TPM

A PROBABILISTIC MODEL FOR INTEGRATION OF STRONG DEPENDENT CUES IN CATEGORY IDENTIFICATION

LUIGI LOMBARDI
BARBARA BAZZANELLA
ANTONIO CALCAGNÌ
UNIVERSITY OF TRENTO

Semantic features are relationally connected to one another and they may show different levels of importance in indexing a familiar basic-level category. The article proposes a new Bayesian model to efficiently capture the relative importance of possibly dependent semantic features in the process of category identification. Unlike the Bayesian model of category identification (naive Bayes identifier), the new model does not require independence between cues and uses a simple property, called strong stochastic dependence (SSD), to efficiently represent dependence-type patterns in semantic features. We applied the model to empirical data from a category identification task administered to a group of 30 participants. Results show that the new Bayesian model is consistently superior to the naive Bayes identifier in predicting the distribution of the participants' responses in the empirical semantic task.

Key words: Integration of multiple cues; Category identification; Bayesian models; Naming to description.

Correspondence concerning this article should be addressed to Luigi Lombardi, Department of Psychology and Cognitive Science, University of Trento, Corso Bettini 31, 38068 Rovereto (TN), Italy. Email: luigi.lombardi@unitn.it

The problem of how to quantify cognitive processes concerning the integration of informative cues in real life contexts has been considered crucial in many human information processing theories of perception (e.g., N. H. Anderson, 1981; Ashby & Townsend, 1986), categorization (e.g., J. R. Anderson, 1991; Nosofsky, 1986), and generalization (e.g., Shepard, 1957; Tenenbaum & Griffiths, 2001). Typically, models of cue integration are tested through behavioral experiments with cues consisting of putatively orthogonal dimensions or independent features such as shape, size, and color (e.g., Ashby & Townsend, 1986). The assumption of independence has at least two important advantages. First, it makes the formal development and expression of a model easier; second, it makes model instantiation and formal operations tractable.

However, the assumption of independence can be unwarranted and matters can be more complicated when cues are relationally connected to one another (e.g., cause/effect) (Love & Markman, 2003). The problem of dependence between cues seems particularly relevant in the context of knowledge representation. Natural concepts can be extremely rich and groups of correlated features can serve as the foundation upon which conceptual representations are formed (e.g., Smith & Medin, 1981). Moreover, theoretical or causal knowledge can also influence the way people relate the features of a category to one another (such as birds flying because they

have wings) and how that knowledge influences the classification of objects (Rehder, 2003; Rehder & Kim, 2006; Sloman, Love, & Ahn, 1998).

The question we address in this paper is how individuals integrate a small number of possibly dependent semantic cues to identify a familiar basic-level category. Suppose that **f** specifies one or more cues (or features) of a category, but the category name (or label), say *c*, is not provided. More precisely, in case of multiple cues we say that **f** represents a *conjunctive combination* of semantic cues for the unlabelled category. *Category identification* (Kemp & Jern, 2014) is the problem of inferring the hidden category label *c* on the basis of semantic cues **f**. For example, if someone mentions that h(er/is) favourite kind of pet is a *faithful friend* and *barks*, I may infer that s(he) probably likes dogs. In the above example, the features faithful friend and barks seem to be highly correlated (concept-dependent correlation). Unfortunately, the majority of the available models for category identification (e.g., Kemp, Chang, & Lombardi, 2010; Lombardi & Sartori, 2007; Medin & Schaffer, 1978; Shepard, 1957) require that the semantic cues are treated as if they make separable contributions to the model's output, that is, they assume independence between features. Here we take a different approach that, instead, triggers category identification on the basis of possible pattern of *dependence* between features. In particular, in this paper we present a novel Bayesian model for category identification that integrates semantic cues on the basis of a simple property called strong stochastic dependence (SSD). The remainder of the paper is organized as follows. First, we briefly review the Bayesian model of category identification based on the assumption of independence between cues. Next, we define the notion of strong stochastic dependence between semantic cues and apply it to reformulate the Bayesian model. We then compare the predictions of the Bayesian model for independent cues and the new model on data collected using a category identification task paradigm. Finally, we close by discussing the results of the novel approach and compare it with a previous well-known model of category classification (Rehder, 2003) and emphasize the commonalities and differences with respect to our new contribution.

