Johansen, Mark K. , Savage, Justin , Fouquet, Nathalie and Shanks, David R. 2015. Salience not status: how category labels influence feature inference. Cognitive Science 39 (7) , pp. 1594-1621. 10.1111/cogs.12206

This is the post-accepted version of this manuscript in press with *Cognitive Science*

23 July, 2014

Salience not status: How category labels influence feature inference

Mark K. Johansen

Cardiff University

Justin Savage

Cardiff University

Nathalie Fouquet

Swansea University

David R. Shanks

University College London

Correspondence: Mark Johansen, JohansenM@cardiff.ac.uk, PHONE: +44 (0)29 2087 0140, FAX: +44 (0)29 2087 4858, School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, United Kingdom CF10 3AT

Abstract

Two main uses of categories are classification and feature inference, and category labels have been widely shown to play a dominant role in feature inference. However, the nature of this influence is unclear, and we evaluate two contrasting hypotheses formalized as mathematical models: the label special-mechanism hypothesis and the label super-salience hypothesis. The special-mechanism hypothesis is that category labels, unlike other features, trigger inference decision making in reference to the category prototypes. This results in a tendency for prototype-compatible inferences because the labels trigger a special mechanism rather than because of any influences they have on similarity evaluation. The super-salience hypothesis assumes that the large label influence is due to their high salience and corresponding impact on similarity without any need for a special mechanism. Application of the two models to a feature inference task based on a family resemblance category structure yields strong support for the label super-salience hypothesis and in particular does not support the need for a special mechanism based on prototypes.

Keywords: Concepts, Decision making, Classification, Feature Inference

Salience not status: How category labels influence feature inference

Arguably the most important aspect of categories is that they facilitate the predictive inference of hidden features. But how does category membership influence feature inference decision making? What does category membership information add? And in what way is this information added?

It is not surprising that category labels as indicators of category membership should receive a lot of attention in the context of feature inference decision making as they are clearly markers for a substantial amount of hidden information and knowledge about category instances. For example, in their classic experiments Gelman and Markman (1986) contrasted the influence of category labels versus perceptual similarity on feature inference. One case comprised presenting participants with two instances and a testing item. One instance was labelled as a bird, looked like a flamingo and was said to have the hidden feature of a right aortic arch, and the second was labelled a bat, looked like a bat and was said to have the hidden feature of a left aortic arch. The test case was labelled a bird and looked like a dove, so while it had the same category label as the flamingo instance (bird), it was perceptually more similar (in the picture) to the bat instance. Participants were then asked to infer whether the test case had a right aortic arch, like the matching category instance (bird), or a left aortic arch, like the more perceptually similar instance of the bat category. Gelman and Markman's main finding was that most adults and the majority of children inferred a hidden feature consistent with an instance's category label, thus matching a perceptually dissimilar but nonetheless identically labelled instance, even though there was another, perceptually more similar item but with a different category label.

This method of contrasting category membership information with similarity has been adapted, notably by Yamauchi and Markman (2000), to the more controlled stimuli in perceptual categorization paradigms, on which the present research is based. In overview, Yamauchi and Markman (2000) also found that category labels dominantly influenced feature inferences even when overall perceptual similarity indicated a different response. In more detail, they used a pure decision making task in which participants were given a category summary containing all of the category instances presented simultaneously (similar to Fig. 1). The key result was that participants made feature inferences consistent with the prototype corresponding to the category label even when the other features in the test case were from the prototype of the other category. Put differently, participants made inference decisions consistent with the central tendency of the category indicated by the category label even when simple similarity was higher to instances of the opposing category.

Yamauchi and Markman's (2000) main conclusion was that their results supported the hypothesis that feature inference decision making focuses on the central tendency of the category indicated by the instance's category label. Specifically, "… information about category membership molds the way people infer a characteristic [feature] of an object. When the category membership of an object is known, people pay particular attention to the feature value most prevalent in the members of the corresponding category [the prototype]" (p. 792). More generally they summarized their results to "indicate category membership is indeed a key determinant of inductive judgment and that category labels are not simply another feature on par with other category features" (p. 793).

Central to the argument that labels are not the same as other features is that even if labels are treated as simply features like other perceptual features they still are able to dominate multiple other features to drive feature inference. That is, one label "feature" still dominates multiple perceptual features. However from the perspective of simple similarity, this argument is based on the assumption that all the features influence similarity roughly equally. In the context of simple similarity, one feature might dominate others because it is truly different from the others in terms of invoking different cognitive mechanisms, or it might dominate others for the more basic reason that it is perceptually more salient and just has a larger influence on similarity than other features but only as a matter of degree. We will examine this seemingly subtle distinction between a special mechanism versus simple salience below in greater detail, but it is worth emphasizing in advance that it results in surprisingly large predicted differences *a priori* in people's behavior.

To reiterate, category labels should and do have a large influence in decision making consistent with their adaptive functionality, but the nature of this influence on decision making is not clear. Put more precisely, the critical questions are: Do category labels trigger systematic prototype-compatible feature inferences? Or are they just especially salient features that have a large effect on similarity assessment to whatever representation is available but no special tendency to produce integrative—i.e., prototype-compatible—decision making?

In our experiment, the instances from two categories were presented simultaneously (Fig. 1), and participants then inferred missing features for test cases (e.g., Fig. 3). This summary decision-making methodology was adapted from Yamauchi and Markman (2000) and Murphy and Ross (1994; 2005), but has been employed in the study of category-based induction (see Murphy, 2002, and Hayes, Heit and Swendsen, 2010, for broad summaries of category-based induction) and of course is widely used in decision making research.

Predictions of the Two Main Hypotheses Formalized as Models

The *label special-mechanism hypothesis* is in part derived from Yamauchi and Markman (2000). It assumes that the category label strongly influences feature inference because it indicates category membership, invokes the integration of the category prototype from the category instances in working memory during decision making, and so results in a strong tendency for prototype-compatible feature inferences. This should occur even if none of the stimulus features match that prototype's features.

We have formalized the label special-mechanism hypothesis (Appendix A) in terms of prototype representation (Blair & Homa, 2001; Homa, Rhoads, & Chambliss, 1979; Homa, Sterling & Trepel, 1981; Posner & Keele, 1968; Smith & Minda, 1998; 2000) as applied to feature inference decision making (related to the applications of prototype representation to category learning data in Yamauchi & Markman, 1998, and Johansen and Kruschke, 2005). For a feature inference test case, the model simply calculates the similarity of the probe to the prototype which has the same category label as the probe, relative to its similarity to the prototypes of all other categories, and then predicts the missing prototype-compatible feature to the extent that the probe is similar to that prototype. So the core aspect of this model is that the category label for the test case uniquely determines how the representation is accessed, unlike other features which merely influence similarity. Overall, this formalization corresponds to a strong tendency for prototype-compatible feature inferences, as shown on a variety of testing cases which we specify below.

The *label super-salience hypothesis,* on the other hand, is directly related to the similarity and attention based account of the role of category labels summarized in part in Sloutsky (2003) from a developmental perspective (see also Sloutsky and Fisher (2004; 2011)) and is related to the assessments of salience in Yamauchi and Markman (2000) and Deng and Sloutsky (2013). The super-salience hypothesis assumes that a physical category label, though a marker of category membership, can also be particularly salient compared to the other stimulus features. Hence the label may influence inference decision making relatively more than the other features due to its greater impact on similarity assessment, without directly determining how the representation is accessed.

We have formalized (see Appendix A) the label super-salience hypothesis in terms of an exemplar model (the generalized context model; Nosofsky, 1984; 1986) as applied to feature inference decision making (see also Johansen & Kruschke, 2005; Yamauchi & Markman, 1998). For a feature inference test case, the model calculates the similarity of the case to the category instances that predict one feature inference relative to similarity to instances that predict a contrasting feature inference, with a resulting response probability in proportion to those similarities. A significant difference from the special-mechanism model is that the category labels are treated just like the other features in determining similarity; their influence, like other features, is modulated by the amount of attention they receive, but they do not have any special-mechanism invoking property of determining the selected category to which similarity is computed. Critically, similarity assessment need not result in any strong tendency for prototype-compatible feature inferences. On the contrary, a prototype-incompatible feature inference may be made if the probe is similar to an exemplar with a prototype-*incompatible* feature.

An important difference between the models, which is a major focus of the experiment reported here, arises in situations where a feature inference probe is more similar to the prototype of one category (category A) than to the prototype of another category (B), but at the same time is very similar to an instance of category A which nonetheless has a target feature typical of the prototype from the other category (B). Let us denote the modal value of category A members for the missing feature as *a* and the modal value for category B members as *b*. Assuming that attention is evenly distributed across all feature dimensions as well as the category label, the super-salience model can predict that participants might infer the value *b* for the missing feature, because similarity to the particular category A instance strongly biases the similarity computation. The special-mechanism model, in contrast, will predict that the value *a* is inferred. It does this because the probe’s label invokes the category A prototype and therefore inherits the feature value attached to the prototype.

Before describing how the experiment contrasted the models for the special-mechanism and super-salience hypotheses, it is useful to emphasize the main process differences between the models as summarized schematically in Fig. 2. Both models generate the response probability of a particular feature inference (as indicated by “?”) for a given test item with a given category label and set of features as indicated by “P(feature 1|given LABEL A and feature x?xx)” at the center of the figure. The super-salience model, shown at the bottom of the figure, determines the probability of feature 1 via similarity to instances with feature 1 *in all categories*, so instance category labels influence similarity in the same way as other features. In contrast, the special-mechanism model, shown at the top of the figure, determines the probability of feature 1 via similarity to the prototype corresponding to the category label in the test item, Label A in this case**,** so the category label has the special property of determining which prototype similarity is calculated in relation to and hence is handled differently from other features.

To contrast the label special-mechanism and label super-salience hypotheses, feature inference test cases should differentiate category integrated, prototype based responding from unintegrated responding based on similarity to particular category instances. To do this we have used a family resemblance category structure as shown in Table 1. The 16 category instances corresponding to the 16 stimuli in Figure 1 are shown at the top of the table where each row indicates a particular category instance. Each instance is associated with a category label (A or B) as shown in the first column and has four feature values (1 or 2) one on each of four feature dimensions corresponding to the remaining four columns, one for each of the four perceptual features composing the stimuli in Figure 1. (See the note in Table 1 for this mapping.) The bottom of Table 1 shows that three kinds of test trials were used designated Label versus Feature, Exception and NonException trials.

Starting by direct analogy with Gelman and Markman’s (1986) classic contrast between category labels versus similarity, Label versus Feature trials are so called because they are composed of a category label from one category, but more features typical of the other category, together with a missing feature. The missing feature is then queried. Consequently, the missing feature value as predicted by the prototype corresponding to the category label is potentially different from the feature predicted by similarity to the category instances, much like in Gelman and Markman's study. For example in Table 1 the Label vs. Feature test case B 11?2 shares more features with the instances of Category A, which have many 1 features, than the instances of Category B, which have many 2 features, so instance similarity tends to predict a 1 feature response on the ? dimension. However, B 11?2 nonetheless has the label for category B, so the category label predicts a 2 feature response matching the category B prototype, B 2222. The top of Fig. 3 shows this example where, in particular, the two boosters feature corresponds to the prototype associated with the label (Planet B) and the one booster feature corresponds to instance similarity for the contrast category. Testing trials of this kind featured prominently in Yamauchi and Markman (2000) and are important here as well.

