

Trial-by-trial identification of categorization strategy using iterative decision-bound modeling

Sébastien Hélie¹ · Benjamin O. Turner² · Matthew J. Crossley³ · Shawn W. Ell⁴ · F. Gregory Ashby²

© Psychonomic Society, Inc. 2016

Abstract Identifying the strategy that participants use in laboratory experiments is crucial in interpreting the results of behavioral experiments. This article introduces a new modeling procedure called iterative decision-bound modeling (iDBM), which iteratively fits decision-bound models to the trial-by-trial responses generated from single participants in perceptual categorization experiments. The goals of iDBM are to identify: (1) all response strategies used by a participant, (2) changes in response strategy, and (3) the trial number at which each change occurs. The new method is validated by testing its ability to identify the response strategies used in noisy simulated data. The benchmark simulation results show that iDBM is able to detect and identify strategy switches during an experiment and accurately estimate the trial number at which the strategy change occurs in low to moderate noise conditions. The new method is then used to reanalyze data from Ell and Ashby (2006). Applying iDBM revealed that increasing category overlap in an

information-integration category learning task increased the proportion of participants who abandoned explicit rules, and reduced the number of training trials needed to abandon rules in favor of a procedural strategy. Finally, we discuss new research questions made possible through iDBM.

Keywords Decision-bound modeling · Response strategy · System switching · Perceptual category learning

Identifying the strategy used by participants in laboratory experiments is crucial in interpreting the results of behavioral experiments. For example, participants can add numbers either by using an algorithm or memory retrieval (if they remember calculating that sum in the past) in an experiment requiring arithmetic skills (Logan, 1988). Without the possibility of identifying which strategy is being used, it is difficult to interpret dependent variables such as response times and response accuracy. Ashby and Gott (1988) first proposed the “randomization technique” to address this issue in perceptual categorization. This article extends their technique in order to detect the presence and moment of changes in response strategy.

In a basic categorization experiment, the participant sees one stimulus on each trial, and assigns each of these stimuli to a category by pressing a response key (e.g., either the key associated with a category A response or the key associated with a B response). Feedback indicating the correctness of the response is typically (but not always) presented after each response. The randomization technique can be used to identify the response strategy used with stimuli that vary on any number of perceptual dimensions, but most applications have used stimuli that vary on only two dimensions.

✉ Sébastien Hélie
shelie@purdue.edu

¹ Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA

² Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA, USA

³ Information and Computing Sciences, SRI International, Menlo Park, CA, USA

⁴ Department of Psychology, University of Maine, Orono, ME, USA

Some examples are shown in Fig. 1. In this figure, each symbol denotes the values of the stimulus on the two stimulus dimensions, and the symbol color denotes the category membership of each stimulus. The diagonal line describes the optimal decision strategy.

Figure 1 color codes each stimulus according to its true category membership. An alternative representation, shown in Fig. 2, is to color code each stimulus according to the response the participant made when that stimulus was presented. We will denote this representation as the *decision space*. Whereas the stimulus space is defined by the experimenter and typically is the same for all participants, each participant generates his or her own decision space. Example decision spaces for two hypothetical participants presented with the Fig. 1 categories are shown in Fig. 2.

In traditional decision-bound modeling (DBM, Ashby, 1992; Maddox & Ashby, 1993), the experimenter fits a number of statistical models to the data in the decision space in an attempt to determine the type of decision strategy that the participant used. DBM assumes that participants partition the perceptual space into response regions. On every trial, the participant determines which region the percept is in, and then emits the associated response. Three different types of models are typically fit to the responses of each individual participant: models that assume the participant guessed at random on every trial, models that assume an explicit reasoning strategy, and models that assume a procedural strategy. None of these models make detailed process assumptions, in the sense that a number of different process accounts are compatible with each of the models (e.g., Ashby, 1992). For example, if a procedural model fits significantly better than an explicit reasoning model, then

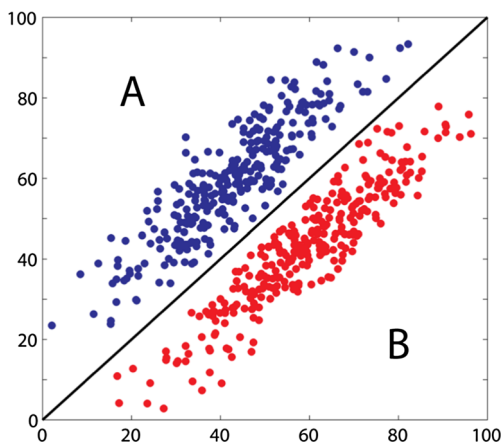


Fig. 1 Example of two-dimensional stimulus space. Each symbol denotes a different stimulus. The *coordinates* of each symbol denote the numerical values of the stimulus on the two stimulus dimensions, and the *symbol color* denotes the category membership of each stimulus. The *diagonal line* describes the optimal decision strategy

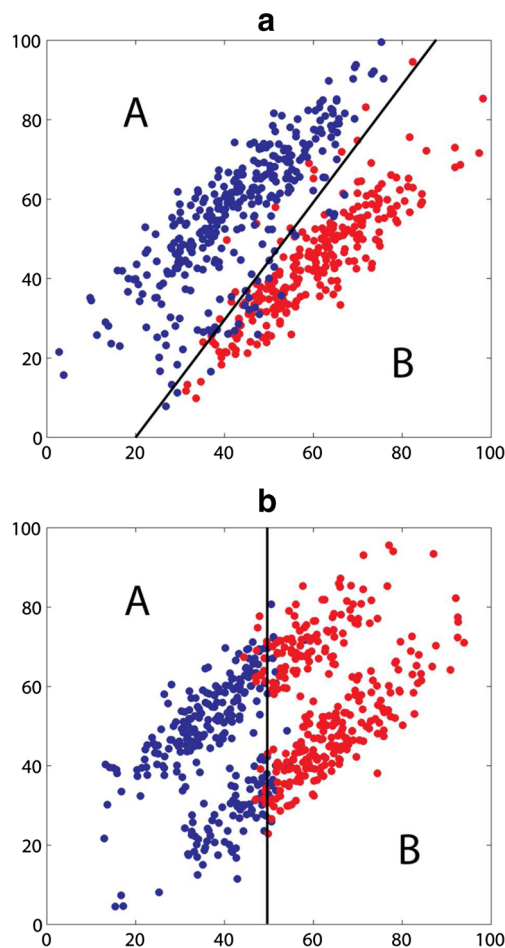


Fig. 2 Hypothetical decision spaces from two participants. As in Fig. 1, the *coordinates* of each symbol denote the numerical values of the stimulus on the two stimulus dimensions. In this case, however, the *color* of each symbol denotes the category decision made by the participant. The *solid line* bound describes the best-fitting decision-bound model

we can be confident that participants did not use a simple explicit rule (e.g., “respond A if the stimulus value on dimension x is low and respond B if the value is high”), but we could not specify which specific non-rule-based strategy was used (e.g., a weighted combination of the two dimensions versus more holistic memory-based processing). In the Fig. 2 examples, the participant whose decision space is shown in panel (a) is classified as using a procedural strategy, whereas the panel (b) participant is classified as using a simple explicit rule.

DBM has been used in hundreds of articles during the past 25 years to identify the response strategies of participants in perceptual categorization experiments. In head-to-head comparisons, it frequently provides better fits to categorization data than the best exemplar models (Ashby & Lee, 1991; Maddox & Ashby, 1993; Maddox et al., 2002), and it successfully classifies participants by strategy use

(i.e., explicit versus procedural) in a way that accounts for more than 20 different empirical dissociations between rule-based (RB) and information-integration (II) categorization tasks (e.g., Ashby & Valentin, 2016).¹ Even so, one serious limitation of DBM is that it assumes that the participant uses the same decision strategy on all trials used during the model-fitting process. Of course, we expect the participant's decision strategy to change during the course of learning, and a number of studies have documented such changes (Haider & Frensch, 1996; Kalish et al., 2005; Logan, 1988). Overall, these results suggest that participants change their response strategy when two conditions are met: (1) a sufficient number of errors have occurred, and (2) the participant is aware of an alternative strategy. Since these conditions are often satisfied in categorization experiments, strategy changes are likely common, and unfortunately difficult to detect.

This article introduces a new modeling procedure, called iterative decision-bound modeling (iDBM), which fits DBMs in a trial-by-trial iterative fashion. The goal of this new approach is to identify: (1) all response strategies used by a participant, (2) changes in response strategy, and (3) the trial number at which each change occurs. The remainder of this article is organized as follows. First, the next section briefly reviews DBM. Second, iDBM is introduced and validated using benchmark simulated data. Third, the data reported by Ell and Ashby (2006) are re-analyzed using iDBM. This study explored the effects of varying amounts of category overlap on response strategies during perceptual categorization. The original article used standard DBM to determine the decision strategy of each participant. The new analysis with iDBM allows us to sort participants in the same way, but it also allows us to estimate how many times each participant switched strategies and the trial number at which each switch occurred. Hence, new conclusions can be drawn about the effects of category overlap on strategy selection. Fourth, the article concludes by discussing new research questions made possible by iDBM.

Decision-bound models

Decision-bound models are a special case of general recognition theory (GRT, Ashby & Soto, 2015; Ashby & Townsend, 1986), which is a multidimensional generalization of signal detection theory. As in GRT, DBMs assume that both the stimuli and the perceptual systems are noisy.