Bayesian Models of Category Identification

Basic Tokens

The Bayesian framework allows for a normative optimal solution of the identification problem as sketched above. In order to clarify the terms of our approach, we first introduce the basic tokens of the models. In particular, we (re)define the notions of category prior probability, cue dominance, and cue validity in the category identification problem.

The prior probability of a basic-level category *c* (category prior probability) is denoted by $p(c)$ and captures a factor like *familiarity* (Mandler, 1980). The critical assumption is that a higher level of familiarity for the category corresponds to a higher prior probability for that category. For example, for an individual the concept *Cat* is usually more familiar than *Okapi*, therefore the prior probability of the former should be higher than that of the latter.

The *cue dominance* of a feature *f* for one category *c* is defined as the conditional probability $p(f|c)$ that an individual generates or lists *f* given *c*. Cue dominance can be understood as a local measure which refers to the strength (or likelihood) of a feature-category association. For

example, in describing the category *Dogs*, the feature *used for hunting* has medium cue dominance as it is used to describe the concept by some, but not all people. By contrast, *barks* is a feature with high cue dominance because it is usually reported by the majority of people.

The *cue validity* of a feature f for one category c is defined as the posterior probability $p(c|f)$ of recovering the category c given that it has feature f. Bayes' formula can be used to express cue validity in terms of category prior probability and cue dominance: $p(c|f) \propto p(c)p(f|c)$. Cue validity is, therefore, used to predict the category with the highest posterior probability. In the categorization literature, cue validity has also been called *diagnosticity* (Rosch, 1978) and refers to the informational value of a feature for one category relative to a family of categories in a semantic domain. For example, if one's task is to retrieve a category name, say *Camel*, the feature *two humps* may prove highly diagnostic as it excels at distinguishing Camel from other concepts of animals (i.e., horse, dromedary, llama).

Multiple Cues

Cue validity is the kernel construct to model category identification. Of course, cue validity can be straightforwardly extended to the case of multiple cues $\mathbf{f} = (f_1, f_2, \ldots, f_m)$. A model for multiple cue integration in category identification has been proposed by several authors (e.g., J. R. Anderson, 1991; Ashby & Alfonso-Reese, 1995; Kemp et al., 2010):

$$\tau_c \equiv p(c|\mathbf{f}) \propto p(c)p(\mathbf{f}|c)$$
$$= p(c)\prod_{j=1}^{m} p(f_j|c). \quad (1)$$

The model described in Equation 1 makes the strong assumption that all the cues $f \in \mathbf{f}$ are conditionally (stochastic) independent given the value (label name) of the category c. Since the structure of this model closely resembles that of a popular model in machine learning called the naive Bayes classifier (Duda & Hart, 1973), we renamed it the *naive Bayes identifier*. When represented as a Bayesian network, this model has the simple structure depicted in Figure 1A. In what follows we propose an alternative representation based on the notion of strong stochastic dependence to model cue integration in category identification.

Strong Stochastic Dependence (SSD)

The core intuition behind SSD is simple. In the knowledge representation of a natural category the highly interconnected network of its semantic features usually entails a relevant level of informational redundancy. In some circumstances this informational redundancy can be reduced to *strong dependence*. In particular, we say that the semantic cues f in $\mathbf{f}$ are strong dependent cues given a target category c if, and only if, exists at least one cue f_r in $\mathbf{f}$, called the *reference cue*, such that

$$p(\mathbf{f}^d|f_r, c) = 1 \quad (2)$$

where $\mathbf{f}^d$ is the set of *derived cues*, that is, the (m-1) array of cues obtained after removing f_r from **f**. When represented as a Bayesian network, strong stochastic dependence has the structure depicted in Figure 1B.