Exception trials had a response feature predicted by the nearest, most similar category instance that was different from the feature predicted by the category prototype corresponding to the instance's category label. For example, the nearest exemplar to the feature inference test case B ?222 in Table 1, that is the category instance B 1222, predicts a different feature than the category prototype, B 2222, features 1 and 2 respectively. The bottom of Fig. 3 show this example where, in particular, the wide wings feature corresponds to the prototype associated with the label (Planet B) and the narrow wings feature corresponds to the feature from the most similar instance. Yamauchi and Markman (2000) did not have testing trials of this type; however, the Exception trials are critical for allowing participants the possibility of exhibiting prototype-*incompatible* feature inferences even in the presence of a large emphasis on the category labels as shown in the Label versus Feature trials.

Lastly, for NonException trials, instance similarity and the category prototype corresponding to the category label predict the same feature. For example, the test case B ?122 has B 2122 as its nearest instance, which predicts the same missing feature, 2, as the category prototype, B 2222. In addition to providing important constraints on the modelling assessment, these trials provided a check that participants weren't simply responding randomly in the task.

We derive predictions from the special-mechanism and super-salience hypotheses for these testing trial types in two ways: First we generate predictions from the hypothesis definitions. Then we evaluate what the formalized versions of the hypotheses as models can predict *a priori* for this category structure and testing trials across a broad range of parameter values. Both the special-mechanism and super-salience hypotheses are consistent with strongly label-consistent feature inferences on Label versus Feature trials. However, the hypotheses predict this for different reasons that can be differentiated by the Exception trials.

The special mechanism hypothesis predicts label-consistent feature inferences because the label, unlike the other features, has the unique tendency to invoke prototype-compatible feature inferences. The special-mechanism hypothesis can predict differences in the tendency for these prototype-compatible inferences as a function of how much attention the instance label receives, but to the degree that the label receives a lot of attention then this translates into a correspondingly strong tendency to predict prototype-compatible feature inferences for the Exception trials. Intuitively, because the label dominates and forces reference to the prototypes, there is effectively no way to predict prototype-incompatible features because a given prototype does not include any.

Despite the apparent intuitive contrast between the hypotheses for this kind of testing trial, the super-salience hypothesis can also predict label-consistent feature inferences but as a result of the dominant influence of the labels in similarity assessment to the category instances, not the category prototypes. In addition, it also can predict differences in the tendency for label-consistent feature inferences as a function of how much attention the instance label receives. But critically the super-salience hypothesis is not constrained to predict prototype-compatible features on Exception trials but is also consistent with prototype-incompatible feature inferences. Intuitively, even if the label is dominant, the test case can be so similar to a particular category instance that a prototype-incompatible feature inference is predicted.

Model simulations

While the formal models were ultimately fitted to all the inference testing trial responses of each participant individually (as reported later), it is useful for visualization to combine test trials of a given type together, which we do both here and in the presentation of the various experimental data sets. That is, both the model predictions and the data were coded in terms of the proportion of label based prototype-compatible responses across the trials of a given type. So, for example, for the four Label versus Feature testing trials in Table 1, this proportion corresponds to the average proportion of prototype-compatible responses (either predicted by the model or observed in the data) consistent with the label present in each testing case across all four cases. Likewise, an average proportion of prototype-compatible responses can be generated across the four Exception trials in Table 1 (and also for the NonException trials). This method of analysis has the important advantage that different patterns of feature inference responses across the testing cases in Table 1 can be represented as different points in a scatter plot of proportions of prototype-compatible responses on Label versus Feature Trials against Exception trials.

To test *a priori* predictions from the models (as specified in Appendix A), we generated a large number of simulated participants. Each simulated participant was produced by selecting random values for each free parameter in the model from a reasonable range of possible parameter values. These free parameters included a parameter for the amount of label attention as well as a separate attention parameter for each of the four feature dimensions. For each set of random parameters, the model generated response proportions for each of the testing cases in Table 1 based on this category structure, and these were averaged together to produce response proportions for each type of testing trial—NonException, Label versus Feature, and Exception. The Label versus Feature and Exception proportions were then used to generate scatter plots in which each point represents a single simulated participant.

The results of 3000 simulated participants for the special-mechanism and super-salience models are shown in the top and bottom panels respectively of Fig. 4, where the *x*-axis is the average proportion of prototype-compatible responses on the Exception trials and the *y*-axis is the average proportion of prototype-compatible responses on the Label versus Feature trials. In addition the gray-scale value (and shape) of each marker corresponds to ranges of average response proportions on NonException trials (tabulated across the four NonException trials) as specified in the figure's key.

What is immediately obvious from the simulations illustrated in Fig. 4 is that the models predict radically different things in this data space. There is very little overlap in terms of what the models can predict except in the middle near guessing and near the top right corner which represents high prototype-compatible response proportions on Label versus Feature *and* Exception trials. This corresponds to situations where most of the attention is allocated to the category labels and where both models make essentially the same predictions.

In addition, the results for NonException trials, coded by symbol gray scale in Fig. 4, more subtly indicate a difference between the models. In the top panel the special-mechanism model predicts progressively higher average proportions on NonException trials (darker markers) as one moves farther to the right in the space, that is higher prototype-compatible responding on Exception trials. Put another way, the special-mechanism model appears to be constrained to predict response proportions near 0.5 on NonException trials to the degree that the average response proportions on Label versus Feature and Exception trials are near 0.5. This is not conceptually surprising in that the special-mechanism model’s ability to predict prototype-compatible responding on Label versus Feature and Exception trials constrains it to also predict prototype-compatible responding on NonException trials as this is based on the same prototypes. The super-salience model, on the other hand, is constrained in almost the opposite way in that low NonException proportions (near 0.5, indicated by lighter markers in the bottom panel) correspond to Label versus Feature and Exception proportions near the center of the space, but high proportions near 1 allow the model to get to a wider range of locations in the data space. Conceptually this makes sense in that the super-salience model is an exemplar model and hence its ability to predict proportions near 0.5 on NonException trials then tends to constrain Label versus Feature and Exception trials also to be near 0.5. So overall, NonException trial responding constrains the models in different ways.

Intuitively the critical aspects of the difference between the models in this data space arise from their different capacities to predict prototype-incompatible responses on Exception trials as a result of similarity to specific instances (left-right in this data space) while at the same time predicting prototype-compatible responses on Label versus Feature trials as a result of a lot of salience-driven attention to the category labels (up-down in the space). To specify the qualitative difference between the models in more detail consider the four quadrants of the data space: the special-mechanism model can account for a range of responding in the right two quadrants (top panel in Fig. 4) while the super-salience model can account for a range of responding in the top two quadrants and at least partly into the bottom left quadrant (bottom panel of Fig. 4).

The super-salience model can account for the full, left-right, range of responding on Exception trials by adjusting how strongly its responding is determined by similarity to a single nearest instance that predicts a prototype-incompatible feature versus similarity to many instances predicting prototype-compatible features. In contrast, the special-mechanism model does not have a systematic way to predict a tendency for prototype-incompatible responding on these trials, on the left of the space, because its responses are derived from the category prototypes and hence correspond to a tendency for prototype-compatible feature inferences, hence on the right of the space. Further, these differences between the models on Exception trials (left-right in the space) interact differently with the Label versus Feature predictions (up-down in the space) in terms of differences in the amount of attention to the category labels.

Lastly, the hypotheses both imply differential responding depending on the relative proportions of attention given to the category label versus other features as formalized in the models by free attention parameters for the label and separately for each feature dimension. However, the hypotheses also suggest that that these influences might be different across the two models, so to assess this we calculated the proportion of label attention to total attention (i.e., to both label and features) for each simulated participant in the context of each model. These proportions were then tabulated as equivalence bands, e.g., proportions of label attention in the range 0.0 to 0.1 were treated as equivalent, 0.2-0.3 as another range, etc.

In Fig. 5 the proportion of label attention is coded by the gray-scale key such that more attention corresponds to a darker data marker (not to be confused with the gray scaling indicating the proportion of NonException trial responding in Fig. 4). For both models, increases in label attention, and darker markers, correspond to stronger prototype-compatible responding on Label versus Feature trials (up-down in the space) but this interacts differently with Exception trial responding for the two models (left-right). Overall, the models make strongly contrasting predictions in this data space, and this difference is borne out by sharply different predictions for the data from individual participants in the following experiment.

Experiment

The family resemblance category structure used here (see Table 1) is closely related to the category structure used by Yamauchi and Markman (2000). Importantly, it allows the testing trial types described above to be contrasted: Label versus Feature, Exception, and NonException trials.

In addition to the “inference decision making” condition, this experiment included a condition designed to clarify the role of the category labels in inference decision making by indirectly manipulating their relative salience. In this “feature label” condition, one of the feature dimensions in the stimuli, the wings of the rocket ships, was removed from the stimuli throughout the entire experiment and replaced with its corresponding written description, “WIDE WINGS” / “NARROW WINGS”. Other than the replacement of a physical feature with a written descriptor, this “feature label” condition was the same as the feature inference condition including the occurrence of category labels and the same relative positioning of the category instances in the category summary.

In the context of the label super-salience and special-mechanism hypotheses, the purpose of the feature label condition was to evaluate the influence of the category label as the only word feature in the inference decision making condition by introducing another word feature. The intent was to competitively reduce the salience of the category labels as the only word features while leaving category membership information intact. This manipulation of having two word labels for each instance, only one of which was a category label, was suggested by Yamauchi and Markman’s (2000) Experiment 3. Though their manipulation was somewhat different in that each instance only had a single label, the results provided some of their most compelling evidence for the label special-mechanism hypothesis: for participants in their feature inference experiment who were told that the word labels associated with each instance indicated a hidden feature, the proportion of prototype-compatible feature inferences was lower than for participants who were told that the labels indicated category membership. This result seems consistent with the label special-mechanism hypothesis in that the label indicating category membership corresponded to more prototype compatible responding than when it indicated a hidden feature. But an alternative, simpler explanation of Yamauchi and Markman's finding is that the instructional manipulation influenced the salience of the category labels independent of any special tendency to induce integrated prototype-compatible inferences. Saying that a label represents an unseen feature might have reduced attention to it by implying that it is less important than a category label and comparable to other visible features, hence moderating the large expected influence on responding (i.e., the super-salience). This might have particularly been the case as the labels "monek" and "plaple" sound more like categories than features.

In the context of the two hypotheses and corresponding models (Appendix A), if category labels have a special status in terms of invoking a tendency for prototype-compatible feature inferences then a change in their salience should have little if any influence on this tendency as long as their physical salience is sufficient to clearly indicate category membership. But if the dominant influence of category labels on feature inference is due to their super-salience then the competitive salience of another feature should reduce this influence. So the purpose of the feature label condition was to allow a further evaluation of the super-salience and special-mechanism hypotheses by using a manipulation designed to reduce the salience of the category labels but critically with Exception trials present to allow participants to indicate nonintegrative responding.

Method

Participants

There were 25 and 31 participants respectively in the inference decision making and feature label conditions of this experiment, most of whom were undergraduate psychology students at Cardiff University.