¹RB tasks are those where optimal accuracy is possible with a simple explicit rule (e.g., "stimuli with thick bars are in A and stimuli with thin bars are in B"), whereas II tasks are those in which optimal responding requires participants to integrate information from more than one stimulus dimension at a pre-decisional level (e.g., as in Fig. 1). For a more thorough discuss, see e.g., Ashby and Maddox (2005).

Hence, every time a stimulus is presented it elicits a new and unique percept, even if the stimulus has been previously encountered. Each percept is represented by a point in a multi-dimensional perceptual space (one dimension for each stimulus dimension), and the set of all possible percepts is represented by a multivariate probability distribution. GRT and DBMs assume that the participant's decision process divides the perceptual space into response regions. On each trial, the decision process notes which region the percept is in and then emits the associated response.

GRT is often applied to identification experiments in which the stimuli are all highly confusable. In this case, errors are often made because of perceptual confusions. As a result, GRT models typically allocate many parameters to the perceptual distributions. For example, it is not uncommon to allow the means of each perceptual distribution to be free parameters and to allow the perceptual distributions associated with the different stimuli to all have different variances and covariances (e.g., Ashby & Soto, 2015). In category-learning experiments like the one illustrated in Fig. 1, perceptual confusions are inevitable. However, most errors are not caused by such confusions, but rather by the application of a suboptimal decision strategy. For this reason, DBM uses a highly simplified perceptual representation relative to the most general versions of GRT. In particular, DBM assumes that the mean of each perceptual distribution equals the stimulus coordinates and that all perceptual distributions have equal variances on every perceptual dimension and that all covariances equal zero. These assumptions leave only one free perceptual parameter—namely the common perceptual variance, denoted by σ_p^2 .

All DBMs make the same perceptual assumptions. Three different classes of models can be constructed depending on what assumptions are made about the decision process. These three classes are described in the following subsections.

Explicit-reasoning models

Explicit reasoning models assume the participant uses an explicit rule that is easy to describe verbally (Ashby et al., 1998). This means that all models in this class only include decision bounds that are perpendicular to one or more stimulus dimensions. The most widely used models in this class assume that participants set a decision criterion on a single dimension. For example, a participant might base his or her categorization decision on the following rule: "Respond A if the bars are narrow; respond B if the bars are wide." These one-dimensional classifiers have two free parameters: a decision criterion along the relevant perceptual dimension (e.g., bar width), and the perceptual noise variance. Figure 2b shows responses from a hypothetical participant that were best fit by a DBM of this type.

More complex explicit-reasoning models assume a two-dimensional rule such as a logical conjunction (e.g., “Respond A if the bars are narrow and the bar orientation is steep; otherwise respond B”) or a unidimensional disjunction (e.g., “Respond A if the bars are of intermediate width; respond B if the bars are narrow or thick”). Both of these more complex model examples have three free parameters: one for each decision criterion, as well as a (common) perceptual noise variance.

Procedural-learning models

Procedural-learning models assume that perceptual information from all relevant dimensions is integrated *before* a decision is made. This sharply contrasts with explicit-reasoning models, which assume participants make separate decisions about each relevant stimulus dimension, and then combine these decisions if more than one criterion is needed. The integration of information in a procedural-learning model can be linear or nonlinear. The most common application assumes linear integration, and the resulting model is known as the general linear classifier (GLC). The GLC assumes that participants divide the stimulus space using a linear decision bound. One side of the bound is associated with an “A” response, and the other side is associated with a “B” response. The GLC has three parameters: the slope and intercept of the linear decision bound, and a perceptual noise variance. The responses of the hypothetical participant illustrated in Fig. 2a are best fit by the GLC.

Note that diagonal bounds encourage procedural response strategies because they are incompatible with explicit verbal rules. For instance, “respond A if the bars are thicker than they are steep” is a verbal rule that describes a diagonal bound, but this rule has low saliency because bar thickness and bar angle are non-commensurable stimulus dimensions.

Guessing models

Guessing models assume that the participant guesses randomly on every trial. All versions assume the probability of responding “A” (and therefore also the probability of responding “B”) is the same for every stimulus. As a result, perceptual noise cannot change these predicted probabilities and so there is no need to account for perceptual noise in the guessing models. Because of this, guessing models do not include a noise variance parameter. Two types of guessing models are common. One version assumes that each response is selected with equal probability. This model has no free parameters. A second model, with one free parameter, assumes that the participant guesses response “A” with probability p and guesses “B” with probability $1 - p$, where p is a free parameter. This model is useful for identifying

participants who are biased toward pressing one response key.

Model fitting

In regular applications of DBM, each model is fit separately to the responses of every individual participant. The data of each participant are usually divided into blocks of at least 50 trials, but blocks of 100 or more trials are common. The models are then fit to single blocks of data. Model parameters are estimated using maximum likelihood (Ashby, 1992), and the best fitting model is typically selected via the Bayesian information criterion (BIC):

$$\text{BIC} = r \times \ln(N) - 2 \times \ln(L) \quad (1)$$

where N is the block size, r is the number of free parameters in the model, and L is the likelihood of the data given the model (Hélie, 2006). The BIC statistic penalizes a model for bad fit and for extra free parameters. To find the best model among a set of competitors, one simply computes a BIC value for each model and then chooses the model with the smallest BIC. It is important to note that we only use BIC to penalize models for additional free parameters. We are not assuming that the true model is one of the fitted models, and therefore we make no attempt to use BIC scores to estimate model probability or likelihood (Neapolitan, 2004).

Limitations of DBM

As mentioned earlier, one inherent limitation of the DBM method is that it assumes that the participant uses the same decision strategy on every trial. This assumption makes the most sense for later blocks in the experiment—after accuracy has reached asymptote. For the earliest blocks, however, this assumption is probably almost always false. To minimize the problems that can arise from strategy switching, applications of DBM that fit models to data collected early in a session usually reduce block size down to 50 trials or so. But even this strategy will fail unless the participant happens to coincidentally only switch strategies on trials 51 and 101. Otherwise, the responses in the first block of trials, for example, will be generated from multiple different decision strategies, and as a result, all decision-bound models will provide poor fits to the data. The next section introduces a new fitting method called iDBM that addresses these limitations.

Iterative decision-bound modeling (iDBM)

Early models that assumed category learning is mediated by multiple systems, including both COVIS and ATRIUM, assumed trial-by-trial switching between the competing

systems (Ashby et al., 1998; Erickson & Kruschke, 1998). No statistical method could possibly identify such frequent strategy switches. However, more recent studies have reported that trial-by-trial strategy switching is extremely difficult—in fact, so difficult that few participants succeed even after training (Ashby & Crossley, 2010; Erickson, 2008). These data suggest that strategy switches are less common than originally thought, and therefore that they might be identified via a statistical modeling approach. iDBM is an attempt to solve this statistical problem.

iDBM is a generalization of classical DBM that attempts to identify every strategy change by each participant. The end result is a list of all strategies used by the participant, along with the trial numbers during which each strategy was used. Thus, iDBM also tries to identify the exact trial number when each strategy shift occurred. This extra information, not provided by classical DBM, may provide a better understanding of individual differences in category learning and allow for stronger tests of category-learning theories and models. For example, COVIS (Ashby et al., 1998) predicts that participants are biased towards the use of simple rules early in category learning, and that only after they have abandoned explicit strategies will they consider procedural strategies, even in II tasks where procedural strategies are optimal. Unfortunately, this strong prediction of COVIS is largely unexplored due to the unavailability of a method capable of observing strategy shifts early in learning.

iDBM uses classical DBM as its starting point. Define the DBMs described above as *basic models*. With stimuli that vary on two dimensions, there are four basic models: (1) guessing, (2) one-dimensional explicit rule models that base their response on the value of the stimulus on the x -dimension (1DX), (3) one-dimensional rule models that base their response on the value of the stimulus on the y -dimension (1DY), and (4) the GLC. iDBM assumes that during the course of the experimental session, each participant switches among basic models some unknown (although reasonably small) number of times. Thus, both the number of strategy switches, and the trial numbers when those switches occur are unknown. For an entire session of data, this leaves open an enormous number of possible combinations of basic models. iDBM uses an iterative approach to sort through all these and find the one combination that provides the best account of the responses of each individual participant.

A basic model is a single DBM. A *switch model* is a model composed of 2 basic models, and a parameter η identifying the switch trial (the trial on which a strategy switch occurs). For example, a switch model may correspond to a participant guessing from the first trial of the block through trial $\eta - 1$ and then switching to a 1DX strategy beginning on

trial η and continuing until the end of the block. With the 4 basic models described above, there are $4 \times 4 = 16$ possible switch models. These include cases where the participant switches between two different instances of the same basic model (e.g., switching from a GLC with a positive slope to a GLC with a negative slope). The number of free parameters for switch models is the sum of the number of free parameters of the basic models composing the switch model plus one (i.e., η). Hence, with the basic models defined in this article, the number of free parameters in switch models varies from 3 (guessing \rightarrow guessing) to 7 (GLC \rightarrow GLC).

The fitting algorithm is described in Table 1. As can be seen, the general idea is to first load the data into a fit window and then fit all the switch models and the basic models to the data in this window. Note that η is included as a regular parameter and estimated at the same time as the model parameters. The best-fitting model by BIC is then selected (Eq. 1). The fit window size is then increased by one trial and the models are all refit. This is repeated until the algorithm settles on the same best-fitting model for s consecutive iterations and the estimated switch trial (η) falls consistently within a predefined range. When these conditions are met, a strategy switch is identified according to the consistently selected model, and the switch trial is set to the average of the estimated switch trials (η) for the consistent run. The fit window is then reduced back to its original size and slid to begin on the identified strategy switch trial (η). The fitting procedure is then repeated until one runs out of data. Note that it is possible that a basic model is consistently identified as the best-fitting model. This would indicate that the participant consistently used a single response strategy throughout the entire session (in this case, η is not estimated).