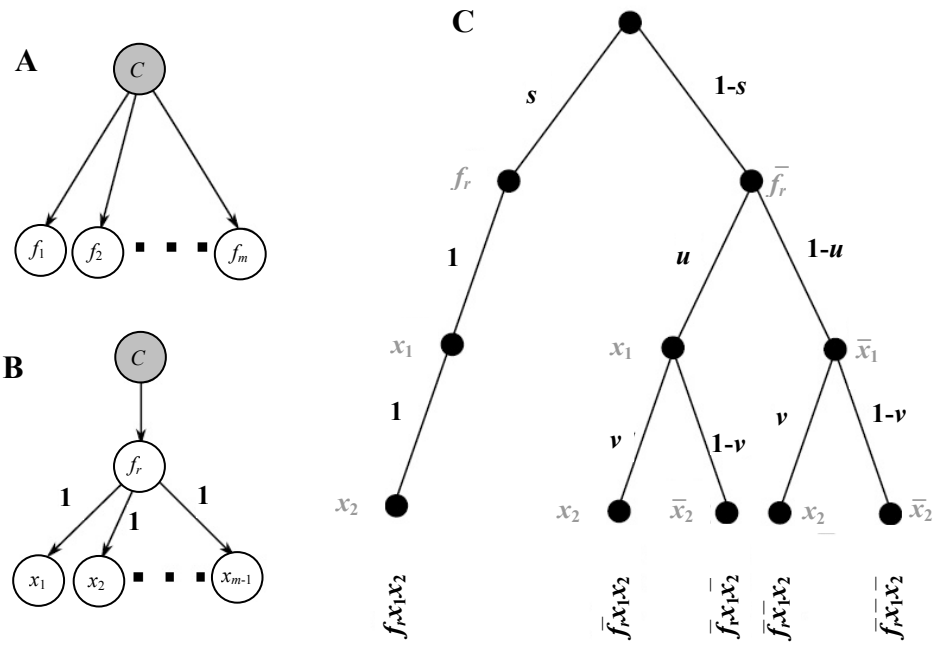

FIGURE 1

A. The structures of the naive Bayes identifier when cues $f_1,\ldots,f_m$ and one category c are considered.
B. Bayesian model for strong dependent cues (SSD) with reference cue f_r and derived cues $x_1, x_2,\ldots, x_{m\text{-}1}$.
C. A SSD branching process with three cues f_r, x_1, x_2
and with branching probability values s, u, and v, respectively.
Note that the SSD branching process satisfies Equation 2 in the text: $p(f_r, x_1, x_2|c) = p(f_r|c) = s$.
For sake of conciseness, the branching process does not represent the zero probability patterns.
The overline symbol denotes the negation of a generic cue (e.g., not x_1).

This network structure captures the main assumption of SSD, namely, that every attribute f in $\mathbf{f}^d$ is dependent (with probability 1) from the reference cue f_r, given the value (label) c of the category of interest. In other words, the derived cues are fully determined given the reference cue of the target category. From the latter it follows that

$$\begin{aligned} p(\mathbf{f}|c) &= p(f_r, \mathbf{f}^d|c) \\ &= p(\mathbf{f}^d|f_r, c)p(f_r|c) \\ &= p(f_r|c) \end{aligned} \tag{3}$$

which leads to the definition of the new Bayesian model for strong dependent cues:

$$\begin{aligned} \pi_c \equiv p(c|\mathbf{f}) &\propto p(c)p(\mathbf{f}|c) \\ &= p(c)p(f_r|c). \end{aligned} \tag{4}$$

In sum, SSD reduces the integration process of multiple cues to one single cue (the reference cue) the value of which is a necessary and sufficient condition to represent the informational content of the description **f**. Figure 1C illustrates an example of stochastic branching process that generates samples of features according to the SSD condition.

The experiment presented in the next section will allow the evaluation of the integration mechanisms by means of which semantic cues activate categories in the mental lexicon. Here we contrast two opposite assumptions about semantic cue integration: stochastic independence (SI) and strong stochastic dependence (SSD). Although both assumptions represent only simplified approximations of the true, but unknown, integration process in category identification, we speculate that in many circumstances semantic features will be still integrated in a way that is less independent than a naive Bayes identifier would predict. In particular, if our main hypothesis is correct, participants' responses should display patterns closer to the interpretation of high informational redundancy (SSD) in the semantic description of a category. The counterhypothesis is that the integration process contains no mechanism to differentially integrate semantic features and that participants' responses display patterns in line with the interpretation of stochastic independence of features. The experiment that follows simultaneously tested the naive Bayes identifier and the Bayesian model for strong dependent cues by comparing empirical data collected from a category identification task (also known as naming to description) to the values predicted via the two models defined in Equation 1 and Equation 4.

