**Supplement to “How robust are probabilistic models of higher-level cognition?”**

Gary Marcus, Dept. of Psychology, New York University

Ernest Davis, Dept. of Computer Science, New York University

This supplement explains some of the calculations cited in the main paper.

It should be emphasized that all of the alternative models that we analyze below are probabilistic models of the same structure as the ones proposed in the papers that we discuss. In the section on "Balance Beam Experiments", we are applying the identical model to a very closely related task. In the sections on "Lengths of Poems" and on the "Communication" experiment, we are determining the results of applying a model of the same structure to the same task using either priors drawn from more accurate or more pertinent data, or a superior decision rule for action.

**Balance Beam Experiments**

Consider a balance beam with two stacks of discs: a stack of weight *p* located at *x=a <* 0 and a stack of weight *q* located at *x = b >* 0. The fulcrum is at 0. Assume that a hypothetical participant judges both positions and weights to be distributed along a truncated Gaussian cut off at 0, with parameters 〈*a*, σx〉, 〈b, σ*x*〉, 〈*p*, σ*w*〉, 〈*q* σ*w*〉, respectively.

The truncation of the Gaussian corresponds to the assumption that the participant is reliably correct in judging which side of the fulcrum each pile of disks is on and knows that the weights are non-negative; it is analogous to the “deterministic transformation that prevents blocks from interpenetrating” described in Hamrick et al. A random sample of points in a Gaussian truncated at 0 is generated by first generating points according to the complete Gaussian, and then excluding those points whose sign is opposite to that of the mean.



Figure 1: Balance beam with a=-3, b=2, p=3, q=4.

Following the model proposed by Hamrick et al., we posit that the participant computes the probability that the balance beam will lean left or right, by simulating scenarios according to this distribution and using the correct physical theory; and he predicts the outcome with higher probability.

If σ*x* and σ*w* are both small, then the participant's perceptions are essentially accurate, and therefore the model predicts that participant will answer correctly. If σ*x* is large and σ*w* is small, then essentially the participant is ignoring the location information and just using the heuristic that the heavier side will fall down. If σ*x* is small and σ*w* is large, then essentially the participant is ignoring weight information and using the heuristic that the disk further from the fulcrum will fall down.

The interesting examples are those in which these two heuristics work in opposite direction. In those, there is a tradeoff between σ*x* and σ*w*. Consider an example where the weight heuristic gives the wrong answer, and the distance heuristic gives the right answer. For any given value of σ*w* there is a maximal value m(σ*w*) such that the participant gets the right answer if σ*x* <m(σ*w*) and gets the wrong answer if σ*x* >m(σ*w*). Moreover m is in general an increasing function of σ*w* because, as we throw away misleading information about the weight, we can tolerate more and more uncertainty about the distance. Conversely, if the weight heuristic gives the right answer and the distance heuristic gives the wrong one, then for any given value of σ*x* there is a maximal value m(σ*x*) such that the participant gets the right answer if σ*w* <m(σ*x*) and gets the wrong answer if σ*x* >m(σ*x*).

The balance beam scenario and the supposed cognitive procedure are invariant under scale changes both in position and in weight, independently. Therefore, there is a two degree-of-freedom family of scenarios to be explored (the ratio between the weights and the ratio between the distances). The scenario is also symmetric under switching position parameters with weight parameters, with the appropriate change of signs.

The results of the calculations are shown below. The simple MATLAB code to perform the calculations can be found at the web site for this research project: http://www.cs.nyu.edu/faculty/davise/ProbabilisticCognitiveModels/

The technique we used was simply to generate random values following the stated distributions, and to find, for each value of σ*x*, the value of σ*w* where the probability of an incorrect answer was greater than 0.5. In cases where

either σ*x* or σ*w* is substantial, this probability is in any case very close to 0.5. The values for m(σ) for large values of σ are therefore at best crude approximations; we are looking for the point where a very flat function crosses the value 0.5 using a very crude technique for estimating the function. A more sophisticated mathematical analysis could no doubt yield much more precise answers; however, there is no need for precision since our sole objective is to show that the values of σ needed are all very large.

The result of all the experiments is that the hypothetical participant (in contrast to the more error-prone performance human participants, both children and many adults) always gets the right answer unless the beam is very nearly balanced or the values of σ are implausibly large. Moreover, for any reasonable values of σ*x,*  σ*w*, the procedure gets the right answer with very high confidence. For instance if σ*x,* = σ*w* = 0.1, in experiment 1 below, the participant judges the probability of the wrong answer to be about 0.05; in experiment 2, less than 10-6; in experiment 3, about 0.03; in experiment 4, about 0.0001. Cases 3 and 4 here correspond to the "Conflict-distance" and "Conflict-weight" examples, respectively, of Jansen and van der Maas.