As an illustration, suppose we choose an initial fit window size of $n = 50$ and we set the number of consecutive consistent model iterations to $s = 10$. iDBM would then take trials 1...50 and fit all the basic models and all the switch models and keep the best model according to BIC. For example, the best fit might be by a switch model that switches from a one-dimensional strategy on dimension X to the GLC (i.e., 1DX \rightarrow GLC) on trial 35 (i.e., $\eta = 35$). The procedure would then be repeated with trials 1...51, and then with trials 1...52, and so on until the same best model is selected at least 10 consecutive times and the variance of the estimated η s is smaller than a predefined criterion (i.e., t in Table 1). At that point, a switch is identified using the selected model (e.g., from 1DX to GLC), and the estimated switch trial (η) is set to the mean of the estimated η s for the last 10 model fits (i.e., the run of consistent model selections). The fit window size is then reset to 50, and slid up to begin on trial η . iDBM then takes trials η ...($\eta + 49$) and starts over (i.e., fits all the basic and switch models, selects

Table 1 Fitting algorithm for iDBM

-
1. Initialize variables:
 - (a) Define the set of models to be fit; e.g., $M = \{\text{basic models} \cup \text{switch models}\}$.
 - (b) Set the first trial of the fit window to $m = 1$.
 - (c) Set the minimum fit window size to n trials.
 - (d) Set the minimum number of consecutive identifications of the same best-fitting model to s .
 - (e) Set the threshold on the variance of η for the s consecutive trials where the same model is identified as the winning model to t .
 - (f) Set up two empty lists called *Winners* and *Final*.
 2. For trials m to $(m + n - 1)$:
 - (a) Fit each model in M and record the BIC.
 - (b) Denote the model with the smallest BIC as the winner and add the name of this model to the end of the *Winners* list.
 - (c) If the length of the *Winners* list $\geq s$ AND the last model in this list is a switch model AND the last s models in the list are all the same model AND $\text{Var}(\eta) < t$ for the last s models in the list then:
 - i. Add the name of the last model in the *Winners* list to the end of the *Final* list, with η set to the mean of the η values of the last s models in the *Winners* list.
 - ii. Set m to the mean η value identified in Step 2ci.
 - iii. Empty the *Winners* list.
 - iv. Reset n to its original value.
 - v. Go back to Step 2.
 - (d) Increment n by 1 (i.e., replace n with $n + 1$).
 - (e) Go back to Step 2 until all data have been used.
 3. All the models that best account for part of the data are listed in *Final*, along with the estimated switch trials (η s).
-

the best model, increases the fit window size by 1, etc.). This procedure is repeated until one runs out of data.

Testing iDBM with simulated data

Testing iDBM by using an optimal responder with changing stimulus distributions

We first tested iDBM by applying it to a set of simulated data that included 600 trials (a typical number for one session of a perceptual categorization experiment). The data were constructed by first assuming that 500 hypothetical participants switched among a number of alternative decision strategies. From trials 1–100, the hypothetical participants all used a guessing strategy (Fig. 3, top left). On trial 101, the participants all switched to a one-dimensional rule strategy on the x dimension, and persisted with this strategy until trial 350 (Fig. 3, top right). Finally, during trials 351–600, the participants all used a linear procedural strategy (Fig. 3, bottom). Noise was added to the data by assuming that the simulated participant abandoned the operative iDBM strategy and made a random guess on 0 %, 20 %, 40 %, 60 %, or 80 % of the trials. Each noise condition was simulated 100 times (i.e., 100 simulated participants for each noise condition).

Note that the way we added noise to the simulated data violates the assumptions of all DBMs, and therefore of all basic models that are used in iDBM.² In DBM, normally distributed noise changes the percept, which means that it changes the coordinates of the stimulus in stimulus space. Because of the normality assumption, small changes are more likely than large changes. All stimuli are presumed to be equally affected, and these changes can cause the model to switch responses on any given trial. But noise induced response changes are much more likely for stimuli near the decision bound than for stimuli far from the bound because a response change can only occur when noise moves a percept to the opposite side of the decision bound. In contrast, in our simulated data, every response was equally likely to be a guess, regardless of how far the percept was from the decision bound. Testing models against incompatible rather than compatible data provides a stronger and more

²Technically the simulated data are compatible with an iDBM in which strategy switching occurs on almost every trial—for example, back and forth between guessing and the GLC. However, this model is non-identifiable since it would likely have more free parameters than data points. To protect against this problem we assume a lower bound on the number of consecutive trials during which a response strategy is used (e.g., 30 trials). With this added restriction, the simulated data are incompatible with all iDBMs (except for the 0 % noise condition).

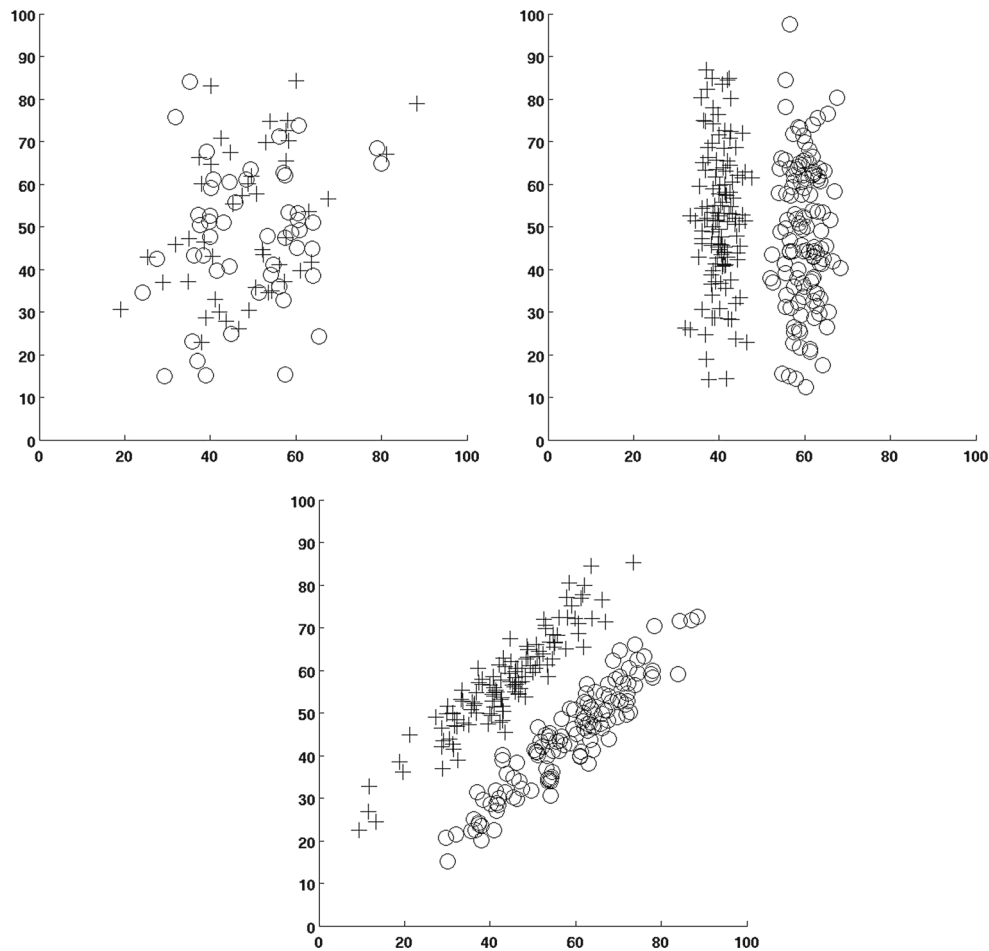


Fig. 3 Simulated data used to evaluate iDBM (before any noise is added). *Symbols* denote different categories. Trials 1–100 used a guessing strategy (*top left*), trials 101–350 used a one-dimensional

rule strategy on the x dimension (*top right*), and trials 351–600 used a procedural strategy (*bottom*). The data from all three panels were concatenated and then noise was added

realistic test of the models, since real data are likely to be somewhat incompatible with all tested models. Our noise conditions were designed in an attempt to create data sets that were increasingly incompatible with iDBM, yet based on a simple and compelling underlying structure.

The algorithm from Table 1 was directly applied to the data from each of the 500 simulated participants in succession with the following settings. All the models described above were fitted (16 switch models + 4 basic models = 20 models), the minimum window size was set to $n = 100$, the minimum number of consecutive model identifications was set to $s = 15$, and the threshold on the variance of estimated η s was set to $t = s + 6 = 21$. Note that none of these parameters were optimized. Instead they were set to reasonable values based on previous experience with the fitting procedure used in the 100×100 decision space. These parameters should thus be adequate for any stimuli that are re-scaled into an arbitrary 100×100 coordinate space.