METHOD

Participants

Thirty undergraduate and postgraduate students (19 women, 11 men; mean age = 26.13) from the Department of Psychology and Cognitive Science of the University of Trento (Italy) participated voluntarily in the experiment.

Stimuli and Procedure

Twenty-four category names were used, randomly selected from a larger pool of 541 categories belonging to the repository of McRae, Cree, Seidenberg, and McNorgan (2005), which furnished the semantic domain under study. The repository contained concepts included in 13 different superordinate categories (i.e., birds, clothes, furniture, fruits, mammals, vehicles) and 2526 distinct features. Each of the 24 categories was described by a different sentence **f** containing a set of three semantic features randomly selected from the set of all features that were associated to that category in the repository. This random sampling procedure guaranteed that the construction of the descriptions was a priori biased for neither independent nor dependent cues. The sampling did not exclude, however, that in the repository the features were naturally associated according to a category-dependent representation, which would reflect the natural correlations among the features of the category in the semantic domain. The 24 descriptions constituted the items of the category identification task (see Appendix A for the list of the category task items).

The task was preceded by a rule-learning phase in which all participants were instructed to correctly elaborate on conjunctive combinations of semantic cues. In this pretest phase, participants had to correctly choose from a set of pictures representing animate or inanimate objects only those that conjunctively satisfied a list of target cues. For example, given the cues *mane* and *sharp teeth*, the subject had to select the picture representing a lion, but not a lioness or puma. For each participant the learning phase ended when s(he) performed 100% correct in a sequence of 10 consecutive trials. The items of the learning phase were all different from those used in the category identification task and selected in order to minimize any interference with the items of the subsequent task.

After the learning phase the category identification task was administered to the participants. The three semantic features were presented orally and in random order to the participants, who were required to retrieve the corresponding category without any emphasis on speed (no time pressure). The required responses were oral. Participants completed the task individually in a quiet room and the experimenter was present during the experimental session. The dependent variable of this study was, for each description **f**, the frequency distribution **y** of responses across the participants. More precisely, the response vector **y** was calculated on the basis of the subset of responses generated by the participants which overlapped with the family of 541 category names represented in the repository. In other words, we did not model categories beyond the semantic domain of the repository. However, we assumed that the type of category names generated by the participants could be well represented by the semantic classes described in the repository.

Parameter Estimates of the Bayesian Models

Equations 1 and 4 specify a formal approach to category identification that relies on the category priors $p(c)$ and the cue dominances $p(f|c)$. These constructs capture the background knowledge that guides the identification process, and are assumed to be specified by the semantic repository of McRae et al. (2005) that includes knowledge about basic-level categories and their features.

In particular, the repository includes familiarity ratings for the entire set of the 541 category names. We normalized the familiarity ratings for the 541 category names to create the prior distribution $p(c)$ required by our models. Of the 541 categories in the repository, Table and Pants are among the two with highest prior probability, and Bayonet and Hyena are among the two with lowest prior probability. The repository also furnished the co-occurrence (541 Categories × 2526 Features) data matrix to estimate the individual conditional probabilities $p(f|c)$ used in the Bayesian models (Equations 1,4). In particular, in the co-occurrence data matrix the entry corresponding to the pair (c,f) contained the number of subjects, $n(c,f) \leq 30$, in McRae's et al. (2005) repository who listed feature f for category c, whereas the value 30 represents the total number of subjects used to construct the counting value in the repository.[1]

We adopted a conjugate approach (e.g., Robert, 2001) for the parameter estimates of the semantic models. Let $\mathbf{f} = (f_1,f_2,f_3)$ be a generic multiple cue used to describe a target category in the stimulus set. For the naive Bayes identifier, a straightforward sampling can be obtained using the posterior Beta distributions

$$p^*(f_j|c) \sim \text{Beta}\left(\cdot \middle| \alpha + n(c,f_j), \beta + 30 - n(c,f_j)\right), \qquad j = 1,2,3 \tag{5}$$

based on the conjugate Beta(α,β) prior and the *posterior mean* estimate

$$P\text{apm}(f_j|c) = \frac{\alpha + n(c, f_j)}{\alpha + \beta + 30}, \qquad j = 1,2,3.$$