Materials and Procedure

All participants were instructed to carefully study the instances from two categories on a category summary sheet (Fig. 1). This summary sheet was constantly available during testing. Participants were asked to infer a missing feature for each of a series of category instances by circling one of the two possible features shown below the instance (e.g., Fig. 3).

The abstract category structure and testing cases are shown in Table 1 together with the abstract-to-physical-feature mapping for the four stimulus dimensions composing the rocket ship stimuli (Fig. 1): wing width, nose cone shape, booster number, and portal orientation. For example, B 1?22 indicates an inference test case where the instance was a member of category B, had narrow wings, two boosters and an up-oriented portal, and participants were asked to infer the missing feature as either a pointed or curved nose. The presentation order of the testing trials is shown in column 3 of Table 1.

The feature label condition was the same as the inference decision making condition except that one of the feature dimensions, the wings, was removed from the physical stimulus and replaced by a word label underneath each rocket ship for all of the category summary instances and testing items throughout the experiment. That is, a rocket with wide wings had the physical wings removed and the written label “WIDE WINGS” placed underneath it, and narrow wings were replaced with “NARROW WINGS”. Otherwise the arrangement of the instances into categories on the summary sheet (Fig. 1) was the same.

Results and Discussion

Inference Decision Making Condition

The prototype-compatible response proportions for the NonException, Label versus Feature, and Exception trials in the inference decision making condition are shown on the left of Fig. 6 with standard error bars, averaged across the trials of a particular type (Table 1). All testing trial data were coded in terms of prototype-compatible responding, that is, in reference to the category label present for each testing case and the category prototypes, A 1111 and B 2222. So for example, 0.92 of the participants in this condition responded to the testing case A ?211, which asked for a response on the first feature dimension, with the prototype-compatible feature 1 from the category prototype A 1111.

Not surprisingly most participants responded to the NonException trials with the prototype-compatible feature, 0.97. This is similar to the 0.89 for this trial type from Experiment 1 of Yamauchi and Markman (2000).

For Label versus Feature trials, most participants responded consistently with the label-based prototypes, 0.77. This indicates even more prototype-compatible responding than on the Label versus Feature trials from Yamauchi and Markman (2000), 0.52.

Lastly, the Exception trials resulted in less than half of the participants’ responses, 0.41, being consistent with the label-based prototype. Yamauchi and Markman (2000) did not have trials of this kind.

In contrast to the conclusions of Yamauchi and Markman (2000), the Exception trial results do not support a special mechanism being invoked by the category label in terms of responding in reference to the category prototypes: More than half of the responses were contrary to the prototype corresponding to the category label even though the majority of other features also matched that prototype. On the other hand, this result is consistent with the label super-salience hypothesis: High salience merely means that the category label has a larger impact on decision making than the other feature dimensions. But if that process is referencing a nearest exemplar, the label and other features both contribute to similarity to that exemplar and hence agree, predicting the same feature.

The Exception trial results do not provide compelling evidence for the special-mechanism hypothesis. But with the average data response proportion of 0.41 fairly close to 0.5, the result is also not clearly different from two-response guessing. This suggests consideration of individual participant data.

The individual data indicate that participants were not just guessing on the Exception trials. Each participant responded to four different Exception trials (Table 1), and the response distribution on the left in Fig. 7 shows the proportion of participants who made a given number (0, 1, 2,...) of prototype-compatible responses across these four trials. The dashed lines in the distribution provide a reference for chance responding as determined by the binomial distribution for each possible number of prototype-compatible responses out of four (based on a response probability of 0.5 for each of the two possible responses on a given trial, *p*(success)=0.5). The response distribution was strongly bimodal and differs dramatically from the binomial distribution reference lines that roughly correspond to guessing. At minimum, these data indicate little evidence of guessing.

Using the data space from the *a priori* model predictions in Fig. 4, a more detailed way of looking at the individual participant data can be had from a scatter plot of the average Exception trial response proportions from Fig. 7 against the average proportions for the four Label versus Feature trials specified in Table 1. This scatter plot in the top panel of Fig. 8 (with identical points slightly offset to indicate data density) suggests quite strong constraints on where participants do and do not tend to fall in this data space. In particular, the pattern of data bears a noticeable resemblance to the *a priori* predictions of the super-salience model at the bottom of Fig. 4 while being in marked contrast to the *a priori* predictions of the special-mechanism model at the top of Fig. 4.

The results of fitting the special-mechanism and super-salience models (Appendix A) to the data from each participant individually are reported in the same Exception against Label versus Feature data space (see Fig. 9). Specifically the data from the top panel in Fig. 8 appear as small diamonds in Fig. 9 while the model fit predictions are shown as circles whose size and gray-scale indicate the RMSD fit value: a large black circle spatially separated from the data diamonds indicates poor predictions while a small white circle centered on a data diamond indicates perfect predictions. Identical model predictions for identical data points are offset slightly to indicate data density (as with the data in the top panel of Fig. 8).

The special-mechanism model (top of Fig. 9) was able to account for the data in the top right hand corner of the data space as was the super-salience model. However, the special-mechanism model was unable to account for the bulk of the data which lie on the left in the scatter plots. Note though that there was one unusual data point in the bottom right corner of the data space which the special-mechanism model accounted for perfectly and the super-salience model accounted for poorly. However overall, the super-salience model accounted for the data at the level of individual participants significantly better than the special-mechanism model (average RMSD 0.16 versus 0.34, $t(48) = -3.73$, $p < 0.001$) and again even the super-salience model’s poorer accounts of some specific individuals still fall in parts of the data space that contain almost all the data unlike the special-mechanism model which failed qualitatively for most of the participants.

Finally, the top panel of Fig. 10 shows the proportion of label attention given by the super-salience model when accounting for each participant in the data space (Fig. 8, top panel), again consistent with the *a priori* predictions at the bottom of Fig. 4. The proportion of label attention was highest in the top right corner and lowest toward the bottom left. (Note that the super-salience model was not able to account for the strange data point at the bottom right hand corner anyway as discussed above so its label attention parameter should be largely ignored here.) Although only observed in some participants, the “super-salience” of the category label should be considered a potential mechanism to account for the dominance of the category label via salience driven selective attention.

Feature Label Condition

An important result for the feature label condition that needs to be emphasized at the outset arises from modelling analysis reported in detail below: the category label attention parameter in the super-salience model was significantly smaller for fits of participants in the feature label condition compared to the inference condition (0.19 versus 0.36 average attention weight parameters, *t*(37) = -2.23, *p* < 0.032, assuming unequal variance). Hence the feature label manipulation did have an influence on the amount of attention the category labels received and by implication their salience. However, the results specifically for the label versus feature trials, as shown on the right in Fig. 6, indicate that the proportion of prototype-compatible responses was not lower than in the standard inference condition on the left. In fact it was (insignificantly) higher. But the Exception trials here resulted in marginally fewer prototype-compatible responses than the decision making inference condition (*t*(37.55) = 1.871, *p*=0.069, assuming unequal variance, or Mann-Whitney *U* = 310.5, *p* = 0.183). And more tellingly, the distribution of Exception trial responding illustrated on the right in Fig. 7 shows a lack of the bimodality found in the inference decision making condition indicating less prototype-compatible responding. In addition, the data space scatter plot in the bottom panel of Fig 8 clearly resembles the *a priori* model predictions in Fig. 4 of the super-salience model much more than those of the special-mechanism model.

The model hypothesis testing results for the feature label condition are shown in Fig. 11. Specifically, the super-salience model shown at the bottom of the figure accounted for the individual participant data significantly better than the special-mechanism model at the top of the figure (average RMSD 0.21 versus 0.44, *t*(60) = -7.80, *p* < 0.001). Even the poorer fits of the super-salience model tend to fall in regions of the data space containing the vast majority of the data (i.e., on the left of the data space). The one exception to this is the unusual data point in the bottom middle of the space (0.5, 025) and neither model was able to account for the data of this participant, in the main because their average response proportion for the NonException trials was very low, 0.25. (This was the only participant in both conditions to score this low on the NonException trials.) Both models were able to account for the data point in the top right corner perfectly by attending solely to the category label. However, the fewer data points in the top right quadrant here in the feature label condition (one, as shown in the bottom panel of Fig. 8) versus in the inference decision making condition (seven, as shown in the top panel of Fig. 8) is consistent with less attention to the label. Likewise the proportions of label attention in the super-salience model fits to individual participants are consistent with this as shown in the data space at the bottom of Fig. 10. Thus the feature label manipulation reduced label salience in a way which can be well accounted for by the super-salience hypothesis.

General Discussion

A category label serves as a marker of category membership, but at least in perceptual categorization, a label is also likely to be a highly salient feature which attracts a lot of attention. The experiment reported here presented a summary of family-resemblance perceptual categories (Fig. 1) and asked participants to make a variety of feature inferences (Table 1) in this purely decision making task with the category summary constantly present. The Label versus Feature trials pitted the category label from one category against several typical features of the other category and showed that category labels have a dominant influence on feature inference relative to other features, consistent with prior research (e.g., Gelman & Markman, 1986; Sloutsky, 2003; Yamauchi & Markman, 2000; but see Deng & Sloutsky, 2013). We have evaluated this influence in the context of two hypotheses and corresponding formal models (Appendix A): The label special-mechanism hypothesis implies that the category label invokes the integration of category information in working memory during feature inference decision making because it directly indicates category membership, unlike other features, and then critically results in a systematic tendency for prototype-compatible feature inferences. In contrast, the more parsimonious label super-salience hypothesis implies that the dominant influence of category labels on feature inference can be explained, not by invoking a special mechanism, but by their salience and corresponding influence on the similarity assessment process. The additional mechanism of integrative decision making producing prototype-compatible feature inferences is not needed. Exception test trials were very similar to particular category instances which nonetheless had atypical, prototype-incompatible features for the queried feature. The results for these trials in particular were consistent with the label super-salience hypothesis and did not support the need for the additional processes in the label special-mechanism hypothesis.

Most importantly, both the *a priori* predictions and individual participant fits of models formalizing these two hypotheses (Appendix A) strongly falsified the special-mechanism hypothesis and were consistent with the super-salience hypothesis. Lastly, an additional feature label condition replaced one of the perceptual features of the category instances with a word feature so as to compete with the salient category labels as the only words present in the inference decision making condition. The modelling results for these feature label condition data indicate that the category labels received less attention in this condition compared to the inference decision making condition, but in addition these results still strongly falsified the special-mechanism model while being consistent with the super-salience model.

A constrained view of these conclusions is that they represent an even more focused version of non-normative *single-category* influence on feature inference in the face of uncertain categories (Murphy & Ross, 1994) down to single nearest instances for specified category membership, in contrast to the multiple influences prescribed by, for example the Rational Model (Anderson, 1991), and normative Bayesian perspectives in general. Further, in this context, it can be argued that this nearest neighbour kind of reasoning arises out of the affordances of the pure decision making task and its summary presentation of categories rather than being indicative of how decision making based on internal representations actually works. However, if that is true for the present research then it also quite strongly constrains the conclusions, not only of Yamauchi and Markman (2000) specifically, but, more generally, the many studies that have evaluated decision making in reference to external summary representations. At minimum, the present research suggests that evaluations of decision making in reference to an external representation need to measure the possibility of nearest neighbour effects especially when drawing conclusions that seem to imply representational integration of category information.

