but few attempts to apply models like these to the problem of inferring people's preferences from observations of their choices. In this paper, we explore a computational approach to preference learning based on inverting a decision-making model and test it as a psychological account. We call this approach *inverse decision-making*.

The inverse decision-making approach is illustrated in Fig. 1a. The figure shows an example in which Alice chooses between three boxed lunch options: (1) eggplant sandwich and a cookie, (2) turkey sandwich and a slice of cake, and (3) tuna sandwich and an apple. The utility function in Fig. 1a (depicted by a bar chart) shows that Alice prefers the eggplant sandwich over the other sandwiches and the cookie over the other desserts. A decision-making model specifies a *decision function* that maps preferences to choices. Given Alice's preferences, any standard model of decision-making will predict that Alice will choose the option with an eggplant sandwich and a cookie. The shading on the nodes of the graph in Fig. 1a indicates what information about Alice's choice is visible to an observer. In this case, that includes the three boxed lunch

* Corresponding author at: Department of Humanities and Social Sciences, Rose-Hulman Institute of Technology, 5500 Wabash Ave, Terre Haute, IN 47803, United States.
E-mail address: jern@rose-hulman.edu (A. Jern).

http://dx.doi.org/10.1016/j.cognition.2017.06.017
0010-0277/

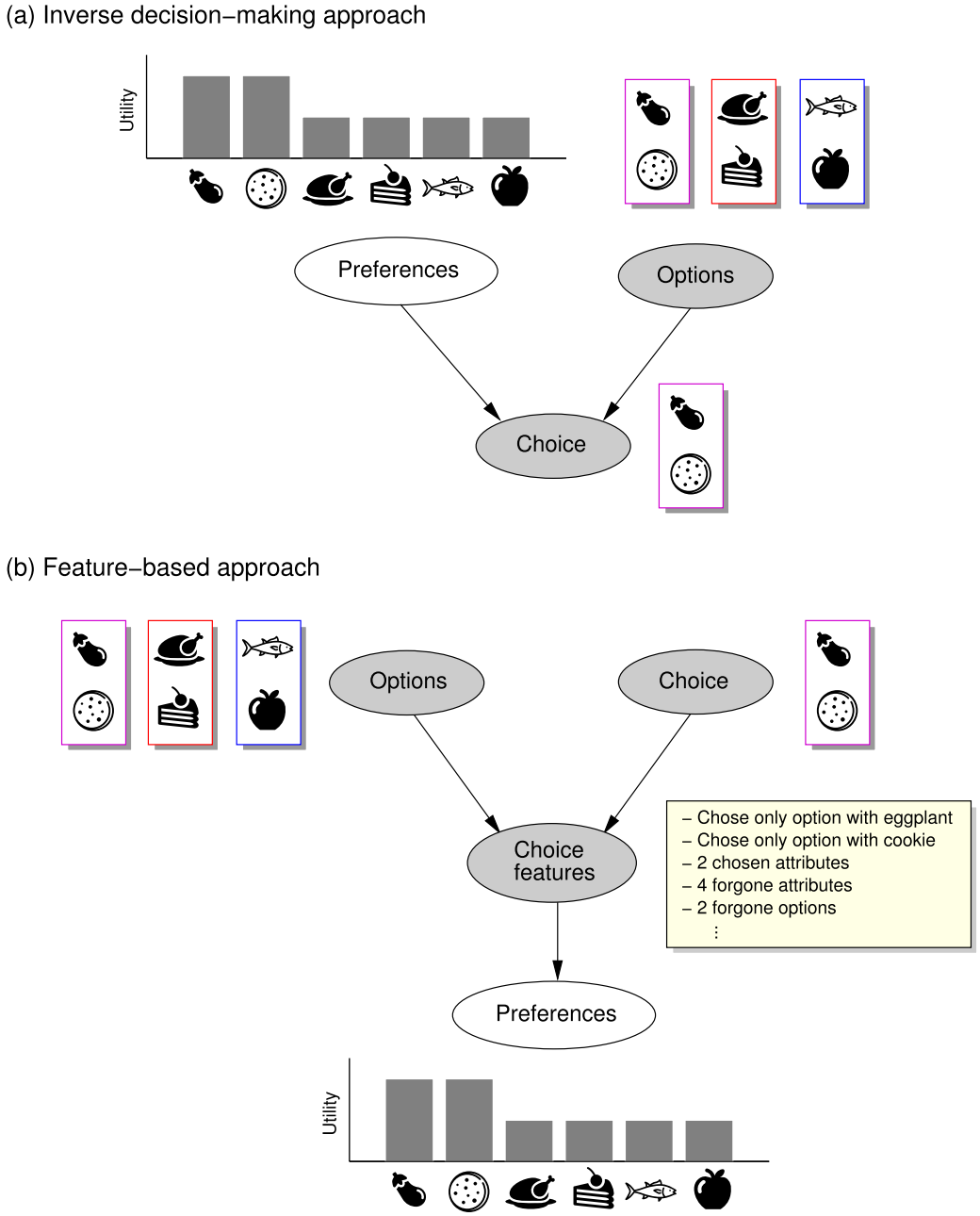

Fig. 1. Two approaches to preference learning applied to Alice's choice of boxed lunch. In both panels, the shaded nodes represent observed information and the unshaded nodes represent inferred information. (a) The inverse decision-making approach specifies a decision function that maps Alice's preferences and choice options to her choice and then inverts this function to infer the preferences that led to her choice. (b) The feature-based approach maps a set of features directly to the preferences that led to Alice's choice.

options and Alice's choice. The unshaded node indicates that Alice's preferences are not visible to an observer. Even so, the observer can invert a decision-making model to make inferences about the unobserved preferences that led to the observed choice.

Fig. 1b shows an alternative *feature-based* approach that does not rely on a decision-making model. Instead, this approach characterizes Alice's choice using a set of features. For example, the features in Fig. 1b indicate that Alice chose the only option with an eggplant sandwich and the only option with a cookie, that her choice had two attributes (eggplant sandwich and cookie), that she passed up four attributes (turkey, tuna, cake, apple), and that she passed up two options (the two boxes that she did not choose). These features carry information about Alice's preferences, and the feature-based approach relies on an *inference function* that maps choice features to preferences. For example, the larger the number of chosen attributes, the less likely it is that she was specifically interested in the eggplant sandwich, and the larger the number of forgone options, the more likely it is that Alice has a strong preference for eggplant.

The inverse decision-making approach has received little attention in the social psychology literature, but the feature-based approach has served as the basis for several influential accounts of interpersonal attribution (e.g., Jones & Davis, 1965; Kelley, 1973; Newtson, 1974). One example of the feature-based approach is Jones's and Davis's (1965) correspondent inference theory (CIT). One choice feature identified by CIT is whether a chosen attribute