This is compatible with the assumption that the frequencies in the co-occurrence matrix of the semantic repository were independently generated according to a Beta-binomial distribution (Gupta & Nadarajah, 2004). Unlike the naive Bayes identifier, the sampling for the SSD model is obtained by means of the distribution of the reference cue

$$p^*(f_r|c) \sim \text{Beta}(\cdot|\alpha + \hat{n}_r, \beta + 30 - \hat{n}_r), \tag{6}$$

with the *posterior mean* estimate

$$P\text{apm}(f_r|c) = \frac{\alpha + \hat{n}_r}{\alpha + \beta + 30}$$

and where

$$\hat{n}_r = \min\{n(c, f_1), n(c, f_2), n(c, f_3)\}$$

which follows from the basic properties of SSD.

Finally, for each of the 24 descriptions **f** in the experiment and for each category *c* in the observed response vector **y**, the estimates of the posterior probabilities τ_c and π_c can be computed by plugging in the estimates for the priors $p(c)$ and the conditionals $p(f|c)$ in Equations 1,4 of the two Bayesian models.

RESULTS

Analysis of Responses

In the category identification task approximately 18% of the responses were left uncoded because they did not correspond to any of the 541 categories in the semantic repository, and all subsequent analyses will consider only the responses that were coded to construct the frequency distribution **y**.

Model Fitting and Evaluation

The likelihood *L* of the array of counts, **y**, can be computed using the posterior probability τ of the naive Bayes identifier (Equation 1) and the posterior probability π of the Bayesian model for strong dependent cues (Equation 4):

$$L_I(\mathbf{y}|\tau) \propto \prod_{yk \in \mathbf{y}} (\tau_{ck})^{yk} \qquad \text{independent model} \tag{7}$$

$$L_D(\mathbf{y}|\pi) \propto \prod_{yk \in \mathbf{y}} (\pi_{ck})^{yk} \qquad \text{dependent model} \tag{8}$$

where y_k is the *k*-th element in the vector **y** of response frequencies and c_k denotes the category label (concept name) associated with the frequency of occurrence y_k in **y**. Similarly, τ_{ck} (respec-

tively π_{ck}) is the posterior probability of category c_k computed using the independent model (respectively the model for strong dependent cues). Larger values for the likelihood function indicate better fits of the model to the empirical data. We used the likelihood ratio (Kass & Raftery, 1995) to measure the evidence in favor or against one of the two models as compared to the other.[2] In particular, we used the Jeffrey's (1961) scale to judge the evidence in favor of or against the naive Bayes identifier brought by the data **y**. For each of the 24 descriptions the likelihood ratio test was estimated by means of the data simulation procedure described in Appendix B.

The results for the two models are presented in Table 1. Figure 2 shows the Kullback-Leibler divergence between the posterior distributions predicted by the models and the distributions of the simulated samples **y***. In Figure 2 the pattern of the independent model largely dominates that of the dependent one, indicating that the divergence between the independent model and the simulated data was larger than that observed for the dependent model. Overall, the fit of the Bayesian model for strong dependent cues was better than the fit of the naive Bayes identifier. In particular, the strength of evidence for the dependent model resulted *decisive* in 15 out of the 20 cases considered in Table 2 (according to a 90% threshold; see Appendix A). In sum, the prediction based on the Bayesian model for dependent cues was striking compared with the naive Bayes identifier and the overall result was in line with the hypothesis that the independence assumption on which the naive Bayes identifier is based may be unwarranted when human semantic data are considered.

Discussion

Modeling how individuals integrate multiple sources of information has been a cornerstone in many theories of cognitive processes (e.g., J. R. Anderson, 1991; N. H. Anderson, 1981; Nosofsky, 1986; Shepard, 1957). Nonetheless, the modeling of integration of semantic features in the identification of natural concepts has received considerably less attention (Kemp et al., 2010; Lombardi & Sartori, 2007). The main purpose of this article was to develop and test a rational model of multiple cue integration based on the hypothesis of strong stochastic dependence of cues. SSD reflects the idea that in the integration process of multiple semantic cues the final integrated value can be reduced to the contribution of a single cue (the reference cue) the value of which is a necessary and sufficient condition to infer the posterior probability of the category. Unlike standard naive Bayes classifier, the new model replaces the assumption of stochastic independence with another complementary strong assumption: SSD.