Simulation results

The optimal model-fitting results with the simulated data would be to detect two strategy switches: (1) a switch from guessing to 1DX on trial 101 and (2) a switch from 1DX to GLC on trial 351. Figure 4 shows the mean number of switches and the trial number of the first and last switches identified by iDBM for each condition. As can be seen, the correct number of switches was identified fairly accurately for the 0 % and 20 % noise conditions, with a mean number of switches of 2.09 and 1.98, respectively. Higher noise conditions resulted in underestimation, with a mean number of switches of 1.61 for noise condition 40 %, 1.57 for noise condition 60 %, and 0.31 for noise condition 80 %. In the 80 % noise condition, switches were only detected in 18 out of the 100 simulations. This explains why the mean number of switches is so low (i.e., there are many zeros). In contrast, at least one strategy switch was detected in all the

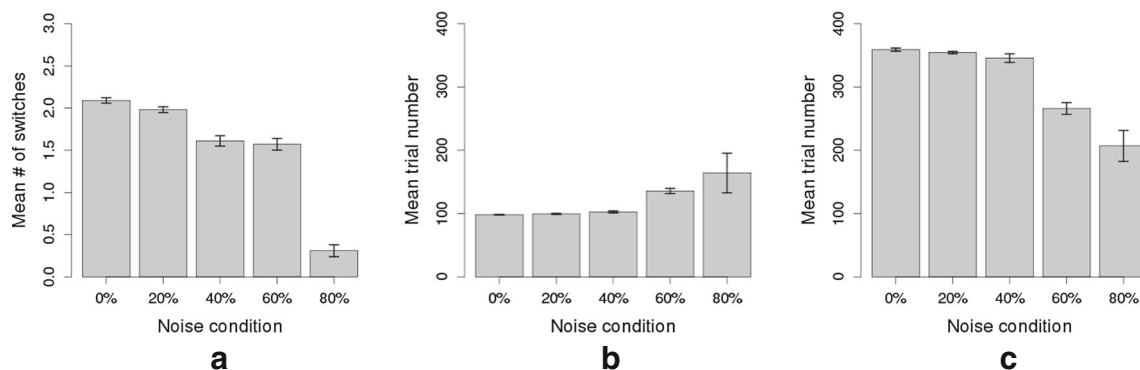


Fig. 4 Mean number of estimated switches for each noise condition. (a) Mean number of switches detected. True number of switches was 2. (b) Estimated trial number of the first switch when at least one switch is detected. True switch trial was trial 101. (c) Estimated trial number

simulations for all other noise conditions. This shows that iDBM underestimated strategy switching in high noise conditions, rather than overestimated, which suggests that with these parameter settings, iDBM is conservative in detecting strategy switches (and therefore avoids false positives).

Figure 4b shows the mean estimated trial number of the first strategy switch for all conditions when at least one strategy switch was detected. As can be seen, the mean estimated trial number of the first switch is accurately estimated for the 0%, 20%, and 40% noise conditions (98.34, 99.63, and 102.81, respectively). This suggests that the first switch (i.e., from guessing to 1DX) is accurately detected for the 40% noise condition, and that the switch that is sometimes not detected is the second switch (i.e., from 1DX to GLC). For higher noise conditions, the estimated first switch trials are higher than expected, with estimates of 135.94 and 164.11 for the 60% and 80% noise conditions, respectively. However, strategy switches were also not always detected in high noise conditions.

Figure 4c shows the mean estimated trial number of the last strategy switch for all conditions when more than one switch was detected, which occurred in 100, 94, 55, 48, and 12 simulations for noise conditions 0, 20, 40, 60, and 80%, respectively. Similar to the first-switch analysis, the trial of the last switch was accurately estimated in the 0, 20, and 40% noise conditions (359.12, 354.46, and 345.71, respectively). However, last switch estimates for higher noise conditions were underestimated, suggesting that the switch from 1DX to the GLC is difficult to detect in high-noise conditions and is often missed. The trial estimates of the last switch were 266.02 and 206.75 for noise conditions 60% and 80%, respectively.

The results described so far focused on the number of detected switches and the trial numbers at which switches were detected. However, it remains to be seen if the identified switches correspond to the correct strategies. As a

reminder, the first identified switch should be Guessing to 1DX and the last identified switch should be 1DX to GLC. The first identified switches for conditions 0, 20, and 40% are shown in Fig. 5.³ As can be seen, the correct first strategy switch was identified in at least 96% of the simulations for all noise conditions. When iDBM incorrectly identifies the strategy switch, the incorrect strategy switch is typically identified as Guessing to GLC.

The last identified switch models (when at least two switches were detected) are shown in Fig. 6. The correct 1DX to GLC switch was identified in most cases for all noise conditions (0, 20, and 40%). For noise conditions of 0 and 20%, the correct strategy was identified 82 and 97% of the time (respectively), and for noise condition 40%, the correct strategy was identified 66% of the time. However, this second switch was only detected in 55% of the simulations in the 40% noise condition (compare with 100 and 94% of the simulations in the 0 and 20% noise conditions, respectively).

Finally we compared the performance of iDBM to classical DBM. To generate a BIC for the best-fitting iDBM we used the switch trials identified by iDBM to separate the data into blocks during which the iDBM analysis suggested that the simulated participant was using a single strategy. Next, we fit the DBM corresponding to that identified strategy to the data in that block. Finally, the likelihood of each block was summed and used to calculate the overall BIC score (Eq. 1). The number of free parameters was the sum of the number of free parameters of the best-fitting model in each block, plus the number of detected switches.

The mean BIC for each noise condition is shown in Table 2. For reference, the BIC score for the classical DBM

³Noise conditions of 60 and 80% were not further analyzed given their failure to accurately identify switch trials.

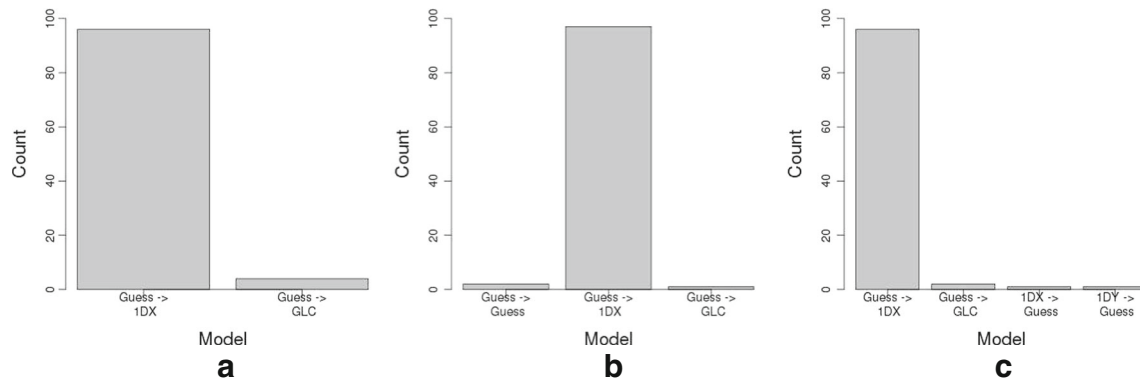


Fig. 5 First strategy switch identified for each noise condition. **a** Noise = 0 %, **b** Noise = 20 %, **c** Noise = 40 %. The noise conditions of 60 and 80 % were omitted due to their failure to accurately identify switch trials

that best fit the entire data set is also shown. This is the score that would result if current DBM practice was followed. As can be seen, the BIC score achieved with iDBM is smaller (better) than the one obtained with classical DBM for noise conditions 0, 20, and 40 %. The advantage of iDBM over DBM increases as noise decreases, and the two methods perform about equally well with noise levels of 60 % and 80 % (when the best-fitting DBM model is the guessing model). This suggests that there may not be much structure left in the data at such high noise levels, which would explain the diminished ability of iDBM to identify switches.

Discussion

The simulation results confirm that iDBM is capable of detecting and identifying both the strategy switch and the trial number when this switch occurs for low to moderate noise conditions. In all but the highest levels of noise (60 and 80 %), the switch from a guessing strategy to a one-dimensional rule was correctly identified in at least 96 % of the cases. Furthermore, the identified switch trial number was close to perfect when noise levels were 40 % or

lower, and over-estimated when the noise level was higher (Fig. 4b). The second switch, from 1DX to GLC was trickier to detect. It was consistently detected at the correct trial number for noise levels of 20 % or less, but less accurately detected at higher noise levels. Finally, the overall fit to the data of iDBM is better than DBM with noise levels of 40 % or lower, and the advantage of iDBM over regular DBM increases as the noise level decreases. iDBM and DBM fit the data about equally well in the highest noise conditions, when there is little structure left in the data.

One reason the switch from 1DX to GLC is difficult to detect at high noise levels is that, as described above, the noise was added to the simulated data in a way that is incompatible with all DBMs, and therefore with all versions of the iDBM. Thus, the more noise we add, the worse all DBMs will fit. The BIC goodness-of-fit measure penalizes models for extra parameters. The 1DX model only has two free parameters and therefore incurs only a small BIC penalty. In contrast, the model that assumes a switch from 1DX to GLC has six free parameters and therefore incurs a much larger penalty. High noise levels almost guarantee that the absolute fit of the switching model can only be modestly

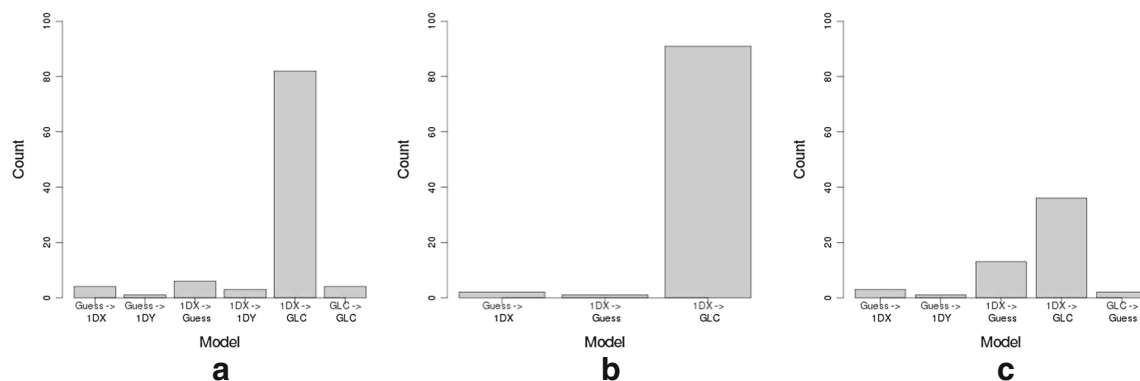


Fig. 6 Last strategy switch identified (when at least two switches were detected) for each noise condition. **a** Noise = 0 %, **b** Noise = 20 %, and **c** Noise = 40 %. Noise conditions 60 and 80 % were omitted due to the iDBM failure to accurately identify switch trials

Table 2 Mean BIC for each noise condition with an optimal responder

Noise level	iDBM	DBM
0 %	226.70	511.35 (GLC)
20 %	552.03	656.35 (GLC)
40 %	756.95	768.94 (GLC)
60 %	833.34	831.78 (Guessing)
80 %	833.61	831.78 (Guessing)

better than the fit of the pure 1DX model, and in many cases, the improvement in fit of the switching model will not be enough to overcome its greater BIC penalty. In the next section, additional benchmark tests are run to assess the ability of iDBM to detect strategy switches of suboptimal responders without changing the stimulus distribution.