Case 1: a=-3, b=2, p=3, q=4 (shown in figure 1). The distance heuristic gives the correct result; the weight heuristic gives the incorrect one.

Results: m(σ*w*) = 1.7 for all values of σ*w* that we tested.

Case 2: a=-3, b=1, p=3, q=4. Distance heuristic is correct; weight heuristic is incorrect.

Results:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| σ*w* | 0 | 1 | 5 | 10 |
| m(σ*w*) | 3.7 | 3.7 | 8 | 10 |

Case 3: a=-1, b=4, p=3, q=1. Distance heuristic is correct; weight heuristic is incorrect.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| σ*w* | 0 | 1 | 2 | 10 |
| m(σ*w*) | 1.2 | 1.7 | 2.9 | 10 |

Case 4: a=-2, b=4, p=3, q=1. Weight heuristic is correct; distance heuristic is incorrect.

|  |  |  |  |
| --- | --- | --- | --- |
| σ*x* | 0 | 1 | 2 |
| m(σ*x*) | 1.5 | 1.4 | 2 |

These results are actually quite unsurprising, given the structure of the model. Consider, for example, the case shown in figure 1 where a=-3, b=2, p=3, q=4. The left side of the balance beam will fall. Suppose that the weights are fixed and we allow a and b to vary. Then the left side of the balance beam falls as long as b < (3/4)|a|. Now suppose further that a is fixed at -3 and that b varies in a Gaussian around 2. If b > 2.25, then the procedure gets the wrong answer. However, the Gaussian is symmetric around b=2, so the probability that b > 2.25 is always less than 0.5. At a certain point, the truncation of the Gaussian at b=0 breaks the symmetry, but that can only be a significant effect when 0 is not much more than 1.5 standard deviation away; i.e. σ*x≈*1.5. (The analogous argument holds if we hold b fixed and allow a to vary in a Gaussian centered at 3, except that in this case the truncation of the Gaussian makes the correct answer *more* likely.)

**Positions of Favorite Lines in Poems**.

Griffiths and Tenenbaum conducted an experiment in which participants were told the position of a “favorite line” of poetry within a poem, and were asked to predict the length of the poem on that basis. As discussed in the main paper, they analyze the results of this experiment based on a model in which participants have prior knowledge of the distribution of lengths of poems; compute a Bayesian posterior using a stochastic in which the “favorite line” is uniformly distributed over poems and, within each poem, uniformly distributed within the lines of the poem; and then use as their prediction the median of the posterior distribution.

However, the second assumption above is demonstrably false. Favorite lines of poetry are not uniformly distributed over the length of a poem; they tend to be either the first or last line.

We will present the data for this claim in turn and compare the predictions based on the true distribution of favorite lines with the predictions given by Griffiths and Tenenbaum’s model and with the participant responses they report.

Data for this survey of favorite passages were taken from the collection "Life/Lines", created by the Academy of American Poets, URL <http://www.poets.org/page.php/prmID/339> downloaded September 28, 2012. The authors of this paper have saved this state of this collection for stable reference; this is available on request, if the web site changes or becomes unavailable.

From this corpus, we extracted a data set, enumerating the poem's author and title, starting line, ending line, and length of favorite passage; and length of the poem containing the passage. This data set is available on the website associated with the study: http://www.cs.nyu.edu/faculty/davise/ProbabilisticCognitiveModels/

Favorite passages are rarely a single line of poetry; the median length of the passages quoted is four lines.

Figure 3 below shows the probability that a line randomly chosen from within a favorite passage lies within the 1st, 2nd, ... 10th decile of the poem. As can be seen, the first and tenth decile are significantly more probable than the middle deciles.



Our model for using this data for predicting the length of a poem from the position of a “favorite line” is based on the following two assumptions:

* The Life/Lines corpus of favorite passages is representative of the general distribution of favorite passages within poems, as a function of the length of the poem.
* Since favorite passages in the corpus are rarely a simple line, we must somehow give an interpretation of the question being asked the participants in a form that can be applied to multi-line favorite passages. A seemingly reasonable model would be to use the following calculation: The hypothetical friend chooses her favorite passage of poetry, then picks a line randomly within that passage, and states the line number of that line. Participants who are told a value of L compute the conditional probability of the length of a poem, given that the value L has been produced following the procedure described above. They then answer as their prediction the median total length, relative to that distribution.