It is worth emphasizing that the special mechanism view of category labels is intuitively compelling in a way that is theoretically challenging and not simply a straw man: It is almost impossible to conceive of categories as being adaptive without the ability to mediate hidden feature prediction via the integration of instance information. But for categories to be more predictive than random sets of instances, it seems unavoidable that they must integrate the instance information such that one feature is a more probable inference than another, for example if more instances have had that feature. Put even more strongly, this basically corresponds to the idea that inferring a missing feature for a category member should normatively reference the category prototype because of the special category-indicating role of the category label. Murphy’s (2002) summary statement about category-based induction is relevant here: “If read literally, almost all the work on category-based induction takes a prototype view of concepts” (p. 265). There are, however, several reasons why a normative status for prototypes in the context of feature inference is potentially misleading:

Even if a category instance is clearly a member of only a single category, an unlikely state of affairs in the real world, there are many cases where high similarity to a particular instance of the category should override category level information in terms of making a feature inference. For example, having been told that a particular instance is a bird would generally make the inference that it flies quite sensible unless the instance was particularly similar to a penguin, emphasizing that there are times when membership in some categories is best ignored. At least sometimes basing feature inferences on strong similarity to specific instances seems very adaptive.

Fundamentally, categories can possess whatever magic ingredient makes them more predictive for feature inference than random groupings of instances without having a special status in terms of invoking special mechanisms relative to other instance features. That is, category labels can be functionally predictive without invoking prototype-compatible inferences because categories already represent a higher level of abstraction than their instances. But this abstraction is represented by an additional feature corresponding to a category label. As such they already have the potential to incorporate additional information which allows them to influence similarity assessment in a functional way that still need not be qualitatively, mechanistically different than other features. As an example, an instance may have the following features some of which are more abstract than others: feathers, wings, clever-looking, member-of-bird-category, edible, etc. In terms of inferring a hidden feature, the alternative to the special-mechanism hypothesis is that category labels are just particularly abstract and perhaps very salient features which are otherwise treated comparably to other features in selective attention-driven similarity assessment.

While this research does not support a special mechanism view of category labels specifically in terms of invoking category prototypes in the context of inference decision making, it does not preclude other kinds of special status for category labels relative to other features. For example, even high label salience arguably gives the label a special status relative to other features. And at a higher level, Yamauchi, Kohn, and Yu (2007) used a mouse tracking paradigm to demonstrate that participants spent more time looking at category labels than other features. More generally there have been a variety of demonstrations that category labels result in behavioral differences from other features (e.g., Yamauchi & Yu, 2008; Yamauchi, 2009), and these have been used to support the idea that category labels indicate category membership and as such serve as an indicator that a rich category structure of featural information is available. Having said that, when category labels are not available and/or category membership is uncertain, people often use feature-based strategies to drive feature inference (e.g., Griffiths, Hayes, & Newell, 2012). So full clarification of the role that category labels play in feature inference will likely require well understood category representations and thoroughly evaluated salience and selective attention to assess whether special mechanisms are required for feature inference beyond those for categorization.

References

Anderson, J. R. (1991). The adaptive nature of human categorisation. *Psychological Review, 98*, 409-429.

Blair, M., & Homa, D. (2001). Expanding the search for a linear separability constraint on category learning. *Memory & Cognition, 29,* 1153-1164.

Deng, W., & Sloutsky, V. M. (2013). The role of linguistic labels in inductive generalization. *Journal of Experimental Child Psychology, 114,* 432-455.

Estes, W. K. (1986). Array models of category learning. *Cognitive Psychology, 18,* 500-549.

Gelman, S., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23,* 183-209.

Griffiths, O., Hayes, B. K., & Newell, B. R. (2012). Feature-based versus category-based induction with uncertain categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 576-595.

Hayes, B., Heit, E., & Swendsen, H. (2010). Inductive reasoning. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 278-292.

Homa, D., Rhoads, D., & Chambliss, D. (1979). Evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 11-23.

Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 418-439.

Johansen, M. K., & Kruschke, J. K. (2005). Category representation for classification and feature inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1433–1458.

Murphy, G. L. (2002). *The Big Book of Concepts.* MIT Press, Cambridge, Massachusetts and London, England.

Murphy, G. L., & Ross, B. H. (1994). Predictions from uncertain categorisations. *Cognitive Psychology, 27*, 148-193.

Murphy, G. L., & Ross, B. H. (2005). The two faces of typicality in category-based induction. *Cognition, 95,* 175-200.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104-114.

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

Nosofsky, R. M., & Zaki, S. R. (2002). Exemplar and prototype models revisited: Response strategies, selective attention, and stimulus generalization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 924-940.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77,* 353-363.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7,* 246-251.

Sloutsky, V. M., & Fisher, A. V. (2004). Induction and categorization in young children: A similarity-based model. *Journal of Experimental Psychology: General, 133,* 166-188.

Sloutsky, V. M., & Fisher, A. V. (2011). The development of categorization. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation, 54,* 141-166.

Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1411-1436.

Smith, J. D., & Minda, J. P. (2000). Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26,* 3-27.

Yamauchi, T. (2009). Finding abstract commonalties of category members. *Journal of Experimental and Theoretical Artificial Intelligence*. 21 (3), 155-180.

Yamauchi, T, & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language, 39*, 124-148.

Yamauchi, T, & Markman, A. B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 776-795.

Yamauchi, T. & Yu, N. (2008). Category labels versus feature labels: Category labels polarize inferential predictions. *Memory & Cognition*. 36 (3), 544-553.

Yamauchi, T., Kohn, N., & Yu, N. Y. (2007). Tracking mouse movement in feature inference: Category labels are different from feature labels. *Memory & Cognition*, 35, 852-863.

Acknowledgments

We thank Takashi Yamauchi and two anonymous reviewers for detailed comments and suggestions on a previous version of this manuscript.

Appendix A

Specification of the special-mechanism and super-salience hypotheses as formal models together with Monte-Carlo simulation and parameter fitting procedures

The special-mechanism and super-salience hypotheses were formalized by extending exemplar and prototype models of categorization to feature inference decision making as in their application to feature inference category learning (Johansen & Kruschke, 2005; Yamauchi & Markman 1998). Specifically the super-salience model was adapted from Nosofsky’s (1986) Generalized Context Model which embodies exemplar representation, and the special-mechanism model was adapted from a multiplicative prototype model (e.g., Estes, 1986) which of course embodies prototype representation.

Super-salience model. The super-salience model used Equation 1 to calculate the similarity, η_{ij}, between a particular test item *i* in the bottom of Table 1 (e.g., an Exception trial) and a particular category instance *j* in the top of Table 1 (e.g., a member of category A). The stimulus feature value for the test item on stimulus dimension *k* in Table 1, x_{ik} in Equation 1,

$$\eta_{ij} = \exp\left(-c \sum_{k \neq respdm} w_k \mid x_{ik} - x_{jk} \mid \right) \qquad (1)$$

is compared to the feature value for the instance in the representation also on dimension *k*, x_{jk}, by taking the absolute differences between them, | x_{ik} - x_{jk} |, which is then multiplied by a dimensional attention value, w_k. These attentionally-weighted dimensional differences are summed across the *k* dimensions where for the category structure in Table 1 *k* = 5, one dimension for the category label plus four other feature dimensions. However, even though there were five different dimensions for this structure (Table 1), a given test item only included four of them, as indicated by *k* ≠ respdm for the summation, where the missing dimension is the response dimension for that test case (the ?’s in the testing items in Table 1). Lastly, the weighted sum is multiplied by –*c*, a similarity scaling parameter for the similarity space, and exponentiated. In this model the category label therefore functions exactly like a feature, although its salience may be greater.

The super-salience model generates the response probability of a given feature for a particular test trial using Equation 2. The probability of feature 1 for testing case *i*, *P*(*f* = 1 | *i*),

$$P(f=1\,|\,i)=\frac{\left(\sum_{j\in f=1}\eta_{ij}\right)^{\gamma}}{\left(\sum_{j\in f=1}\eta_{ij}\right)^{\gamma}+\left(\sum_{j\in f=2}\eta_{ij}\right)^{\gamma}} \qquad (2)$$

is the sum of the similarities to instances that have *f* = 1 on the response dimension ($j\in f=1$) raised to the power *γ*, the response determinism parameter, and then divided by the same sum as the numerator plus the sum of similarities to instances that have *f* = 2. (The response determinism parameter specifies how strongly a given amount of evidence gets pushed toward the response probability extremes of 0 or 1.)

Special-mechanism model. The special-mechanism model for feature inference also used Equation 1. But instead of calculating similarity to each of the category instances (at the top of Table 1), it calculates similarity between a test item and each of the category prototypes, A 1111 for category A or B 2222 for category B.

The response probability equation for the special-mechanism model, given in Equation 3, looks similar to Equation 2 except with similarity to prototypes rather than exemplars. That is, the probability of feature 1 for test instance *i*, P(*f* = 1 | *i*), is the similarity to the prototype

$$P(f=1\,|\,i)=\frac{(\eta_{i1})^{\gamma}}{(\eta_{i1})^{\gamma}+(\eta_{i2})^{\gamma}} \qquad (3)$$

with a 1 feature on the response dimension, η_{i1}, raised to power *γ*, divided by that similarity plus the similarity to the prototype with a 2 feature on the response dimension, also raised to power *γ*.

Finally, P(*f* = 1| *i*) for each test case was converted to the proportion of prototype-compatible responding for that test case (to match the tabulation of the participant data in Figs. 6 and 7) depending on the category label present for that test case: If the label was for category A (in Table 1) then the proportion of prototype-compatible responding was directly P(*f* = 1| *i*). However, if the category label was for category B (in Table 1) then the proportion of prototype-compatible responding was 1 - P(*f* = 1| *i*) as there are only two categories in this category structure.

Fitting procedure. As specified above, both models have seven parameters: one label attention parameter, four feature attention parameters, one similarity scaling parameter and one response determinism parameter. However for both models, the similarity scaling parameter is underconstrained/redundant if all five attention parameters are free. Alternatively, the similarity scaling parameter can be treated as a free parameter and the five attention parameters constrained to sum to 1 with only four of them free parameters. Lastly, the response determinism parameter is underconstrained/redundant for the prototype model (Nosofsky & Zaki, 2002).

The Monte-Carlo simulations for the special-mechanism and super-salience models based on the category structure in Table 1 were generated by sampling random values for the free parameters in Equations 1-3 for each simulated participant: The label and feature attention parameters were randomly sampled from the interval [0,1] while the specificity parameter, *c*, and the response determinism parameter, γ, were sampled from the interval [0,10]. Both models were applied to the same simulated participants. The predictions for a given set of parameters were tabulated for each set of testing trials—NonException, Label versus Feature, and Exception—by averaging across the response proportions for trials of a given type (e.g., the four NonException trials in Table 1). The results for these simulations are shown in Figs. 4 and 5.

The models were fitted to individual participant data by adjusting the nonredundant free parameters in the above equations via a hill-climbing procedure to minimize the discrepancy between the data and the model predictions for all of the individual testing trials in Table 1 as determined by root-mean-squared deviation. Multiple starting points were used for the hill-climbing procedure to determine best fitting parameters. Note that although the model predictions were tabulated by calculating average response proportions across trials of a given type to match the way the data are tabulated in Fig. 8, the models were fitted to all 12 individual testing trials at the bottom of Table 1 for each participant.