Testing iDBM by using a suboptimal responder without changing the stimulus distributions

In the previous set of simulations, the stimulus distributions were changed on trials 101 and 351 and the optimal responder was assumed to change strategy to optimally respond to the new stimulus distributions. However, in most experiments, the stimulus distributions do not change, and only the participant's response strategy changes (which can be suboptimal). To remove the possible confound of changing the stimulus distributions, another set of benchmark simulations was run. In this second set of simulations, the stimulus distributions were II (Fig. 3, bottom) for all 600 trials. However, the simulated participants used a guessing strategy from trials 1 to 100, a 1DX strategy from trials 101 to 350, and a GLC strategy from trials 351 to 600. Note that an optimal responder should have used a GLC response strategy throughout. The same simulation methodology and parameters were used as in the previous set of benchmark simulations. However, only noise levels of 0, 20, and 40 %

were included, since the first set of simulations suggest that there is so little structure in the data when the noise levels are 60 or 80 % that all models perform poorly.

Simulation results

The optimal model fitting results with the simulated data would be to detect two strategy switches: (1) a switch from guessing to 1DX on trial 101 and (2) a switch from 1DX to GLC on trial 351. Figure 7 shows the mean number of switches and the trial number of the first and last switches identified by iDBM for each condition. As can be seen, the correct number of switches was identified fairly accurately, with a mean number of switches of 2.06, 1.84, and 1.76 for noise conditions 0, 20, and 40 %, respectively. At least one strategy switch was detected in all the simulations. This shows that iDBM again underestimated strategy switching in noisy conditions, rather than overestimated, which further supports that with these parameter settings, iDBM is conservative in detecting strategy switches.

Figure 7b shows the mean estimated trial number of the first strategy switch for all conditions. As can be seen, the mean estimated trial number of the first switch is accurately estimated for all noise conditions (100.05, 103.29, and 124.93, for noise conditions 0, 20, and 40 %, respectively). This suggests that the first switch (i.e., from guessing to 1DX) is accurately detected for the 40 % noise condition (although the switch trial is slightly over-estimated), and that the switch that is sometimes not detected is the second switch (i.e., from 1DX to GLC).

Figure 7c shows the mean estimated trial number of the last strategy switch for all conditions when more than 1 switch was detected, which occurred in 99, 83, and 45 simulations for noise conditions 0, 20, and 40 %, respectively. The trial of the last switch was accurately estimated in the 0 % and 20 % noise conditions (355.58 and 345.69, respectively), and under-estimated in the 40 % noise condition

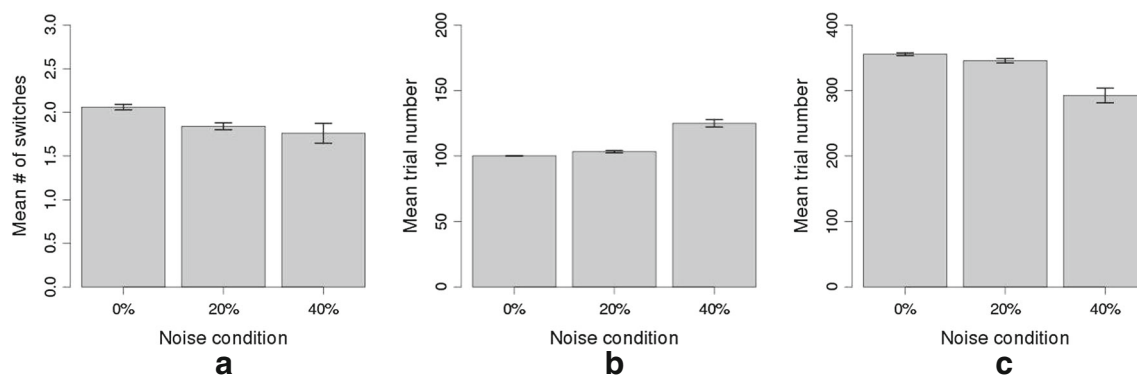


Fig. 7 Mean number of estimated switches for each noise condition. **a** Mean number of switches detected. True number of switches was 2. **b** Estimated trial number of the first switch. True switch trial was

trial 101. **c** Estimated trial number of the last switch when more than 1 switch was detected. True switch trial was trial 351. In all panels, *error bars* correspond to standard errors of the means

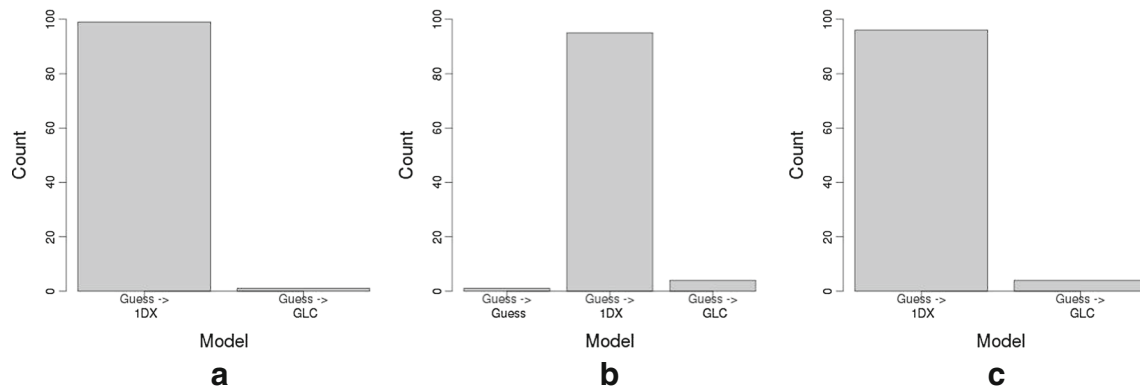


Fig. 8 First strategy switch identified for each noise condition. **a** Noise = 0 %, **b** Noise = 20 %, **c** Noise = 40 %

(292.56). This result is consistent with the first set of simulations and suggests that the switch from 1DX to the GLC is difficult to detect in high noise conditions and is often missed.

As a reminder, the first identified switch should be Guessing to 1DX and the last identified switch should be 1DX to GLC. The first identified switches are shown in Fig. 8. As can be seen, the correct first strategy switch was identified in at least 95 % of the simulations for all noise conditions. Similar to the first set of simulations, incorrect identifications were typically identified as Guessing to GLC.

The last identified switch models (when at least two switches were detected) are shown in Fig. 9. The correct 1DX to GLC switch was identified in most cases for noise conditions 0 and 20 % (91 and 95 %, respectively) but not for a noise level of 40 %. For a noise level of 40 %, only 2.2 % of the simulations identified the correct switch. Most identified switches (89 %) were erroneously labeled as 1DX to guessing.

Finally, the mean BIC for each noise condition is shown in the first two columns of Table 3 (1DX \rightarrow GLC). As can be seen, the BIC score achieved with iDBM is smaller (better) than the one obtained with classical DBM for noise

values of 0 % and 20 %, and the advantage of iDBM over DBM increases as noise decreases. However, DBM has a slightly better fit for a noise level of 40 %. In this condition almost half the trials are random guesses, which causes all models to fit poorly and therefore BIC favors models with fewer parameters. This also explains the diminished ability of iDBM to identify switches in this condition.