The data are noisy, and it is not clear how best to smooth them, but the results of the above procedure seem to be more or less as follows:

If L=1, predict 31.
If L is between 2 and 11, predict 12.
If L is between 12 and 200, predict L+2.

The odd result on L=1 reflects the fact that there are a number of long poems for which only the first line is quoted. This statistic is probably not very robust; however, it is likely to be true that the median prediction for L=1 is indeed significantly larger than for L=2, and may even be larger than the overall median length of poem (22.5).

The claim that, for L between 12 and 200, the prediction should L+2 is based on the following observation: There are 50 passages in which the position of the median line is greater than 12, and in 31 of those, the median line is within 2 of the end of the poem. Thus, given that the position L of the favorite line is between 12 and 200, the probability that the poem is no longer than L+2 is much greater than 0.5.

For L > 200, this rule probably breaks down. The last lines of very long poems do not tend to be particularly memorable.

**Comparison**

Table 2 shows the predictions of the two models and the mean participant responses at the five values L=2, 5, 12, 32, and 67 used by Griffiths and Tenenbaum in their experiment. We also show the predictions of the two models at the value L=1, since there is such a large discrepancy; Griffiths and Tenenbaum did not use this value in their experiment, so there is no value for participant responses. This comparison is shown graphically in Figure 3. The table and figure show the prediction of the number of lines that remain *after* the Lth line, which is the quantity actually being "predicted"; as discussed in the main paper, both of these models and any other reasonable model will predict that the poem has at least L lines.

The numerical data for the predictions of the Griffiths and Tenenbaum model and for the participant responses were generously provided by Tom Griffiths.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Model | L=1 | L=2 | L=5 | L=12 | L=32 | L=67 |
| Participant responses |  --- |  8 | 10 |  8 |  8 | 28 |
| Distribution of lengths | 10 |  9 |  9 |  4 |  11 | 19 |
| Favorite passages | 30 | 10 |  7 |  2 |  2 |  2 |

Table 1: Prediction of poem lengths



Figure 3: Prediction of poem lengths in alternative models. The solid lines show the predictions given by the two models. The circular dots are the participant responses.

**Frank and Goodman Communication Experiment: Additional Discussion**

In this section we discuss the derivation of the various predictions shown in Figure 4 of the main paper. We also discuss an additional possible variation of the analysis, mentioned in footnote 6 of the main paper.

Throughout this section, we will be concerned with the following task, which in Frank and Goodman's experiment was presented to participants in the "listener" condition. Participants are shown a diagram displaying, from left to right, a blue square, a blue circle, and a green square. They are told that a speaker, looking at the same diagram, has used the word "blue" to identify one of the objects, and are asked to place bets on which object he intended.

Frank and Goodman propose the following general structure for the cognitive model of the listener; we will follow this general structure in all of our models. There is a prior distribution P(object) corresponding to the inherent likelihood that the speaker would choose one or another object as a participant of discourse. There is a likelihood P(word|object) that, having chosen a given object, the speaker will decide to use a particular word. The listener then uses Bayes' law to compute a probability P(object|word). He then uses a decision rule to place his bets.

**Frank and Goodman's model**

Frank and Goodman instantiated the various pieces of this model as follows:

The priors corresponded to the salience of the objects, and were determined by asking participants in the "salience" condition which object they would want to refer to. The values obtained were P(left)=0.2, P(middle)=0.4, P(right)=0.4. We will use these same priors in all our models.

The likelihoods were determined in two ways, one experimental and one theoretical, which gave results in close agreement. The experimental method was to ask participants in the "speaker" condition to place bets on whether they themselves would use "blue" or "square" in referring to the left object; or "blue" or "circle" in referring to the middle object; or "green" or "square" in referring to the right object. The theoretical method was to assume that, for any given object O, the probability that a speaker will use word W to refer to O is proportional to the utility of that choices, which is proportional to the specificity of W for O, where the specificity is the reciprocal of the number of objects in the scenario that W describe. The words "blue" and "square" each refer to two objects; the words "green" and "circle" each refer to one. Therefore the specificity of "blue" for the left or middle object is 1/2; the specificity of "square" for the left or right object is 1/2; the specificity of "circle" for the middle object and of "green" for the right object is 1. Therefore the likelihoods for "blue" are: P("blue"|left) = 0.5/(0.5+0.5) = 0.5; P("blue"|middle)=0.5/(1+0.5) = 0.333. Of course P("blue"|right) = 0.

