Supplement to Informed Guessing in Change Detection

Stephen Rhodes, Nelson Cowan, Kyle O. Hardman, and Robert H. Logie

August 15, 2017

Contents

1 Additional Models

The informed guessing model reported in the main manuscript assumes that participants compute the posterior probability of a change having occurred given their state of uncertainty (i.e. all items in memory match whole display items or a match has not been found to a single central probe) and then guess in accordance with that probability (g). That is, for a posterior probability of a change of 70% the observer is predicted to guess *change* on 7/10 trials where they are uncertain. The optimal behavior would be to select the most likely option every time, that is to guess *change* for every trial where the posterior probability exceeds 50%. In probability learning experiments, given enough time, the majority of participants exhibit this kind of optimal response selection (Shanks, Tunney, & McCarthy, 2002). As discussed in the main manuscript, there are reasons to believe that observers may be responding somewhat more optimally than the basic informed guessing model allows (although other possibilities are also discussed in the main manuscript). We assessed this with two additional models.

1.1 Mixture

To allow for more optimal guessing behavior than that encompassed in the informed guessing models reported in the main manuscript, we fit an additional model with a mixture parameter, P^{OG}, which determined the probability of responding optimally versus probability matching. Optimal responding takes the following form;

$$\gamma = \begin{cases} 1, & \text{if } g > 0.5 \\ 0.5, & \text{if } g = 0.5 \\ 0, & \text{if } g < 0.5. \end{cases}$$

Guessing responses were assumed to be optimal with probability P^{OG} and probability matching with probability $1 - P^{OG}$. Thus, guessing behavior, determined by ψ in the model descriptions given in the Appendix, is given by the mixture of the two forms;

$$P_i^{OG} \gamma_{ijlm} + (1 - P_i^{OG}) g_{ijlm}.$$

The other aspects of the model were the same as that described in the Appendix of the main manuscript. The same hierarchical priors were placed on the mean and standard deviation parameters from which participant values of k, a, and u were sampled. In addition, individual (logit transformed) probabilities of optimal information use, P_i^{OG}, were sampled from a normal distribution with the following hierarchical priors;

$$\mu^{(P^{OG})} \sim \text{Normal}(0, 10^2)$$
$$\sigma^{(P^{OG})} \sim \text{Gamma}(1.01005, 0.1005012).$$

1.2 Logistic Rule

Another way to incorporate more optimal information use than that implied in the basic informed guessing model is via the logistic rule (see, e.g., Friedman & Massaro, 1998; Shanks et al., 2002). Here the observer guesses *change* with the probability;

$$o = \frac{1}{1 + e^{-\operatorname{logit}(g)\lambda}},$$

where λ determines the extent to which the observers' guessing behavior is optimal. As $\lambda \to 0$, guessing depends less and less on information in memory, or d ($o \to 0.5$); when $\lambda = 1$ the observer is probability matching ($o = g$) and as $\lambda \to \infty$, behavior becomes increasingly optimal. Figure 1 demonstrates the relationship between the probability of detection, d, and o for various values of λ in the whole display and single probe formulations of the change detection task. As described in the main