Parameter space analysis

As mentioned earlier, the parameters of iDBM were set for an arbitrary 100×100 stimulus space using previous experience in our lab. In order to explore how this parameter setting affects the behavior of the model, we ran a parameter space analysis by individually doubling and halving the values of parameters n (minimum fit window size), s (minimum number of consistent consecutive model identifications), and t (variance threshold on estimated η). One hundred simulations were run for each parameter value with the same II category structures and strategy switches used above with a noise level of 20 %. The resulting BIC values are shown in Table 4. As can be seen, the window size had a negligible effect on how well iDBM fits the simulation

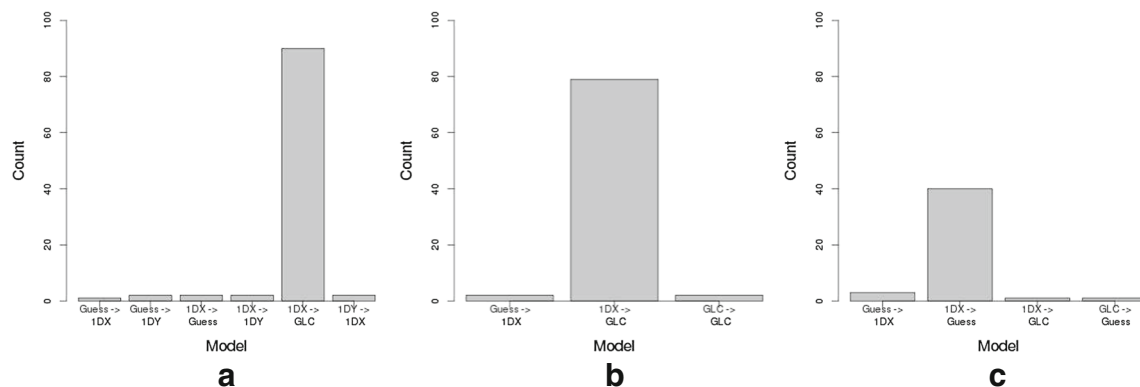


Fig. 9 Last strategy switch identified (when at least two switches were detected) for each noise condition. **a** Noise = 0 %, **b** Noise = 20 %, and **c** Noise = 40 %

Table 3 Mean BIC for each noise condition with a suboptimal responder

Noise level	1DX → GLC		1DX → 1DY		GLC → GLC	
	iDBM	DBM	iDBM	DBM	iDBM	DBM
0 %	222.1	532.8	–	–	–	–
20 %	573.0	656.1	954.1	831.8	617.3	773.1
40 %	790.7	772.0	–	–	–	–

data, and the number of consistent consecutive model identifications and variance on estimated η had a small effect. Note however that in all cases the BIC for the parameter settings used in this article is lower than those in the parameter space analysis (BIC = 573.0). The selection of parameter values is thus reasonable and we therefore recommend that future work uses these same parameter values after re-scaling the stimulus space into an arbitrary 100×100 coordinate system.

Additional simulations

The previous simulations assessed iDBM's ability to identify switches between strategies of different types, but participants sometimes switch between strategies of the same type (e.g., from 1DX to 1DY). To explore this possibility, additional simulations were run with the same Π distributions as above (Fig. 3c) for all 600 trials. There were two conditions: (1) The simulated participants used a guessing strategy from trials 1 to 100, a 1DX strategy from trials 101 to 350, and a 1DY strategy from trials 351 to 600, and (2) The simulated participants used a guessing strategy from trials 1 to 100, a GLC strategy with a positive slope from trials 101 to 350 (which is optimal), and a GLC strategy with a negative slope from trials 351 to 600. One hundred simulations were run for each condition with a noise level of 20 %.

iDBM detected 1.33 and 2.21 switches in conditions 1 and 2, respectively. In both cases, the first switch (away from guessing) was easily detected on mean trial 105.8 and 103.6 (for each condition, respectively). The second switch was always detected in Condition 2 [GLC(pos) → GLC(neg)], but only detected in 19 % of the simulations

in Condition 1 (1DX → 1DY). When the second switch was detected, the trial number was accurately estimated, with mean switch trials of 350.4 and 358.6, respectively. In Condition 1, the correct first switch was correctly identified in 99 % of the cases, and in Condition 2 the correct first switch correctly identified in 79 % of the cases. For the second switch, the correct switch was identified in 68 % of the simulations in Condition 1 (when a second switch was detected), and 62 % of the simulations in Condition 2. The BIC for these additional simulations is shown in Table 3.

Discussion

The simulation results obtained with suboptimal responders are consistent with those obtained with optimal responders. iDBM can accurately detect switches from guessing to 1DX strategies, and can detect changes from 1DX strategies to GLC under moderate or low levels of noise. The main difference is that suboptimal responders are slightly more sensitive to noise, so iDBM does not perform as well as DBM when 40 % of the trials are random guesses.

The parameter space was also explored by doubling and halving the values of the parameters corresponding to the minimum fit window size, number of consistent consecutive model identifications, and threshold on the variance of estimated η . The results show that the window size has a negligible effect on iDBM's fit to the data, and that the other two parameters had a small effect. These simulations also confirmed that the parameter values used in this article were reasonable and outperformed the other tested values.

Finally, a last set of simulations was run in which simulated participants were switching between two rule-based strategies or two procedural strategies. The results showed that the trial numbers of the switches were accurately estimated in both sets of simulations and that the correct switches were identified, but that the switch between 1DX and 1DY often went undetected, which made the fit of iDBM worse than DBM in this condition. This result is likely caused by the random noise in the data, since re-running the simulation without noise yields a BIC of 204.9 for iDBM (compared to 749.1 for DBM). Because of the noise, many stimuli are classified in the same category with 1DX and 1DY, which makes that switch difficult to

Table 4 Mean BIC in the parameter space analysis

Parameter	Half	Double
n	615.5	611.6
s	604.8	625.1
t	624.6	604.3

Note. The parameter space analysis was run with a noise level of 20 %. For comparison, the iDBM BIC with the regular parameter settings was 573.0 (see Table 3)

detect. In the next section, iDBM is used to re-analyze (real) empirical data reported by Ell and Ashby (2006).

Revisiting the data of Ell and Ashby (2006)

Ell and Ashby (2006) conducted three experiments to assess the effects of category overlap on strategy selection. Specifically, they were interested in whether high levels of category overlap might encourage participants to use one-dimensional rules in an II categorization task, rather than the optimal procedural strategy. The five conditions of their first two experiments are illustrated in Fig. 10. Note that the optimal strategy is identical in all conditions, and the conditions differ only in the amount of category overlap. In the Low Overlap condition (top left panel), a one-dimensional rule and the optimal procedural strategy both achieve perfect accuracy. In contrast, in the High Overlap condition (bottom left) the best one-dimensional rule achieves an accuracy of only 58 % correct, whereas the optimal procedural strategy yields an accuracy of 70 % correct.

Following COVIS (Ashby et al., 1998), Ell and Ashby hypothesized that human participants are biased towards

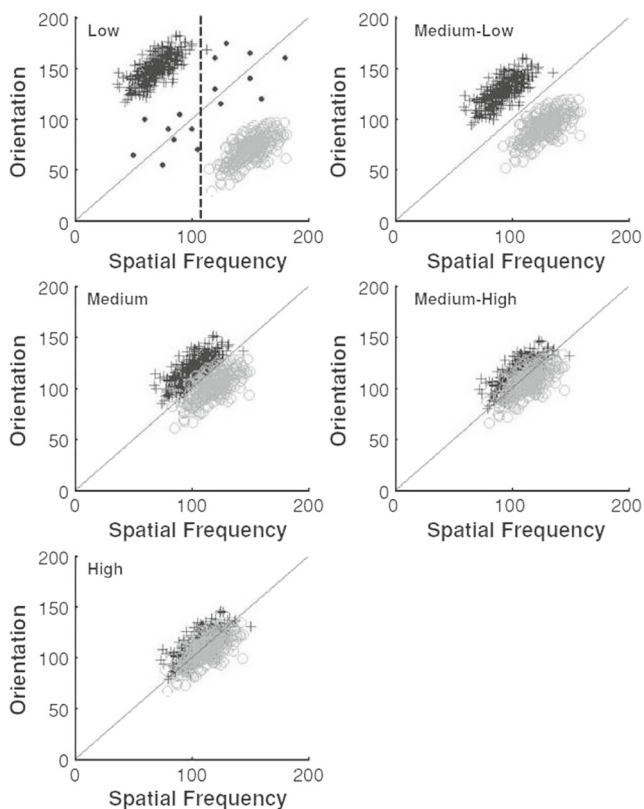


Fig. 10 II categories used in Experiments 1 and 2 of Ell and Ashby (2006). In the Low Overlap condition, the stimuli denoted by the *filled circles* were only presented (without feedback) at the end of training during a test phase. Reproduced with permission (c) Springer

using explicit rules, but that the larger the accuracy advantage for the procedural strategy over a one-dimensional rule, the more likely that participants will abandon rules in favor of a procedural strategy. In Fig. 10, the accuracy advantage of the optimal procedural strategy over the best one-dimensional rule is 0, 8, 21, 16, and 12 %, in the Low to High Overlap conditions, respectively. Thus, Ell and Ashby predicted that the number of participants using a one-dimensional rule in each condition would be ordered as follows: Low > Medium–Low > High > Medium–High > Medium, and also that the number of participants using a procedural strategy should be ordered in the opposite direction. The results of Experiment 1 closely matched these predictions, with an observed ordering of Low = High > Medium–Low = Medium–High > Medium. Note that the only discrepancy occurred in the High Overlap condition. However, performance was quite poor in this condition, and Ell and Ashby (2006) did not fit guessing models. So what looked like rule-use might actually have been guessing. Experiment 2 was a replication of the High Overlap condition from Experiment 1, but with longer training and monetary compensation for added motivation. The results were similar to those obtained in Experiment 1.

In Experiment 3, the stimuli from the Medium–Low, Medium, Medium–High, and High Overlap conditions of Experiment 1 were rotated in stimulus space 45° counter-clockwise. Note that this turns the optimal strategy into a one-dimensional rule, and also means that the previous optimal procedural accuracy is now the accuracy of the best one-dimensional rule (i.e., 100, 96, 78, and 70 % correct, respectively). Importantly, adopting a procedural strategy does not provide an advantage in any of these conditions, so participants should use a one-dimensional rule in all conditions. Figure 11 shows the resulting stimulus distributions. As expected, over 80 % of the participants adopted a rule in all conditions of Experiment 3 (Ell & Ashby, 2006).