Applying Bayes' law, the posterior probabilities, given that the speaker has said "blue", are

P(left|"blue") = P("blue"|left)\*P(left) / (P("blue"|left)\*P(left) + P("blue"|middle)\*P(middle)) = (0.5\*0.2)/(0.5\*0.2+0.333\*0.4) = 0.43

P(middle|"blue") = P("blue"|middle)\*P(middle) / (P("blue"|left)\*P(left) + P("blue"|middle)\*P(middle)) = (0.333\*0.4)/(0.5\*0.2+0.333\*0.4) = 0.57

Finally, Frank and Goodman assumed that participants would split their bets proportional to the posterior probability; thus they split their bets 43:57:0. This agreed well with the their experimental data.

**Alternative decision rules**

If participants' reasoning is carried out using the rule that agents follow the maximum expected utility (MEU) rule, then the results of the reasoning are very different. The MEU rule states that an agent should always choose the action with the greatest expected utility; it is universally considered normatively correct.[[1]](#footnote-1) Either placing bets proportionally to the probability of an outcome, or choosing actions with a probability proportional to their expected utility, is a suboptimal strategy. Frank and Goodman's calculation of the likelihoods, above, is based on the assumption that the listener assumes that the speaker is following the second of these suboptimal strategy, choosing with probability 2/3 to use "circle" to refer to the blue circle, rather than doing so with probability 1, as the MEU rule would prescribe. Their calculation of the bets placed by the listeners assumes that the listeners themselves are following the first of these suboptimal strategies, hedging their bets, rather than placing all their money on the more likely outcome, which is what the MEU rule would prescribe.

There are thus four possibilities:

1. The listener believes that the speaker will choose his word proportionally, and the hearer splits his bet. This is the model adopted by Frank and Goldman.
2. The hearer believes that the speaker will choose his word proportionally, and the hearer, following MEU, places all his money on the more likely posterior probability.
3. The hearer believes that the speaker will definitely choose the more informative word, following MEU, but the hearer himself splits his bet.
4. The hearer believes that the speaker will follow MEU, and he himself follows MEU. In a theory that posits that agents are rational, this is surely the most appropriate of the options.

Possibilities (3) and (4) actually have the same outcome. If the speaker follows MEU, then the probability that he will refer to the middle object as "blue" is 0, since he has the option of referring to it as "circle". Hence, applying Bayes' rule,

P(left|"blue") = P("blue"|left)\*P(left) / (P("blue"|left)\*P(left) + P("blue"|middle)\*P(middle)) = (0.5\*0.2)/(0.5\*0.2+0\*0.4) = 1.0

P(middle|"blue") = P("blue"|middle)\*P(middle) / (P("blue"|left)\*P(left) + P("blue"|middle)\*P(middle)) = (0\*0.4)/(0.5\*0.2+0\*0.4) = 0

Since there is no probability that the speaker will describe the middle object as "blue", if he uses the word "blue", it is entirely certain that he means the left object. Therefore, whether the listener is following MEU or hedging his bets, he should put all his money on the left object.

In possibility (2), the posterior probabilities will be the same as in Frank and Goodman's calculation above: P(left|"blue") = 0.43, P(middle|"blue") = 0.57. However, the listener, following MEU, should place all his money on the more likely outcome, which is the middle object.

**Choice of words**

All of the above calculations are based on the assumption that the listener is aware that the only four words available for the speaker are "blue", "green", "square", and "circle". However, this information is not given to the listener in the experimental setup described, and it is by no means clear how he comes to this awareness. There are certainly other single words that could be used by the speaker; either synonyms or near synonyms for these such as "turquoise", "emerald", "rectangular", and "round"; words referring to positions such as "left", "middle" and "right"; and other words, such as “shape”, “picture”, or “salient”. If the space of words were increased then that would certainly change the theoretical calculation of the specificity of these words, and it might well change the behavior of the participants in the "speaker" condition.

