Cognition 150 (2016) 133–149

Contents lists available at ScienceDirect

Cognition

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

Interference effects of categorization on decision making ☆

Zheng Wang [a,*], Jerome R. Busemeyer [b]

[a] School of Communication, The Ohio State University, 3145 Derby Hall, 154 N. Oval, Columbus, OH 43210, United States
[b] Indiana University, United States

ARTICLE INFO

Article history:
Received 20 October 2014
Revised 26 January 2016
Accepted 28 January 2016

Keywords:
Categorization
Decision
Interference effects
Law of total probability
Entanglement
Superposition
Quantum probability
Markov model
Signal detection model

ABSTRACT

Many decision making tasks in life involve a categorization process, but the effects of categorization on subsequent decision making has rarely been studied. This issue was explored in three experiments ($N = 721$), in which participants were shown a face stimulus on each trial and performed variations of categorization-decision tasks. On C-D trials, they categorized the stimulus and then made an action decision; on X-D trials, they were told the category and then made an action decision; on D-alone trials, they only made an action decision. An interference effect emerged in some of the conditions, such that the probability of an action on the D-alone trials (i.e., when there was no explicit categorization before the decision) differed from the total probability of the same action on the C-D or X-D trials (i.e., when there was explicit categorization before the decision). Interference effects are important because they indicate a violation of the classical law of total probability, which is assumed by many cognitive models. Across all three experiments, a complex pattern of interference effects systematically occurred for different types of stimuli and for different types of categorization-decision tasks. These interference effects present a challenge for traditional cognitive models, such as Markov and signal detection models, but a quantum cognition model, called the belief-action entanglement (BAE) model, predicted that these results could occur. The BAE model employs the quantum principles of superposition and entanglement to explain the psychological mechanisms underlying the puzzling interference effects. The model can be applied to many important and practical categorization-decision situations in life.

1. Introduction

The fields of categorization and decision making are empirically mature and theoretically well developed, but to a large degree, they have evolved in a parallel and independent manner. Little is known about the interactions between these two basic cognitive tasks – that is, how a categorization task changes performance on a subsequent decision task.[1] In many situations in life, decision makers need to make categorizations before deciding on an action. For example, a doctor needs to categorize a biopsy as cancerous or not before making treatment decisions; a judge needs to categorize a defendant as guilty or not before assigning a punishment; a police officer needs to categorize a driver as intoxicated or not before making an arrest; a military operator needs to categorize an agent as an enemy or not before making an attacking decision. In all these examples, it seems necessary to infer a category before choosing an action. Suppose the decision maker has to *report* this category inference before making the decision. How does this overt report of the category affect the later decision? For example, would the probability that a police officer shoots a suspect be changed if she or he had to report seeing a weapon possessed by the suspect first?

In the work described below, participants were presented with a face and were asked to categorize it first and then decide on an action. However, the general categorization-decision paradigm is not limited to these particular details, and as mentioned above, there are many important and practical examples of categorization-decision situations in real life. In general, any task that has the following four characteristics falls into this paradigm: (1) a stimulus providing information is presented, after which (2) a categorical inference is made based on the stimulus, followed by (3) a decision about an action, and (4) the action has consequences that depend on both the action and the true state of the category.

To explore the relation among these tasks, three experiments were conducted, and three theoretical explanations – a Markov

☆ This research was supported by the US NSF SES-1153726, SES-1153846, the US AFOSR FA 9550-12-1-0397, and FA 9550-15-1-0343.

* Corresponding author.
E-mail addresses: wang.1243@osu.edu, zhengjoycewang@gmail.com (Z. Wang).

[1] Maddox and Bohil (1998) examined the effects of decision making variables such as prior probabilities and payoffs on a categorization task, but here we examine how a categorization task affects a subsequent decision task. More closely related is the effect of categorization on subsequent feature inferences, such as has been discussed by Murphy and Ross (1994), Griffiths, Hayes, and Newell (2012), and Chaigneau, Barsalou, and Sloman (2004). In the General Discussion section, we relate our research to these other lines of work.

http://dx.doi.org/10.1016/j.cognition.2016.01.019
0010-0277/

model, a signal detection model, and a quantum cognition model based on quantum probability rules – are discussed and compared. Only the quantum cognition model *a priori* predicted an interference effect of categorization on subsequent decision making that systematically occurred in the experiments.

2. The categorization-decision paradigm

2.1. The categorization-decision interference

Townsend, Silva, Spencer-Smith, and Wenger (2000) initiated an investigation of the category-decision paradigm. On each trial, participants were shown one of 34 faces that were assigned to a "good guy" or "bad guy" category based on some facial features (e.g., width of faces), and then asked to decide whether to "attack" the face or "withdraw" from it. Fig. 1 illustrates some examples of the faces used in our new experiments, which were similar to those employed by Townsend et al. As shown, it was fairly easy to discriminate the two types of faces, but the task was made difficult because the assignment of faces to a category was probabilistic: The narrow faces were assigned to the "bad guy" category on 60% of the trials and to the "good guy" category on the remaining trials; likewise, the wide faces were assigned to the "good guy" category on 60% of the trials and to the "bad guy" category on the remaining trials.