New analyses with iDBM

While the Ell and Ashby (2006) data were a useful first test of the hypothesis that the probability of using a procedural strategy depends on the magnitude of its accuracy advantage over the best one-dimensional rule, the interpretation of the Ell and Ashby results were limited by their modeling procedures. Those procedures allowed a count of the number of participants whose responses were better described by a procedural strategy or a one-dimensional rule at the end of training, but they were incapable of determining how many strategy switches were made by each participant or of estimating when the last switch occurred. A stronger version of the Ell and Ashby (2006) hypothesis would be that the switch trial from a one-dimensional rule to a procedural strategy will occur sooner when the accuracy advantage

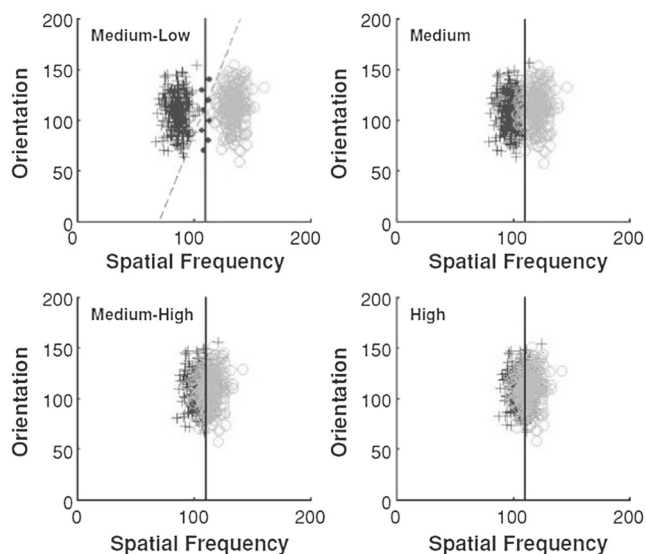


Fig. 11 RB categories used in Experiment 3 of Eil and Ashby (2006). In the Medium–Low Overlap condition, the stimuli denoted by the filled circles were only presented (without feedback) at the end of training during a test phase. Reproduced with permission (c) Springer

of the procedural strategy over the rule is increased. iDBM allows for a test of this hypothesis. In addition, guessing models were not fit to the data in the original article, and since guessing seems likely—especially when there is high overlap between the categories and optimal accuracy is low—guessing models were included in the new analyses.

The Eil et Ashby data were re-scaled in the 100×100 space and the algorithm from Table 1 was applied to the data from all three experiments using the same parameter settings as for the simulated data. The results for Experiment 1 were as follows: (1) For the Low Overlap condition, all participants except one began by guessing, and then switched to a one-dimensional rule. The exception participant began with a procedural strategy but then switched to a one-dimensional rule. The mean trial number on which participants switched to a one-dimensional rule was 23. (2) For the Medium–Low Overlap condition, all participants except one began by alternating between guessing and RB strategies but eventually switched to the optimal procedural strategy. The mean switch trial to adopt the procedural strategy was 274. The exception participant used a procedural strategy throughout the experiment. (3) In the Medium condition, 3 of the 5 participants used the correct procedural strategy throughout the experiment and stayed with that strategy for the duration of the experiment. The other two participants unstably switched among one-dimensional rules, guessing, and a procedural strategy during the first training session (600 trials), but stably switched to the correct procedural strategy on the first trial of Session 2 (trial 601). (4) In the Medium-High condition, four of the five participants began by unstably switching between a

one-dimensional rule and a guessing strategy but eventually switched to the correct procedural strategy on mean trial 670 (except for short periods of guessing at the beginning of each session). The other participant kept switching among all three strategies (i.e., one-dimensional rule, guessing, procedural) and never settled on a stable strategy. (5) In the High Overlap condition, one participant successfully switched to the correct procedural strategy on trial 417 of Session 2 (trial 1017). The other four participants never settled on any one strategy.

The new iDBM analyses showed the following ordering for mean switch trial numbers: Medium (244) < Medium–Low (274) < Medium–High (670) (see Fig. 12). These results are consistent with a stronger version of the Eil and Ashby (2006) hypothesis that the Medium Overlap condition should lead to the earliest switching to the correct procedural strategy. However, switches in the Medium–Low condition happened earlier (and more frequently) than in the Medium–High condition. This difference between the new iDBM results and the previous results may be caused by the inclusion of guessing models. Tasks that are objectively more difficult (such as the Medium–High condition) may discourage some participants and lead to more guessing. This could prevent participants from stably switching to a procedural strategy.

Another interesting result of the new analysis concerns the High and Low Overlap conditions. As in the original Eil and Ashby (2006) analysis, iDBM concluded that almost no participants selected a procedural strategy in either of these conditions. However, iDBM suggests that this happened for different reasons in the two conditions. In the Low Overlap condition, iDBM concludes that participants settled fairly quickly on a one-dimensional rule, which is sufficient for perfect accuracy. This is predicted by COVIS (Ashby et al., 1998), which assumes that humans are biased toward using rules. In contrast, iDBM concludes that participants in the High Overlap condition never settled on a procedural strategy because the task was too difficult. Instead, they

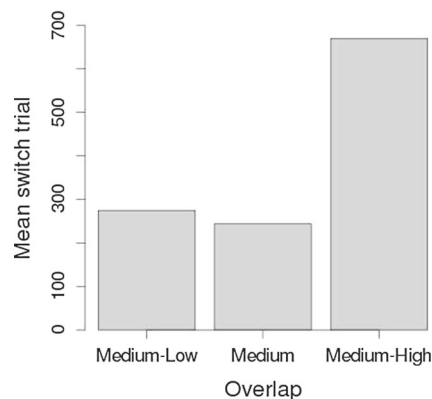


Fig. 12 Mean switch trial to the correct strategy in Eil and Ashby (2006)'s Experiment 1 as identified by the iDBM

oscillated among one-dimensional rules, procedural strategies, and random guessing. This is a more extreme version of what happened in the Medium–High condition, where many participants became discouraged by the difficulty of the task. This interpretation is further supported by the results of Experiment 2 (a replication of the High Overlap condition from Experiment 1), in which all but one participant eventually switched to the correct procedural strategy (on mean trial 1,055). Only one participant never settled on any one strategy (compared with four in the High Overlap condition of Experiment 1). Participants in the High Overlap condition of Experiment 1 had more than 2000 trials of training, so the fact that participants could select the correct procedural strategy in Experiment 2, but not in the High Overlap condition of Experiment 1 suggests that the added monetary incentive in Experiment 2 may have increased the participants’ motivation enough for them to identify the correct procedural strategy.

The data from Experiment 3 were also fit. Similar to Ell and Ashby (2006), we found that most participants in all conditions used the correct rule-based strategy. More specifically: (1) For the Medium Overlap condition, all but one participant used the correct one-dimensional rule for the entire duration of the experiment. The exception began by guessing before correctly selecting a one-dimensional rule on trial 219. However, this participant later abandoned the rule strategy during the third training session to use instead a procedural strategy on trial 1,444. This was probably caused by fatigue. (2) In the Medium–Low Overlap condition, 6 of the 8 participants used the correct rule strategy for the duration of the experiment. The other 2 participants began by alternating between guessing and procedural strategies but switched to the correct rule strategy on mean trial 72. (3) In the Medium–High Overlap condition, four of the five participants used the correct rule strategy from trial 1. However, one of these four participants responses were sometimes unstable, so there do appear to be some short episodes of guessing interspersed throughout the training. The other participant started by guessing but selected the correct rule strategy on trial 19. (4) In the High Overlap condition, one participant used the correct rule strategy for

the duration of the experiment. The other four participants started with guessing and procedural strategies, were sometimes unstable, but all ended up selecting the correct rule strategy on mean trial 1,351. These results are consistent with the previous analysis of Ell and Ashby (2006) and with the predictions of COVIS (Ashby et al., 1998) that humans are biased towards using rules. Our results also suggest that errors entice participants to switch strategies (Kalish et al., 2005), so higher overlap increases the number of trials needed to settle on a stable strategy, even when the optimal strategy is a one-dimensional rule.

Finally, Table 5 shows the mean BIC estimated with iDBM and classical DBM for each condition in each experiment of Ell and Ashby (2006). Experiments 1 and 2 used II stimuli, so the GLC is the optimal strategy. As can be seen, the results obtained are similar to those obtained with the simulated data (Table 2). For low overlap conditions (Low and Medium–Low), iDBM provides a better fit than classical DBM. For higher overlap conditions, the fits are similar for iDBM and classical DBM, with a small advantage for classical DBM due to the smaller number of free parameters. Higher overlap produced more errors, which corresponds to higher noise levels in the simulated data. Experiment 3 used RB categories, and the optimal strategy was IDX. Again, the results show that the advantage of iDBM over classical DBM increases as overlap (and errors) decreases. However, in this case iDBM fits the data better than classical DBM for every level of overlap.