However, it should be stressed here that our criticism about the common accepted assumption of independence in the formal representation format of concepts does not amount in any way to the conclusion that *all* concept descriptions based on semantic cues must necessarily be characterized by strong dependent patterns or that in general other probabilistic classifiers cannot account for correlated features (see for example, Friedman, Geiger, & Goldszmidt, 1997; Rosseel, 2002, for some relevant instances of probabilistic classifiers for correlated features). On the contrary, our target has been focused on empirically testing the allegedly uninfluence of accepting stochastic independence in the construction of formal theories of category identification. In order to evaluate the reliability of the independence assumption we contrasted two opposite Bayesian models that represent only formal approximations of the true, but unknown, semantic integration process. Overall our results showed that a) independence between semantic features

TABLE 1
Percentage of results in favor of the Bayesian model for strong dependent cues based on the 1000 simulated samples

Item	% Decisive $\log_{10}(R_{DI}) > 2$	% Strong $1 < \log_{10}(R_{DI}) \leq 2$	% Substantial $0.5 < \log_{10}(R_{DI}) \leq 1$	% Poor $0 < \log_{10}(R_{DI}) \leq 0.5$
2	31.8	16.4	10.4	11.5
3	90.1	8.8	1.1	0.0
4	95.8	1.5	0.8	0.2
5	100.0	0.0	0.0	0.0
7	98.1	0.5	0.1	0.4
8	99.8	0.0	0.0	0.1
10	100.0	0.0	0.0	0.0
11	99.9	0.0	0.1	0.0
12	22.8	17.9	11.9	14.8
14	84.1	5.7	2.8	2.2
15	99.9	0.1	0.0	0.0
16	98.6	0.8	0.5	0.0
17	70.7	10.6	5.6	2.7
18	99.2	0.3	0.0	0.2
19	97.0	1.0	0.2	0.3
20	78.0	12.1	3.5	2.5
21	97.3	0.8	0.4	0.3
22	99.7	0.1	0.0	0.0
23	97.3	1.3	0.6	0.2
24	100.0	0.0	0.0	0.0

Note. We used the Jeffrey's scale to interpret the result of the likelihood ratio test. R_{DI} denotes the ratio between the likelihoods of the Bayesian model for strong dependent cues and the naive Bayes identifier. The analysis for items 1,6,9, and 13 has not been performed as the response vectors **y** for these items contained only one category value *y* (constant response across participants) which, in turn entails that the performances of the two models are equivalent.

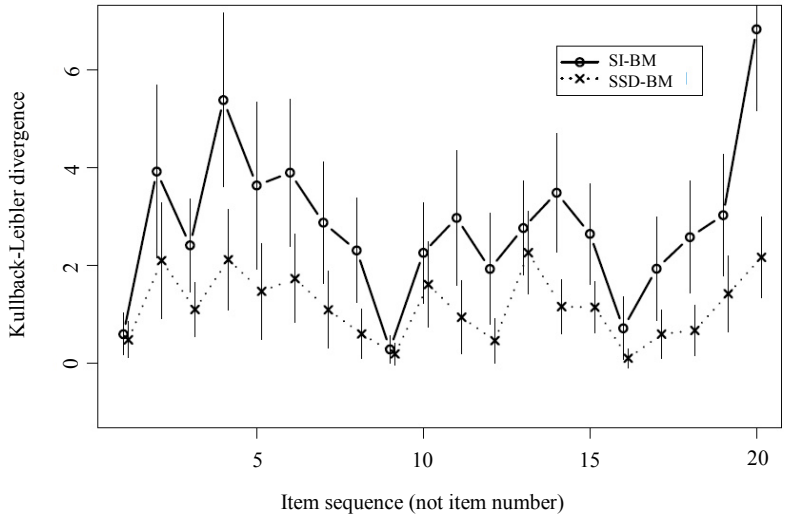

FIGURE 2
Kullback-Leibler divergence plots. Average representation based on the 1000 simulated samples, the vertical segments indicate the average divergences' standard deviations.
SI-BM = naive Bayes identifier; SSD-BM = Bayesian model for strong dependent cues.

seems more the exception than the rule b) although SSD represents the most extreme instance of the broad range of realizations of cue dependency, it often suffices in explaining the observed data pattern.