Table 1. Family Resemblance Abstract Category Structure and Test Cases in the Experiment

Abstract features	Trial types	Trial order
A 2 1 1 1		
A 1 2 1 1		
A 1 1 2 1		
A 1 1 1 2		
A 2 1 1 1		
A 1 2 1 1		
A 1 1 2 1		
A 1 1 1 2		
B 1 2 2 2		
B 2 1 2 2		
B 2 2 1 2		
B 2 2 2 1		
B 1 2 2 2		
B 2 1 2 2		
B 2 2 1 2		
B 2 2 2 1		
B 1 ? 2 2	NonException	3
A ? 2 1 1	NonException	5
B ? 1 2 2	NonException	10
A 2 ? 1 1	NonException	12
A 2 1 2 ?	Label v. Feature	4
B 1 1 ? 2	Label v. Feature	7
A 2 2 ? 1	Label v. Feature	9
B 1 2 1 ?	Label v. Feature	11
A 1 ? 1 1	Exception	1
A ? 1 1 1	Exception	2
B 2 ? 2 2	Exception	6
B ? 2 2 2	Exception	8

Note. The abstract category structure is composed of the 16 instances specified at the top of the left column, and the testing items are at the bottom where a "?" indicates the dimension on which participants were asked to infer the missing feature. The assignment of abstract to physical stimulus dimensions was: category label category (A)="Planet A" and category (B)="Planet B"; dimension 1 was wing width, 1=narrow and 2=wide; dimension 2 was nose cone shape, 1=curved and 2=pointed; dimension 3 was booster number, 1 or 2; and dimension 4 was portal orientation, 1=down and 2=up. The second column specifies test trials of equivalent types as specified in the main text. The actual order of the test cases is specified in the third column.

Figure 1. Category summary sheet with instances from two categories corresponding to the abstract category structure in Table 1.

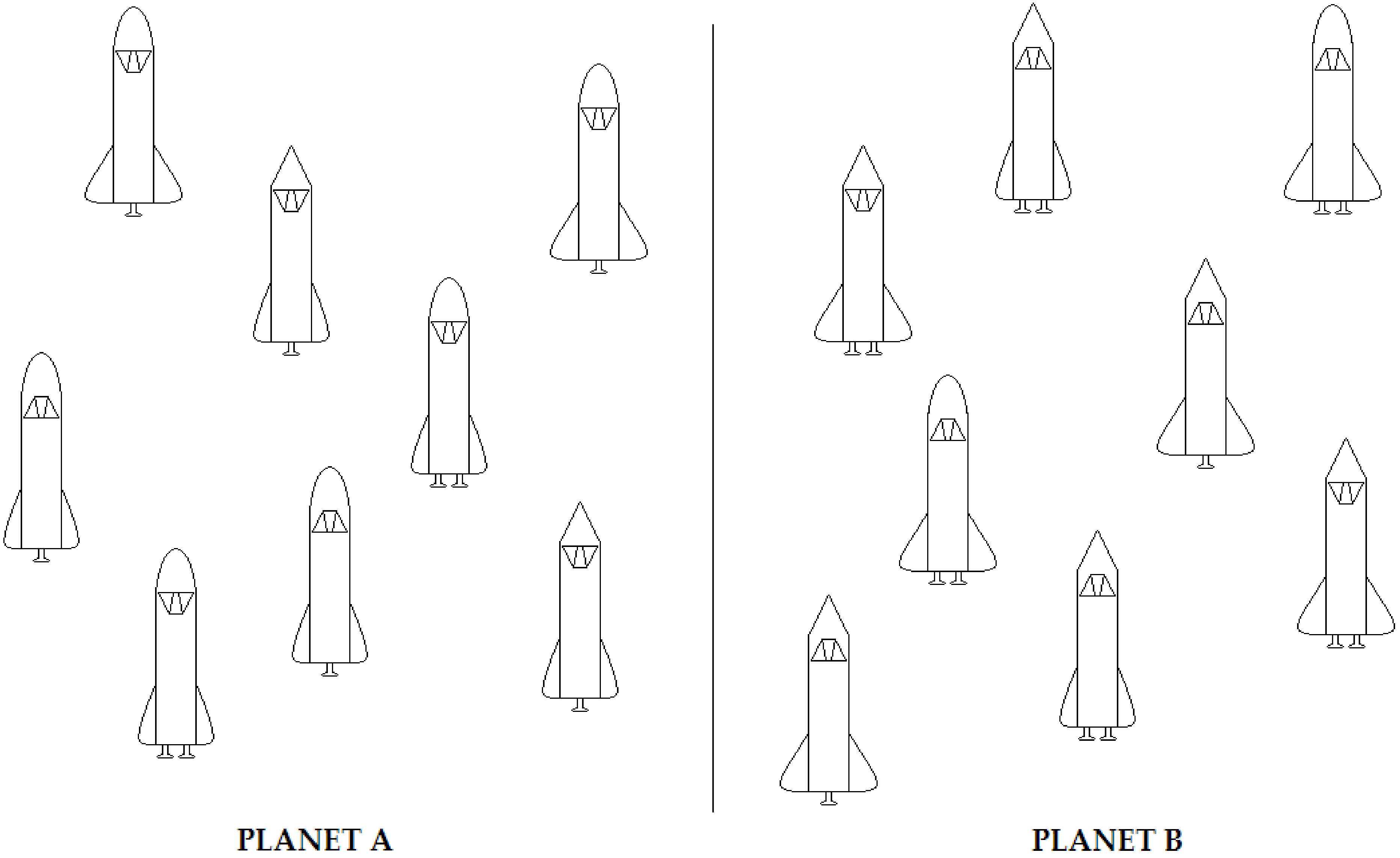


Figure 2. Schematic of the difference between the special-mechanism and super-salience models.

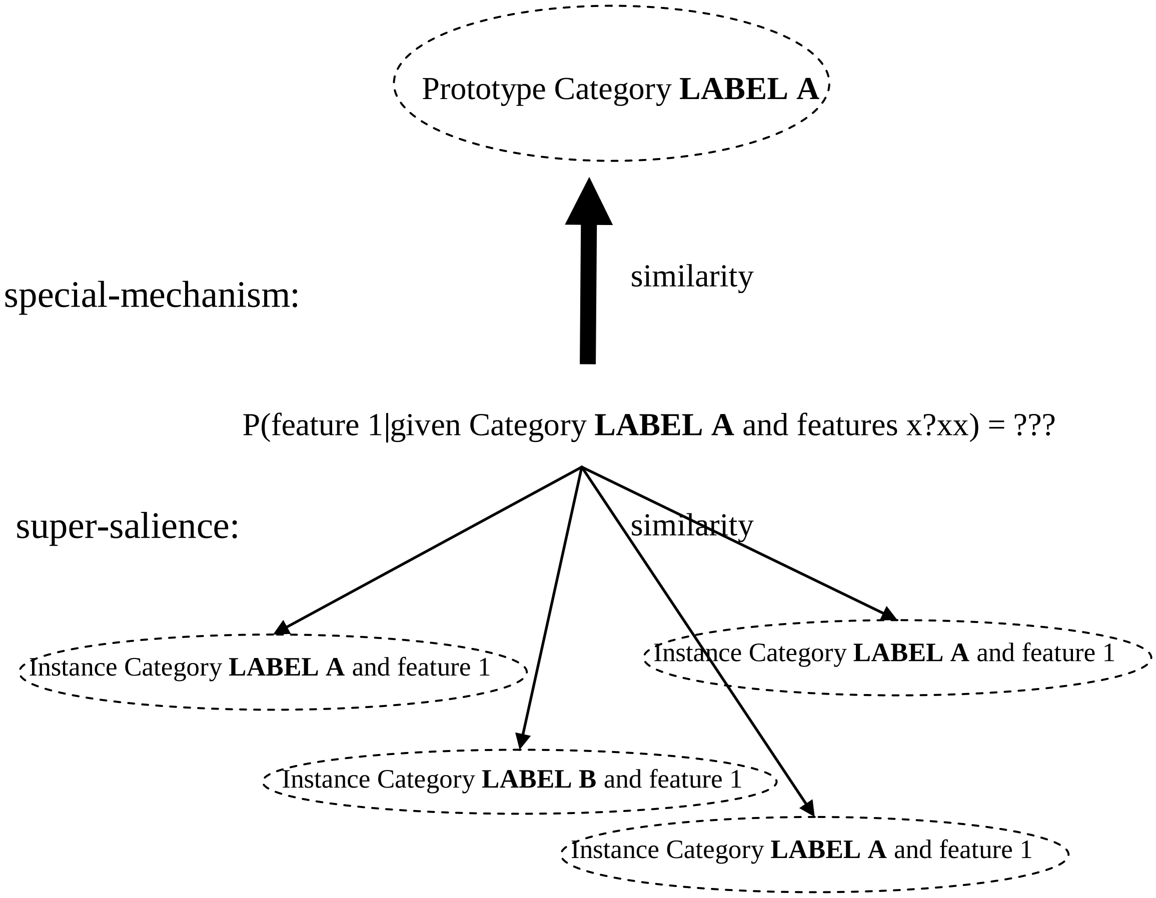
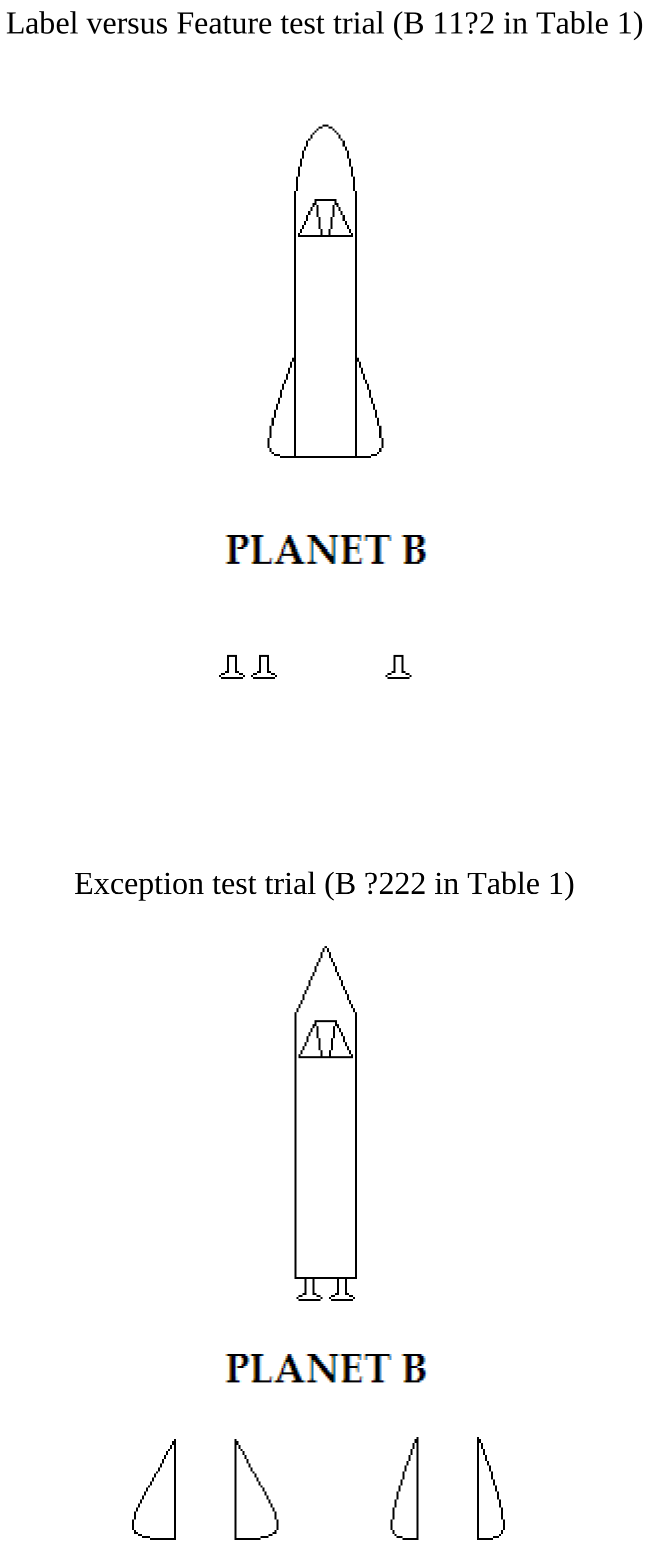