Suppose, for instance, that the space of words under consideration is expanded to include "round", which has specificity 1 with respect to the middle object. If we follow Frank and Goodman's model in assuming that both speaker and hearer follow suboptimal decision rules, the calculation proceeds as follows: The likelihood P("blue"|left) = 0.5 as before. The likelihood P("blue"|middle) = 0.5/(0.5+1+1) = 0.2. Applying Bayes' law, we obtain the posteriors

P(left|"blue") = P("blue"|left)\*P(left) / (P("blue"|left)\*P(left) + P("blue"|middle)\*P(middle)) = (0.5\*0.2)/(0.5\*0.2+0.2\*0.4) = .55

P(middle|"blue") = P("blue"|middle)\*P(middle) / (P("blue"|left)\*P(left) + P("blue"|middle)\*P(middle)) = (0.2\*0.4)/(0.5\*0.2+0.2\*0.4) = .45

The two posterior probabilities have thus switched places. Now, of course, the decision just to add the word "round" was entirely arbitrary, but so was the original set of words.[[2]](#footnote-2) The reader may enjoy calculating the effect of this extension on the alternative models.

**Gweon, Tenenbaum, and Schulz experiment on infants' reasoning about samples**

The paper discussed in this section came to our notice too late to include a discussion in the main paper; however since it vividly illustrates the kinds of arbitrary assumptions that can be made in the application of Bayesian models, we have included it here, in this supplement.

Gweon, Tenenbaum, and Schulz (2011) carried out the following experiment: 15 month old infants were shown a box containing blue balls and yellow balls. In one condition of the experiment, 3/4 of the balls were blue; in the other condition, 3/4 were yellow. In both conditions, the experimenter took out three blue balls in sequence, and demonstrated that all three balls squeaked when squeezed. The experimenter then took out an inert yellow ball, and handed it to the baby. The experimental finding was that, in condition 1, 80% of the babies squeeze the yellow ball to try to make it squeak, whereas in condition 2, only 33% of the babies squeeze the ball.

The explanation of this finding given by Gweon, Tenenbaum, and Schulz is as follows. The babies are considering two possible hypotheses about the relation of color to squeakiness: Hypothesis A is that all balls squeak; hypothesis B is that all and only blue balls squeak; (the obvious third alternative that only yellow balls squeak is ruled out by the observation, and therefore can be ignored). Thus if A is true, then the yellow ball will squeak; if B is true, it will not. The babies are also considering two possible hypotheses about the experimenter's selection rule for the first three balls. Hypothesis C is that the experimenter is picking at random from the set of all balls; hypothesis D is that she is picking at random from the set of all balls that squeak,

The intuitive explanation then, is this. If most of the balls in the box are blue, then the fact that the experimenter has pulled three blue balls is unsurprising, and thus reveals nothing about her selection strategy. If most of the balls are yellow, however, then that suggests that she is pursuing a particular selection strategy to extract the three blue balls. That selection strategy must be selecting balls that squeak, since that is the only selection strategy under consideration. Then the fact that selecting balls that squeak gives all blue balls suggests that only the blue balls squeak.

We can formalize this argument as follows: Let z be the fraction of blue balls in the box. Let p=Prob(A), q=Prob(C), so Prob(B) = 1―p and Prob(D) = 1―q. Assume that A vs. B is independent of C vs. D and that they are both independent of z. Let S be the event that the experimenter selects three blue balls (note that under either hypotheses A or B, the blue balls will all squeak). Then the baby can calculate the probability of A given the observation of the three squeaky blue balls as follows.

Prob(S|A,z) = z3

Prob(S|B,C,z) = z3

Prob(S|B,D,z) = 1

 Therefore

 Prob(S|z) = pz3 + (1―p)qz3 +(1―p)(1―q)

Prob(A,S|z) = pz3

So Prob(A|S,z) = Prob(A,S)/Prob(S) = pz3 / (pz3 + (1―p)qz3 +(1―p)(1―q))

It is easily seen that this is an increasing function of z; for instance as z goes to 0, this fraction goes to 0; as z goes to 1, the denominator approaches 1, so the fraction converges to p. If we choose p=q=1/2, then Prob(A|S,z=3/4) = 0.37, whereas Prob(A|S,z=1/4) = 0.03.

Thus, if most of the balls in the box are blue (z=3/4), there is a reasonable chance (though less than ½) that all balls squeak; if most of the balls in the box are yellow (z=1/4), it is extremely likely that only blue balls squeak.

There are many difficulties here. The most obvious is that the choice of strategies is limited to C and D. Another obvious alternative is (E) that the experimenter is deliberately selecting only blue balls. *A priori* this would seem more likely than D, since the experimenter can easily see which balls are blue, but cannot see which balls squeak, If one adds this to the model as a third possibility, then Prob (A|S,z) is still an increasing function of z, but the dependence is much weaker.