Finally, as a model for semantic cue integration our proposal is closely related to the causal-model theory of conceptual representation and categorization (Rehder, 2003; Rehder & Burnett, 2005; Rehder & Kim, 2006). In particular, SSD has a close relationship with the *common cause network* described by Rehder and Kim (2006). In a common cause network with k distinct features, one feature (the *primary cause*) is described as the cause of all the other k-1 features (the *peripheral features*) in the network. One obtains a causal-model for the integration of strong dependent cues if the parameters representing the weights that connect the peripheral features with the primary cause are all set to one (1) in the network. In this latter case it is straightforward to show that the likelihood equation for the common cause network reduces to Equation 3. However, the causal-model theory of categorization has been developed for modeling data collected from researches studying the acquisition of artificial categories in the laboratory in which stimuli are characterized by few selected features with uniform prior probability distribution. From its part our Bayesian models are feature-based models of non-artificial concept descriptions which are represented by high-dimensional co-occurrence data matrices and explicit Bayesian rules to rigorously integrate the likelihoods with the priors.

NOTES

1. In the McRae's et al. (2005) repository all positive co-occurrences $n(c,f)$ are represented by at least five subjects (cutoff value = 5). The cutoff avoids the presence of idiosyncratic features in the repository.
2. For sake of conciseness, but with some abuse of notation, we removed the multinomial coefficients from the likelihood equations (Equations 7,8). Of course, the removal of the coefficients does not affect in any way the computation of the likelihood ratio.

AKNOWLEDGMENTS

We thank Ken McRae and his colleagues for creating and sharing the data set of semantic concepts. Charles Kemp, Marco Baroni, and Luigi Burigana made helpful suggestions about early version of this paper.

REFERENCES

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409-429.
Anderson, N. H. (1981). *Foundations of information integration theory*. New York, NY: Academic Press.
Ashby, F. G., & Alfonso-Reese, L. A. (1995). Categorization as probability density estimation. *Journal of mathematical Psychology*, *39*, 216-233. doi:10.1006/jmps.1995.1021
Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154-179.
Duda, R. O., & Hart, P. E. (1973). Pattern classification and scene analysis. New York, NY: Wiley.
Friedman, N., Geiger, D., & Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, *29*, 131-163. doi:10.1023/A:1007465528199
Gupta, A. K., & Nadarajah, S. (2004). Handbook of Beta distribution and its applications. New York, NY: Marcel Dekker.
Jeffrey, H. (1961). Theory of probability (3rd ed.). Oxford, UK: Oxford University Press.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *430*, 773-795. doi:10.1080/01621459.1995.10476572

Kemp, C., Chang, K., & Lombardi, L. (2010). A probabilistic account of category and feature identification. *Acta Psychologica*, *133*, 216-233. doi:10.1016/j.actpsy.2009.11.012

Kemp, C., & Jern, A. (2014). A taxonomy of inductive problems. *Psychonomic Bulletin & Review*, *21*, 23-46. doi:10.3758/s13423-013-0467-3

Lombardi, L., & Sartori, G. (2007). Models of relevant cue integration in name retrieval. *Journal of Memory and Language*, *57*, 101-125. doi:10.1016/j.jml.2006.10.003

Love, B. C., & Markman, A. B. (2003). The nonindependence of stimulus properties in human category learning. *Memory & Cognition*, *31*, 790-799. doi:10.3758/BF03196117

Mandler, G. (1980). Recognizing: The judgement of previous occurrence. *Psychological Review*, *87*, 252-271. doi:10.1037/0033-295X.87.3.252

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers*, *37*, 547-559. doi:10.3758/BF03192726

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, *85*, 207-238. doi:10.1037/0033-295X.85.3.207

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology*: *General*, *115*, 39-57. doi:10.1037/0096-3445.115.1.39

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1141-1159. doi:10.1037/0278-7393.29.6.1141

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. *Cognitive Psychology*, *50*, 264-314. doi:10.1016/j.cogpsych.2004.09.002

Rehder, B., & Kim, S. W. (2006). How causal knowledge affects classification: A generative theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 659-683. doi:10.1037/0278-7393.32.4.659

Robert, P. R. (2001). The Bayesian choice (2nd ed.). New York, NY: Springer.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Erlbaum.