Prototype Category **LABEL A**
similarity
special-mechanism:
P(feature 1|given Category **LABEL A** and features x?xx) = ???
super-salience:
similarity
Instance Category **LABEL A** and feature 1
Instance Category **LABEL A** and feature 1
Instance Category **LABEL B** and feature 1
Instance Category **LABEL A** and feature 1

Figure 3. Examples of testing trials from the experiment (Table 1) with a choice between two possible response features.



Label versus Feature test trial (B 11?2 in Table 1)

Exception test trial (B ?222 in Table 1)

Figure 4. *A priori* predictions from the models corresponding to the label special-mechanism (top) and label super-salience (bottom) hypotheses for the test cases (Table 1) based on 3000 simulated participants each using random parameter values as described in Appendix A. p.c. = prototype-compatible. The gray-scale of the markers corresponds to average response proportion ranges for the NonException trials.

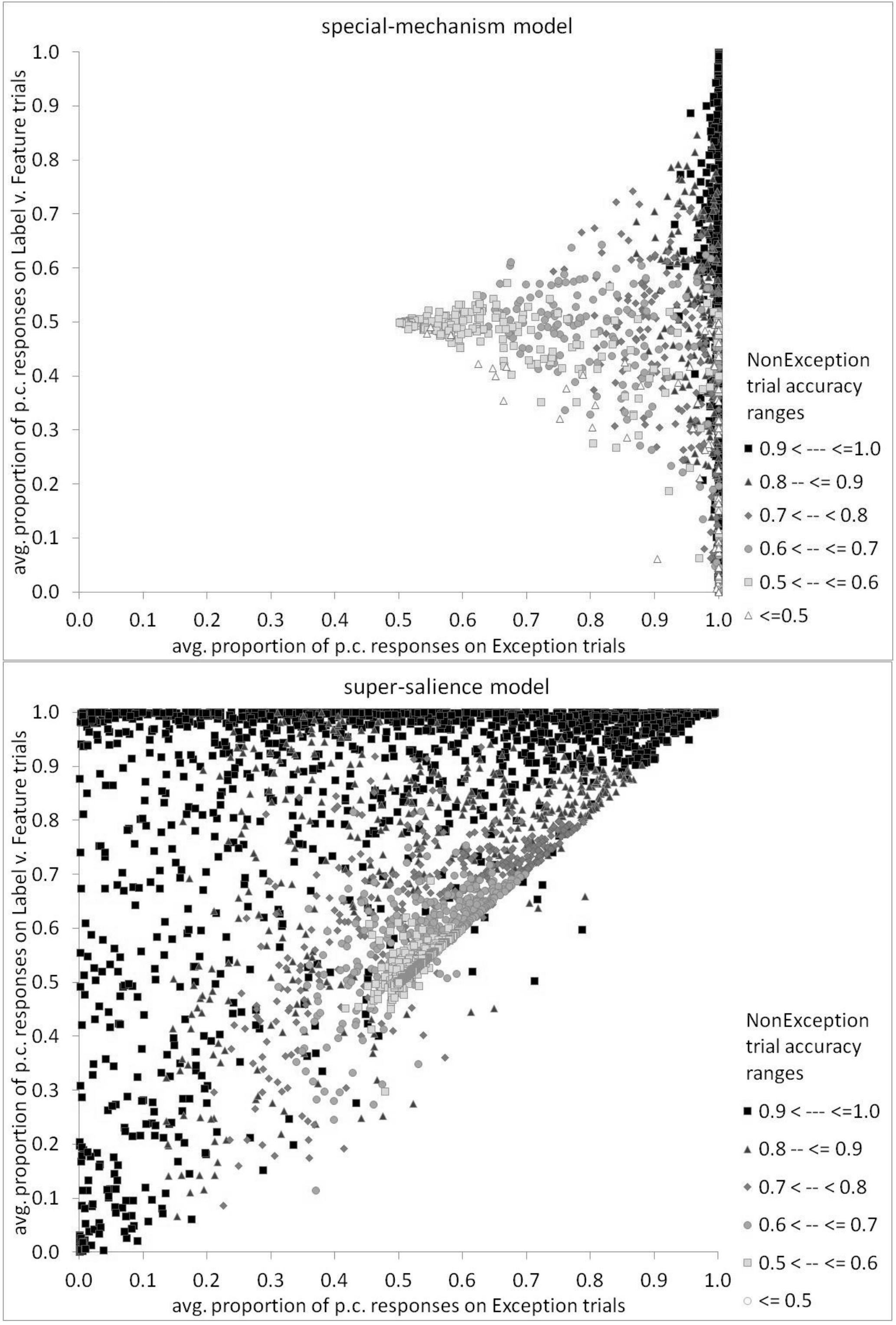


Figure 5. *A priori* label attention predictions from the models corresponding to the label special-mechanism (top) and super-salience (bottom) hypotheses for the test cases (Table 1) based on 3000 simulated participants (the same 3000 as in Fig. 4) each using random parameter values as described in Appendix A. p.c. = prototype-compatible. The gray-scale of the markers is the proportion of label attention, unlike in Fig. 4.

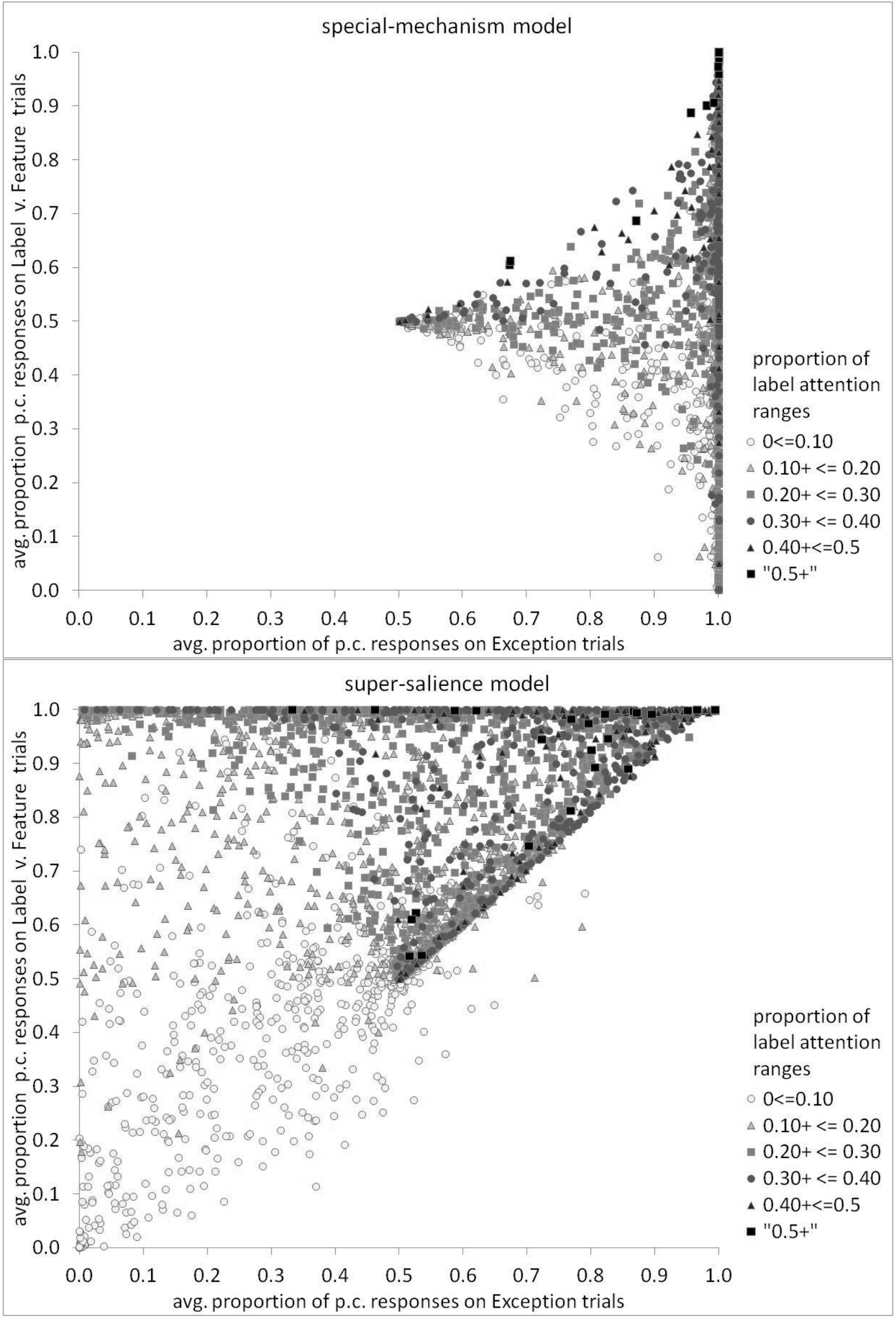


Figure 6. Results from the inference decision making and feature label conditions in terms of average prototype-compatible response proportions by trial type (with standard error bars). See the main text for the definitions of the NonException, Label versus Feature, and Exception trial types.