We can redo the calculation as follows. Let q=Prob(C), r=Prob(D), so Prob(E) = 1―(q+r). We have now

Prob(S|A,C) = z3

Prob(S|A,D) = z3

Prob(S|A,E) = 1

Prob(S|B,C) = z3

Prob(S|B,D) = 1

Prob(S|B,E) = 1

 Therefore

 Prob(S) = (q+pr)z3 + 1 – q - pr

Prob(A,S) = p(q+r)z3 + p(1―(q+r))

So Prob(A|S) = Prob(A,S)/Prob(S) = (p(q+r)z3 + p(1―(q+r)))/ ((q+pr)z3 + (1―p)r + (1―(q+r)))

If we choose p=1/2, q=r=1/3 then the results are

Prob(A|S,z=3/4) = 0.43

Prob(A|S,z=1/4) = 0.34

The babies are thus making a quite strong distinction on the basis of a fairly small difference in posterior probability.

If we take into account that D is inherently less plausible than E, then the difference becomes even smaller. Suppose for instance that C and E are equally plausible and that D is half as plausible as either; thus q=2/5, r = 1/5. Then

Prob(A|S,z=3/4) = 0.46

Prob(A|S,z=1/4) = 0.40

As r goes to zero, the dependence of the posterior probability of A on z also vanishes.

Another problem is the assumption that the selection strategy applies only to the first three blue balls. If the strategy is "Select balls that squeak" and the experimenter is pursuing this strategy in choosing the yellow ball as well, then the probability is 1 that the yellow ball will squeak. It is hard to see why the baby should assign probability 0 to this interpretation of what the experimenter is doing. Gweon, Tenenbaum, and Schulz address this possibility in a footnote as follows:

The model mirrors the task design in distinguishing the sampling phase from the test phase. Because the yellow ball was treated differently from the blue ball(s) (i.e. given directly to the children and not manipulated by the experimenter), we do not treat the yellow ball as part of the sample in the model.

But how do they know that the baby takes the same view of the matter? Alternatively, the baby might reason that, since the experimenter didn't bother to squeeze the yellow ball, it must not squeak.

One can also question the independence of A and B from z. An optimistic baby might reason that, after all, a box *should* contain a lot of squeaky balls; therefore if a box contains a lot of yellow balls, that in itself suggests that yellow balls should squeak. A pessimistic baby might make the reverse argument.

 Another difficulty is that equating the fraction of babies who think that the ball will squeak with the probability that any given baby thinks that the ball will squeak is hard to justify. Each individual baby, after all, either does or does not squeeze the ball; if babies are all perfectly rational, all have the same priors, and the same decision rule (e.g. squeeze the ball if the *a posteriori* probability that it will squeak is greater than 1/3) then under any given circumstance, they should either all squeeze the ball or all not squeeze the ball. There are a number of possible answers here; they may have different priors on the hypotheses, or they may assign different ratios of the utility of getting a squeak vs. the disutility of making the effort to squeeze the ball. But even so, the translation of the probability into cross-participant action frequency is tricky.

Gergely and Jacob (2013) propose an alternative explanation of these results, resting on the idea, which has independent experimental evidence, that infants distinguish between instrumental and communicative actions in people who are interacting with them (in this case the experimenter). However, their analysis relies on the same anomalous choice of the selection criterion under consideration, and therefore is subject to the same critique.

**References**

Gergely, G. & Jacob, P. (2013). Reasoning about instrumental and communicative agency in human infancy. In F. Xu and T. Kushnir (eds.) Rational Constructivism in Cognitive Development, Amsterdam:Elsevier, 59-94.

Jansen, B. R. J., & van der Maas, H. L. J. (1997). Statistical test of the rule assessment methodology by

latent class analysis. *Developmental Review*, 17, 321–357.

Russell, S. and Norvig, P. (2010) *Artificial Intelligence: A Modern Approach*. 3rd edn., Prentice Hall.

1. The normatively proper method of using bets to elicit subjective probabilities, following de Finetti, is to ask participants at what odds they would take one side or another of a bet (see (Russell and Norvig 2010) 489-490). [↑](#footnote-ref-1)
2. One principled approach to this issue would be to use all possible adjectives and nouns that could apply to these objects among the thousand most common words in English. In that case, the question would arise how the frequency of these words should be combined with the specificity measures proposed by Frank and Goodman in computing the overall prior. We have not carried out this calculation. [↑](#footnote-ref-2)