Rosseel, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, *46*, 178-210. doi:10.1006/jmps.2001.1379

Shepard, R. N. (1957). Stimulus and response generalization for psychological science. *Psychometrika*, *22*, 325-345. doi:10.1007/BF02288967

Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*, 189-228. doi:10.1207/s15516709cog2202_2

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral Brain Sciences*, *24*, 629-641. doi:10.1017/S0140525X01000061

APPENDIX A

Table 2 reports the list of items used in the category identification task.

TABLE 2
Items of the category identification task

Item	Category name	1st attribute	2nd attribute	3rd attribute
1	Anchor	made-of-metal	found in water	used on boats
2	Vulture	beh-lays eggs	is large	has wings
3	Birch	has bark	has leaves	is tall
4	Bracelet	bought/sold in stores	is round	made of metal
5	Buggy	has a handle	is old fashioned	made of metal
6	Elephant	lives in Africa	has a trunk	is grey
7	Rooster	is white	is red	a bird
8	Grater	used for shredding things	found in kitchens	used for food
9	Grater	has holes	used for grating cheese	is sharp
10	Jet	is fast	an airplane	used for passengers
11	Shelves	used for storing books	used for storage	is white
12	Cow	has four legs	used for producing milk	mammiferi
13	Coconut	grows on trees	has an outside	has milk
14	Olive	is red	is edible	is small
15	Football	made of pig skin	is oval	used by throwing
16	Wall	is solid	used for holding pictures	is flat
17	Cockroach	is small	is black	beh-flies
18	Pin	is small	has a head	is silver
19	Boots	is brown	is black	is long
20	Mole	a rodent	is short sighted	lives in ground
21	Tiger	lives in India	beh-roars	is orange
22	Tray	made of plastic	is rectangular	used for carring drinks
23	Fox	is wild	hunted by people	has four legs
24	Pumpkin	grows on vines	is round	used for Jack O'Lanterns

APPENDIX B

For each description/response-set pair (**f**,**y**) the likelihood ratio test was estimated by means of the following data simulation procedure:

S1. First, a collection of 1000 samples $\boldsymbol{\theta}^*$ were generated from the posterior Dirichlet distribution $\mathrm{Dir}(\cdot|\alpha+\mathbf{y})$ with Jeffrey's prior $\alpha = 1$. Next, for each sample $\boldsymbol{\theta}^*$ we simulated a new vector of counts $\mathbf{y}^*$ by sampling from the multinomial distribution $\mathrm{Mult}(\cdot|\boldsymbol{\theta}^*,N)$ with N being the sum $N = \Sigma_k y_k$ of the observed counts in **y**.

S2. *Naive Bayes identifier*: For each pair (f_j,y_k) with f_j in **f** and y_k in **y**, 1000 simulated conditional probability values $p^*(f_j|c_k)$ were generated by independently sampling from the posterior Beta distribution detailed in Equation 5 with Jeffrey's priors $\alpha = \beta = 1$. Therefore, for each simulated sample we derived the posterior probability array τ^* (Equation 1).

S3. *Strong dependent model*: From the simulated conditional probabilities generated in Step 2 we kept only those corresponding to the reference cues f_r: $p^*(f_r|c_k)$ (for all y_k in **y**). In particular, given a description **f**, the reference cue f_r for a target category c_k is given by

$$f_r = \arg\min_{f\in\mathbf{f}}\{n(c_k f)\}.$$

Therefore, for each of the 1000 simulated reference probabilities the posterior probability array π^* was derived on the basis of Equation 4.

S4. Finally, for each of the 1000 combined samples $(\mathbf{y}^*,\tau^*,\pi^*)$ the likelihood ratio test, $\log_{10}(L_D/L_I)$, was computed by using Equations 7,8.

In the data simulation procedure we did not model uncertainty for the category priors $p(c)$ as the repository of McRae and colleagues (2005) does not contain sufficient details to guess the stochastic process that generated the familiarity ratings.