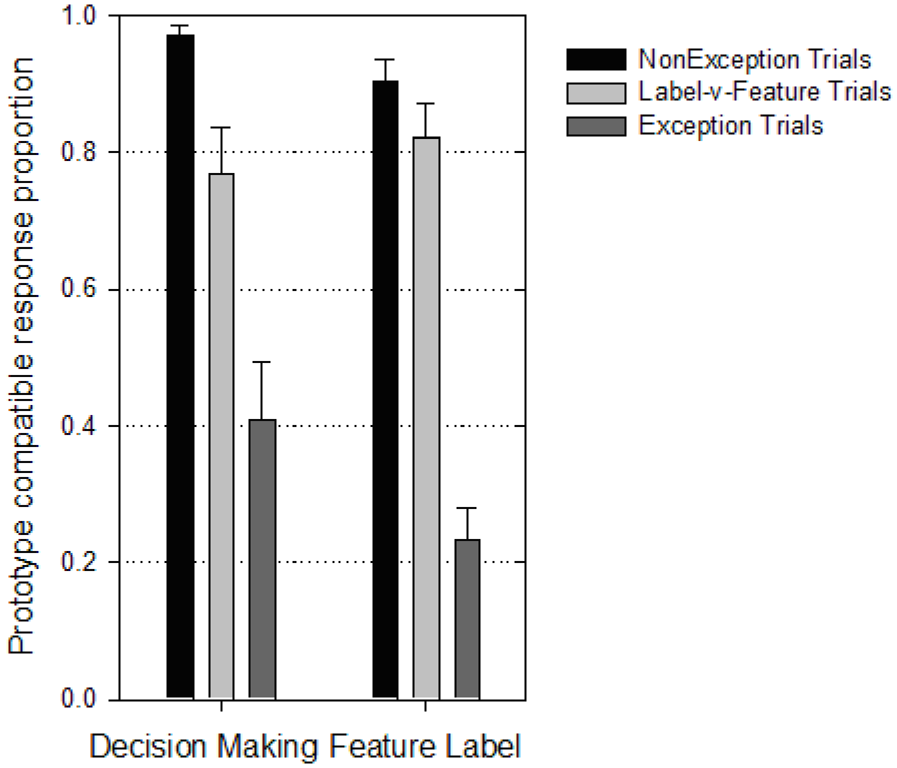


Figure 7. Distributions of the number of prototype-compatible responses across the four Exception trials (Table 1) in the decision making inference and feature label conditions. The dashed reference lines indicate the proportion of participants expected to make a given number of prototype-compatible responses by chance out of the four Exception trials as determined by the binomial distribution, *p*(success)=0.5, if responding corresponded to random guessing between the two possible choices on each of the trials.

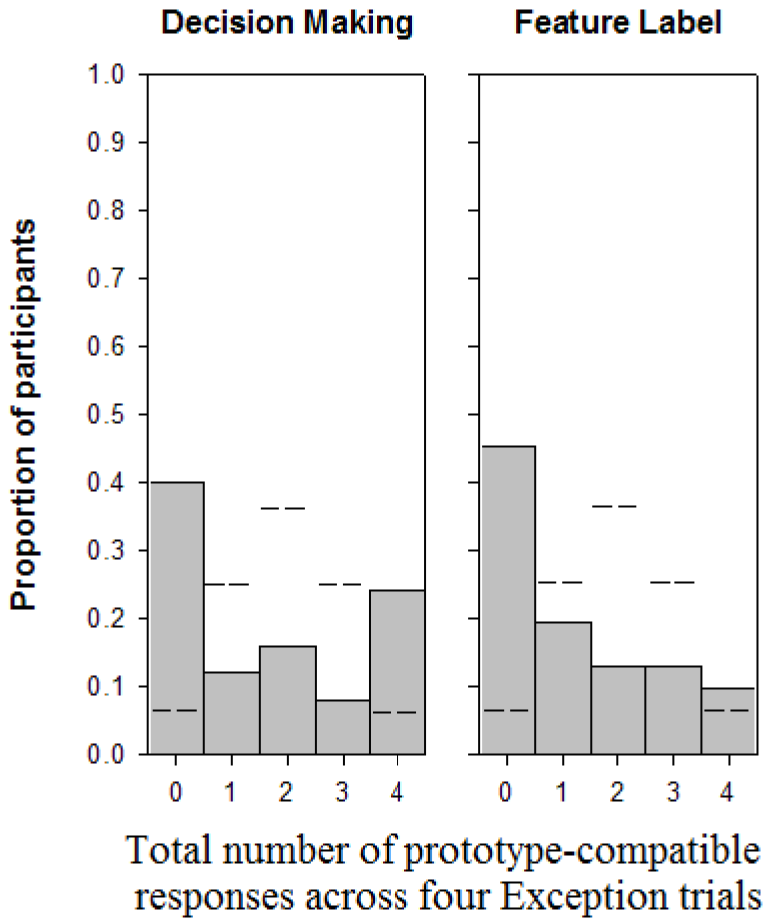


Figure 8. Average proportion of prototype-compatible (p.c.) responses across the trials of a given type, Exception and Label versus Feature (Table 1), plotted against each other for the inference decision making (top) and feature label (bottom) conditions. A small, systematic offset has been added to some data points that had identical values so data density can be seen, for example, the four data points in the top right hand corner of the top panel.

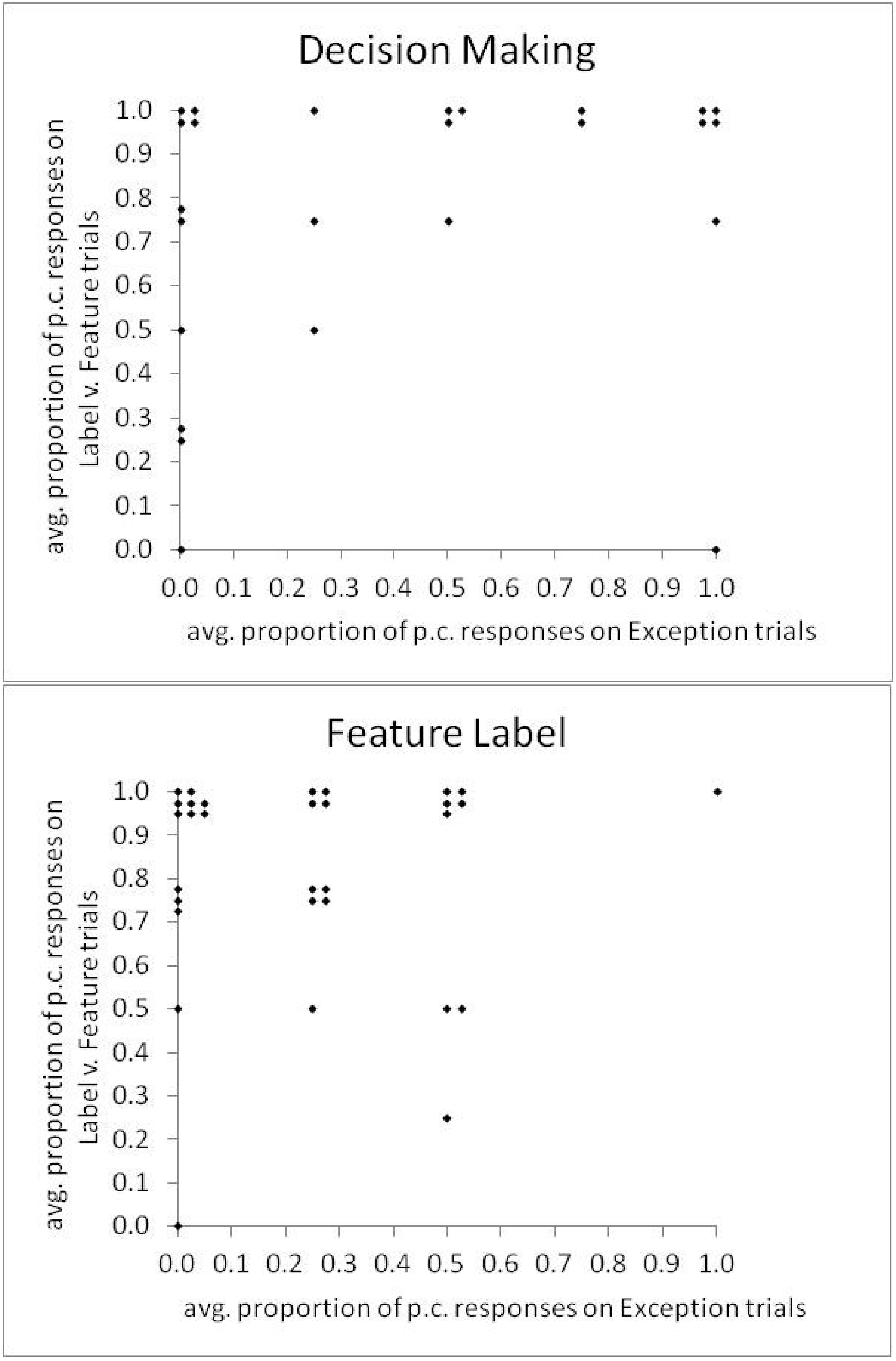


Figure 9. Special-mechanism and super-salience model fits to individual participant data from the inference decision making condition. The data are indicated by small black diamonds while the models' predictions are indicated by circles whose size and gray-scale shade indicate how well the model fit a given participant as measured by RMSD. A model accounting for a participant's data perfectly corresponds to a small light-colored circle centered on a small diamond (e.g., the four data points in the top right hand corner). When a model did not account for a participant well this corresponds to a large dark circle far away from the small diamond indicating the data. Some of the data and model predictions have been shifted slightly so as not to obscure each other. p.c. = prototype-compatible.

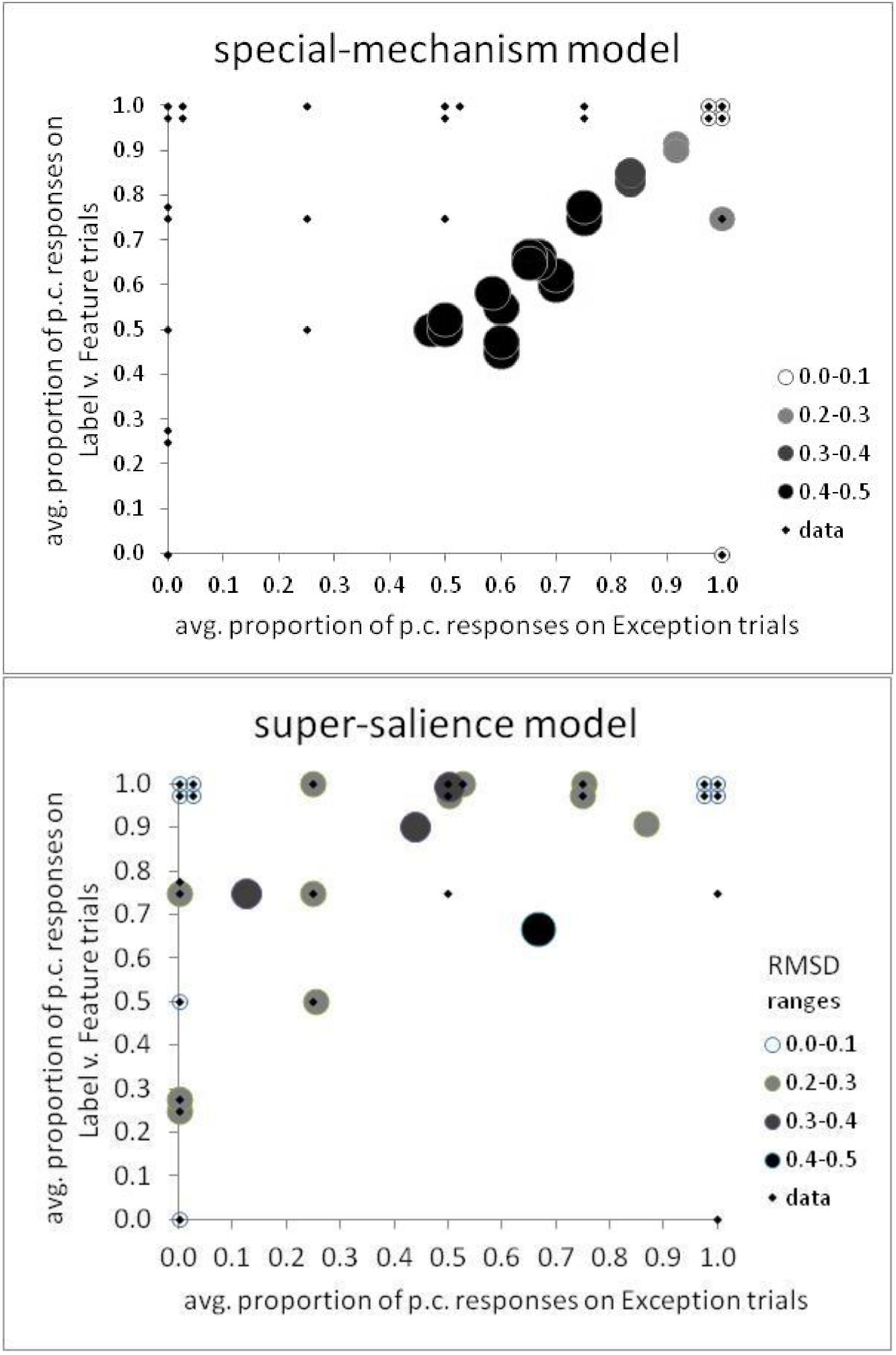


Figure 10. Proportion of label attention in different ranges for the fit of the super-salience model to the data for each participant in the Exception against Label versus Feature data space for the inference decision making condition (top panel), corresponding to the data space in the top panel of Fig. 8), and the feature label condition (bottom panel) corresponding to the data space in the bottom panel of Fig. 8).