General discussion

This article introduced a new modeling approach, called iDBM, that iteratively fits decision-bound models to category-learning data in order to identify: (1) all response strategies used by a participant, (2) changes in response strategy and, (3) the trial number at which each change occurs. Unlike classical DBM, the new approach does not assume that the participant uses the same decision strategy on every trial, and therefore does not require that the data be arbitrarily separated into blocks before model

Table 5 Mean BIC for each experiment and condition in Ell and Ashby (2006)

Overlap	Exp. 1 (II)		Exp. 2 (II)		Exp. 3 (RB)	
	iDBM	DBM	iDBM	DBM	iDBM	DBM
High	3,093.0	3,073.3	2,735.9	2,705.7	2,372.6	2,390.8
Medium-High	2,548.4	2,503.0	–	–	1,654.1	1,714.7
Medium	1,361.4	1,337.1	–	–	970.2	1,011.7
Medium-Low	561.9	586.4	–	–	176.4	221.3
Low	175.9	185.7	–	–	–	–

fitting. The new method was validated by testing its ability to identify the response strategies used in noisy simulated data. The benchmark simulation results show that iDBM is able to detect and identify strategy switches during an experiment and accurately estimate the trial number at which the strategy change occurs in low to moderate noise conditions. Strategy changes in high noise conditions were more difficult to detect, but when detected they were typically correctly identified. Importantly, iDBM rarely if ever produced false alarms. When the data are too noisy, strategy switches are missed rather than incorrectly detected. The benchmark simulations also show that the fit advantage of iDBM over classical DBM increases as the noise in the data decreases.

The new method was then used to re-analyze data reported by Ell and Ashby (2006). This new analysis allowed much stronger conclusions to be drawn about how category overlap affects strategy selection. In particular, the results of iDBM suggest that increasing the accuracy advantage of procedural strategies over one-dimensional rules not only increased the proportion of participants abandoning rules, but also reduced the number of training trials needed to abandon rules in favor of a procedural strategy. This result is in agreement with previous research (Spiering & Ashby, 2008). The analysis further suggests that when both the rule-based and procedural strategies yield low categorization accuracy, participants tend to get discouraged and either guess or oscillate among a number of response strategies without ever settling on a stable strategy. These conclusions are consistent with previous work on strategy shifts (e.g., Kalish et al., 2005), and were absent from Ell and Ashby's (2006) original article because earlier available methods did not allow such detailed hypotheses to be investigated.

Generality of the method and limitations

The applications considered in this article were all to experiments that used the randomization technique (i.e., to experiments where the stimuli in each category were random samples from some underlying probability distribution). However, iDBM could be applied to any type of categorization experiment where DBMs can be fit and tested, and therefore is not restricted to randomization experiments (for some non-randomization examples, see e.g., Ashby & Lee, 1992). The only extra requirement is that trial-by-trial responses are available.

iDBM was developed as a straightforward extension of DBM. Even so, the iterative algorithm described in Table 1 is general enough that it can be used with any models that can be fit using maximum likelihood and the BIC goodness-of-fit statistic. Table 1 refers to a set of basic models, but nothing in the algorithm assumes these are DBMs. So the algorithm should work for any set of basic models (e.g., exemplar models or prototype models). The switch models

would then be defined by all possible pairs of basic models, along with an η parameter specifying the switch trial. After that, the fitting procedures would be identical to those described here. The inclusion of non-DBM models should be explored in future work.

The choice of which models to include in the set of basic models is likely to significantly affect the fitting procedure. For example, Donkin et al. (2015) recently showed that adding more models when using DBM can change the interpretation of the results. It is important to keep in mind that the conclusions reached depend on which models belong to the set of basic models, and that it is possible that some other untested model might fit better than any of the basic models. Hence, the strategy identified by iDBM is the best among the tested alternatives, but one should refrain from concluding that the detected strategy is the "true" strategy.

Another important limitation of the proposed method is that iDBM does not address the possibility of small improvements or tuning of parameters within a strategy. With iDBM, small changes in the way a strategy is applied (e.g., a small change in the rule criterion) will likely go undetected, while larger changes will be detected as changes in strategy. We are not claiming that these adjustments in the way strategies are applied are not occurring. This simplification is a consequence of iDBM's focus on strategy shifts, and strategy tweaking is outside the scope of the proposed method. However, a future version of iDBM may attempt to address both continuous learning within a strategy and abrupt changes in strategy simultaneously.

Lastly, like all iterative fitting procedures, iDBM can get trapped in local minima. Hence, it is possible that on any given iteration, the selected switch model is not the true best-fitting model included in the set of switch models. However, the iterative nature of iDBM may be helpful in avoiding this pitfall because the same switch model needs to be consistently selected for a number of consecutive iterations, with a small variance on estimated η , before a switch is identified. Because the starting point of the fitting procedure is varied, it is less likely that iDBM would systematically fall in the same local minimum for several trials in a row. Hence, while the risk of local minima is not eliminated, it is unlikely to be a serious problem when using iDBM.

Future work

This article shows how iDBM could be used to explore the effects of category overlap on strategy selection, but many other research questions can also be explored using this new approach. For example, as mentioned earlier, COVIS predicts that in II tasks, participants will initially either guess, or experiment with simple explicit rules before eventually adopting a procedural strategy. Although this prediction is

nearly 20 years old, it has never been tested. Our re-analysis of the Ell and Ashby (2006) data suggests the prediction may have some validity, but obviously, a much more extensive test is needed. Fortunately, iDBM makes such a test possible.

Another issue that iDBM opens up for possible study is the frequency of strategy switching. All current multiple systems models of category learning, including COVIS and ATRIUM (Erickson & Kruschke, 1998), predict trial-by-trial switching. If trial-by-trial switching did occur, then iDBM would surely underestimate the switching frequency. This is because iDBM is conservative with respect to identifying a strategy switch. Even so, iDBM should be able to answer the question of whether strategy switches are rare or common. Furthermore, iDBM could be used to compare the frequency of within- versus between-system switches. For example, is it more common for participants to switch between two explicit strategies (e.g., from 1DX to 1DY) than to switch from an explicit to a procedural strategy (e.g., from 1DX to GLC)? If one-dimensional rules and procedural strategies are mediated by different neural systems, then it seems plausible that switching between two neural systems should be mediated differently than switching between two strategies that are both mediated by the same neural system. Previously, these issues could not be studied because there was no method for identifying a strategy switch. We developed iDBM to fill this void, and hope that the possibility of identifying strategy switches may usher in a new era of insight into classification performance.

Acknowledgments This research was supported in part by NIMH grant #2R01MH063760-09A1 to SH and FGA, and NSF grants #1349677-BCS and #1349737 to SH and SWE (respectively). Correspondence concerning this article should be addressed to Sébastien Hélie, Department of Psychological Sciences, Purdue University (e-mail: shelie@purdue.edu).

References

- Ashby, F.G. (1992). Multivariate probability distributions. In Ashby, F. (Ed.) *Multidimensional Models of Perception and Cognition*. Hillsdale, NJ: Erlbaum.
- Ashby, F.G., Alfonso-Reese, L.A., Turken, A.U., & Waldron, E.M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*(3), 442–481.
- Ashby, F.G., & Crossley, M.J. (2010). Interactions between declarative and procedural-learning categorization systems. *Neurobiology of Learning and Memory*, *94*(1), 1–12.
- Ashby, F.G., & Gott, R.E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33–53.
- Ashby, F.G., & Lee, W.W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, *120*(2), 150–172.
- Ashby, F.G., & Lee, W.W. (1992). On the relationship among identification, similarity, and categorization: Reply to Nosofsky and Smith (1992). *Journal of Experimental Psychology: General*, *121*(3), 385–393.
- Ashby, F.G., & Maddox, W.T. (2005). Human category learning. *Annual Review of Psychology*, *56*, 149–178.
- Ashby, F.G., & Soto, F.A. (2015). Multidimensional signal detection theory. In Busemeyer, J., Townsend, J., Wang, Z., & Eidels, A. (Eds.) *Oxford Handbook of Computational and Mathematical Psychology*. New York: Oxford University Press.
- Ashby, F.G., & Townsend, J.T. (1986). Varieties of perceptual independence. *Psychological Review*, *93*, 154–179.
- Ashby, F.G., & Valentin, V.V. (2016). Multiple systems of perceptual category learning: theory and cognitive tests. In Cohen, H., & Lefebvre, C. (Eds.) *Handbook of Categorization in Cognitive Science*. 2nd edn., (pp. 547–572). New York: Elsevier.
- Donkin, C., Newell, B.R., Kalish, M., Dunn, J.C., Nosofsky, R.M., Donkin, C., & et al. (2015). Identifying strategy use in category learning tasks: a case for more diagnostic data and models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 933–948.
- Ell, S.W., & Ashby, F.G. (2006). The effects of category overlap on information-integration and rule-based category learning. *Perception and Psychophysics*, *68*(6), 1013–1026.
- Erickson, M.A. (2008). Executive attention and task switching in category learning: evidence for stimulus-dependent representation. *Memory and Cognition*, *36*(4), 749–761.
- Erickson, M.A., & Kruschke, J.K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General*, *127*(2), 107–140.
- Haider, H., & Frensch, P.A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*, 304–337.
- Hélie, S. (2006). An introduction to model selection. *Tutorials in Quantitative Methods for Psychology*, *2*, 1–10.
- Kalish, M.L., Lewandowsky, S., & Davies, M. (2005). Error-driven knowledge restructuring in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 846–861.
- Logan, G.D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527.
- Maddox, W.T., & Ashby, F.G. (1993). Comparing decision bound and exemplar models of categorization. *Perception and Psychophysics*, *53*(1), 49–70.
- Maddox, W.T., Ashby, F.G., & Waldron, E.M. (2002). Multiple attention systems in perceptual categorization. *Memory and Cognition*, *30*(3), 325–339.
- Neapolitan, R. (2004). *Learning Bayesian Networks*. Upper Saddle River, NJ: Prentice Hall.
- Spiering, B.J., & Ashby, F.G. (2008). Initial training with difficult items facilitates information integration, but not rule-based category learning. *Psychological Science*, *19*(11), 1169–1177.