The category was important because participants were rewarded on 70% of the trials for attacking faces that were assigned to the bad guy category and punished on 70% of the trials for attacking faces that were assigned to the good guy category. Likewise, they were rewarded on 70% of the trials for withdrawing from faces assigned to the good guy category and punished on 70% of the trials for withdrawing from faces assigned to the bad guy category. Participants were given six blocks of training, during which they first categorized a face and then decided on an action, and afterwards feedback was provided on both the category and the decision. The key manipulation occurred during a transfer test phase, during which each person received two additional blocks with three types of trials: (1) categorization and then decision (C-D) trials exactly like the original training, (2) categorization (C-alone) trials in which only a categorization was made with feedback, and (3) decision (D-alone) trials in which only a decision was made with feedback. For example, on a D-alone trial, the person was shown a face, simply decided to attack or withdraw, and received feedback on the decision. Of course, the categorization of the face on the D-alone trial remained highly relevant to the action decision, and it seems some implicit inference about the category was necessary before participants made the decision even though they did not have to explicitly report this inference.

Using this paradigm, one can examine within each participant how the overt report of the category interferes with the subsequent decision by comparing the probability of attacking on the D-alone trials (denoted as $p(A)$ for a face type) with the total probability of attacking on the C-D trials (denoted as $p_T(A)$ for the same face type). The latter is simply the probability of attacking on C-D trials pooled across trials when the categorization response is ignored. It can also be expressed using the classical law of total probability, which states that the probability to attack (A) equals the probability that the person categorizes a face as a good guy (G) and then attacks plus the probability that the person categorizes the face as a bad guy (B) and then attacks: $p_T(A) = p(G \cap A) + p(B \cap A)$. If these two ways of determining the probability of attacking on D-alone and C-D trials agree for a participant, $p(A) = p_T(A)$, then we say that the law of total probability is empirically satisfied. Based on a chi-square test, Townsend et al. (2000) found that 25% of 138 participants produced statistically significant violations of this law. Apparently, the seemingly innocuous overt report of a category changed how a subsequent decision was made. Specifically, we define an *interference effect* of categorization on decision making as the difference between the probabilities of an event when it is measured alone versus when it is measured after another event, such as, in our context, the probability of attacking on the D-alone trials and the total probability of attacking pooled across the C-D trials.

Busemeyer, Wang, and Lambert-Mogiliansky (2009) further investigated this paradigm and discovered a more surprising result. Their study involved 26 participants, and each participant received both C-D trials and D-alone trials. As shown in the first two rows of Table 1, when a face was most frequently assigned to the good guy category (we denote this type of face as type *g* faces), there was almost no interference effect. However, when a face was most frequently assigned to the bad guy category (we denote this type of face as type *b* faces), the probability of attacking was significantly greater for the D-alone condition as compared to the C-D condition, violating the law of total probability ($p(A) > p_T(A)$ for type *b* faces). More surprisingly, the probability of attacking in the D-alone condition, which left the good or bad guy categorization unresolved, was even greater than the probability of attacking given that the person had already categorized the face as a bad guy in the C-D condition ($p(A) > p(A|B)$) for type *b* faces! It is surprising that for some reason, the overt categorization response interfered with the action decision by reducing the tendency to attack faces that most likely belonged to the bad guy category.

2.2. Candidate models for the categorization-decision paradigm

There are several models that can be considered for the application to the *general* categorization-decision paradigm (not just the particular example used in the current study). Below we briefly summarize five candidates. The first two, the optimal and probability matching models, are oversimplified but provide useful baselines for considering competing models for the paradigm. They predict no interference effects. The next two, Markov and signal detection models, are more general cognitive models, but they fail to predict any interference effects either in an *a priori* manner. The last is a quantum cognition model, which *a priori* predicts that an interference effect could occur.

2.2.1. Optimal model

The optimal model describes the optimal behaviors. According to the optimal model, the decision to attack should depend only on the face. If a type *b* face is presented, then it is always optimal to attack, and if a type *g* face is presented, then it is always optimal to withdraw. This follows from the fact that the probability of reward for attacking equals the probability that the type of face is assigned to the bad guy category (.60) times the probability that a reward is given for attacking a bad guy (.70), plus the probability that the same type of face is assigned to the good guy category (.40) times the probability that a reward is given for attacking a good guy (.30). That is, for a type *b* face, the total probability of being awarded for attacking equals $.60 \cdot .70 + .40 \cdot .30 = .54$, and the probability of reward for withdrawing is $1 - .54 = .46$, so the optimal model predicts that participants should always decide to attack when a type *b* face is presented. Likewise, the optimal model predicts that the participant should always decide to withdraw when a type *g* face is presented. These predictions hold regardless of whether the trial is a C-D trial or a D-alone trial, because the categorization response provides no new information for making the action decision. Therefore, the optimal model predicts no interference effect for the categorization-decision paradigm.