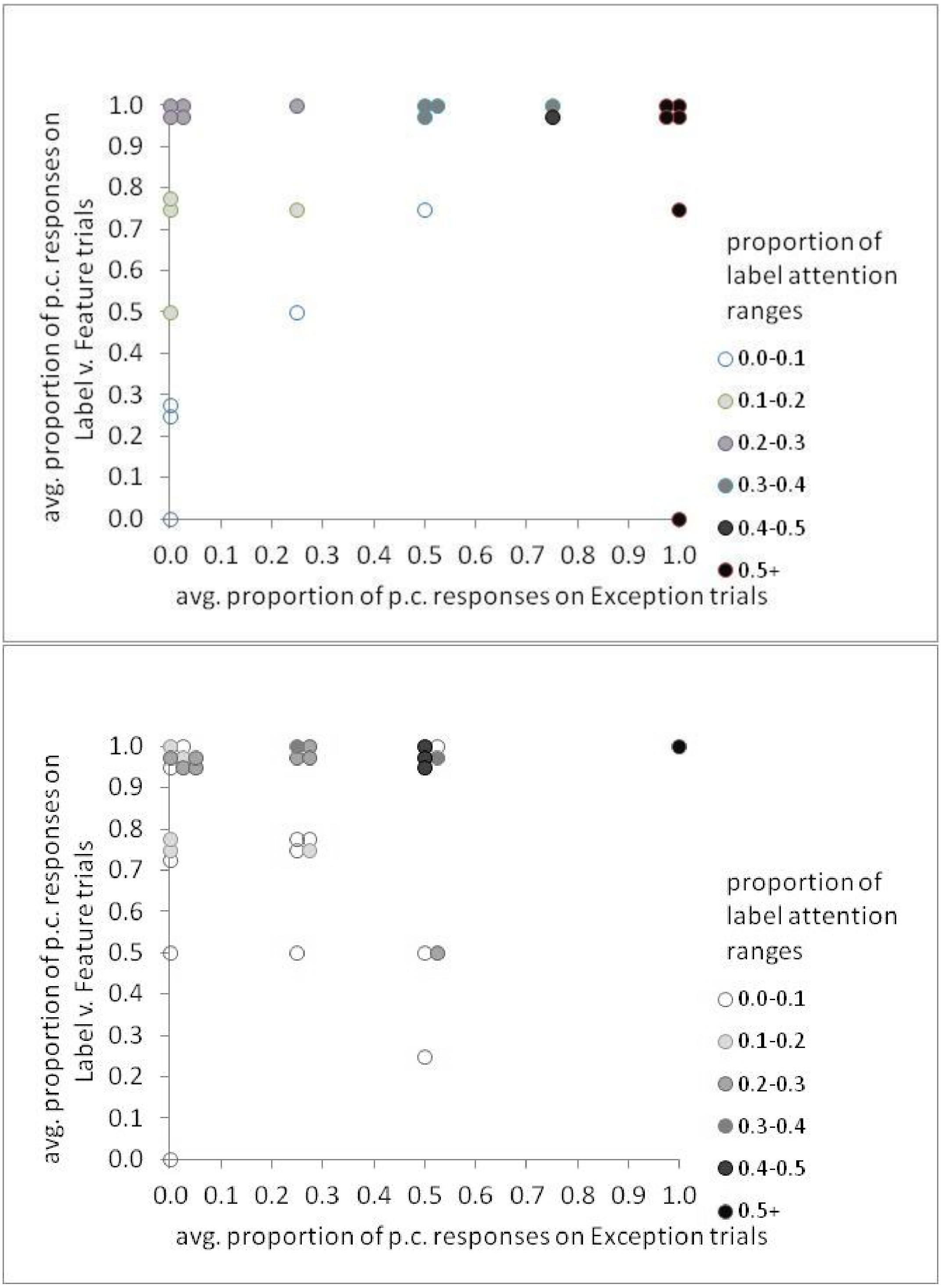
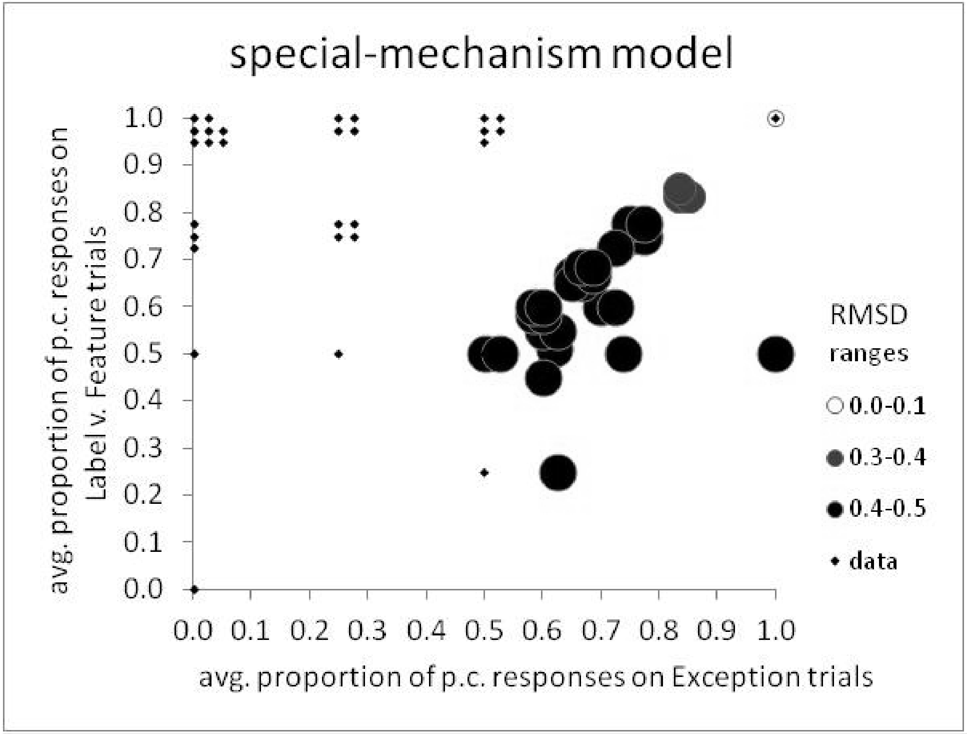
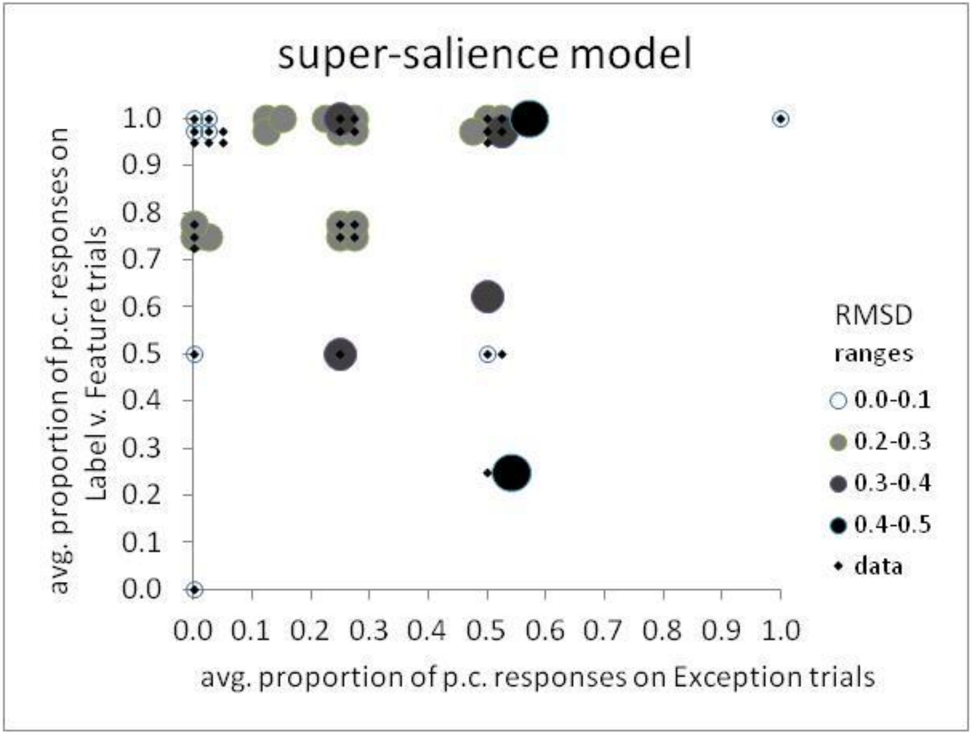


Figure 11. Special-mechanism and super-salience model fits to individual participant data from the feature label condition. The data are indicated by small black diamonds while the models' predictions are indicated by circles whose size and gray-scale shade indicate how well the model fit a given participant as measured by RMSD. When the model accounted for a participant's data perfectly this corresponds to a small light-colored circle centered on a small diamond (e.g., the one data point in the top right hand corner). When a model did not account for a participant well this corresponds to a large dark circle far away from the small diamond indicating the data. Some of the data and model predictions have been shifted slightly so as not to obscure each other. p.c. = prototype-compatible.


special-mechanism model
avg. proportion of p.c. responses on Label v. Feature trials
avg. proportion of p.c. responses on Exception trials
RMSD ranges
○ 0.0-0.1
● 0.3-0.4
● 0.4-0.5
• data


super-salience model
avg. proportion of p.c. responses on Label v. Feature trials
avg. proportion of p.c. responses on Exception trials
RMSD ranges
○ 0.0-0.1
● 0.2-0.3
● 0.3-0.4
● 0.4-0.5
• data