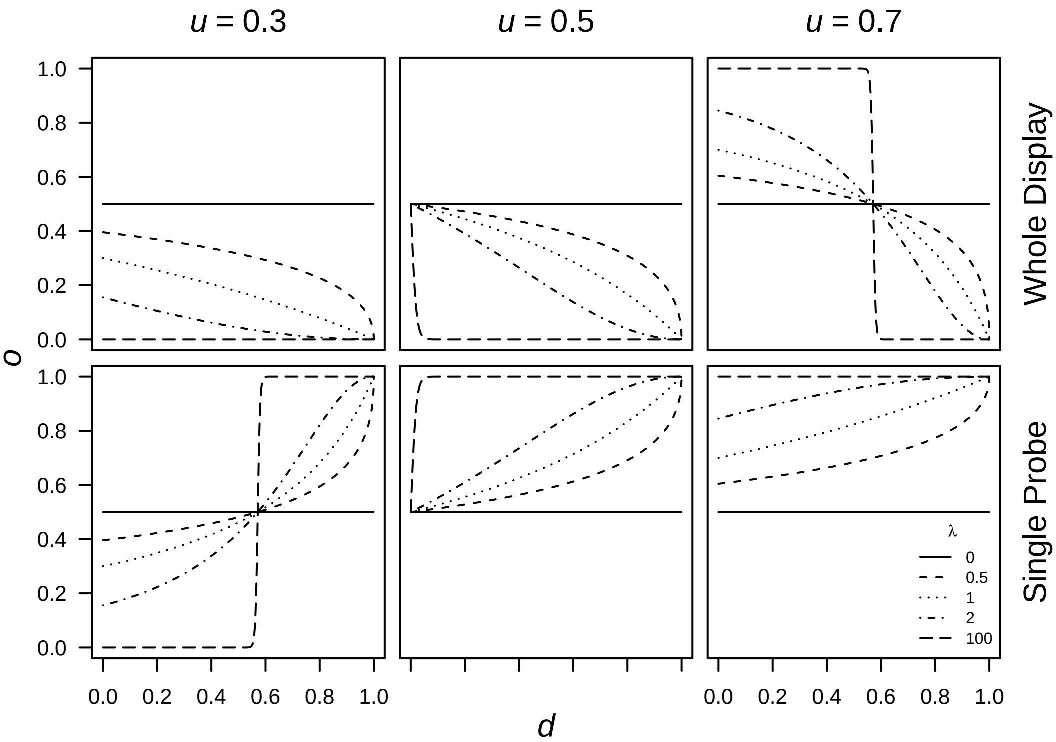

Figure 1: Values of guessing, o, for the single central probe and whole display ($C = 1$) change detection tasks according to the logistic rule as a function of d for different values of λ across various base-expectations, u.

manuscript, more information in working memory (i.e. a greater d) decreases the likelihood of guessing *change* with a whole display and has the opposite effect with a single central probe. The effect of λ is to increase the precision/ optimality with which g is translated into guessing behavior. As shown in Figure 1, a large λ introduces a threshold where guessing behavior gravitates to the most likely response option.

As λ is constrained to be positive, log transformed values of this parameter for each observer were sampled from a normal distribution with the following vague priors on its mean and standard deviation:

$$\mu^{(\lambda)} \sim \text{Normal}(0, 10^2)$$
$$\sigma^{(\lambda)} \sim \text{Gamma}(1.01005, 0.1005012).$$

1.3 Results

Table 1 presents posterior quantities for the hierarchical parameters of these two models, along with their DIC and WAIC values, for Experiments 1–4. Comparison of the DIC values in this table with those in Table 1 of the main manuscript reveals that these variable guessing models generally outperformed the (probability matching) informed guessing model. However, as discussed in the manuscript, other possibilities remain tog be ruled out (e.g. a systematic tendency for observers to overestimate the number of items in memory). Nevertheless, the population-level parameters of the mixture model suggest that participants guessed optimally on 28, 50, 24, and 11% of trials in Experiments 1–4, respectively. Table 1 shows, however, that there is a great deal of variability in these population parameters (zero is a credible value for P^{OG} in Experiments 3 and 4). For the logistic rule model estimates of the mean λ parameter all fall above 1 (the 95% HDIs also exclude 1, see Table 1), suggesting that in all 4 experiments responding was very slightly more optimal than that predicted by probability matching ($\lambda = 1$).

In line with the findings of Hardman and Cowan (2016), there is a great deal of variability between participants in the extent to which information is utilized in the change detection task. Figure 2 depicts this variability. The left panels of this figure present individual estimates of the P^{OG} parameter, transformed back to their natural scale. The values of this mixture parameter are, for the most part, below 0.5, suggesting that the majority of participants responded optimally less than 50% of the time. The right panel presents individual λ estimates from the logistic model, which show a similar picture. For the most part the posterior medians fall above 1, and 9, 10, 13, and 9 participants credibly fall above 1 for Experiments 1–4, respectively (that is, their 95% HDIs exclude 1).

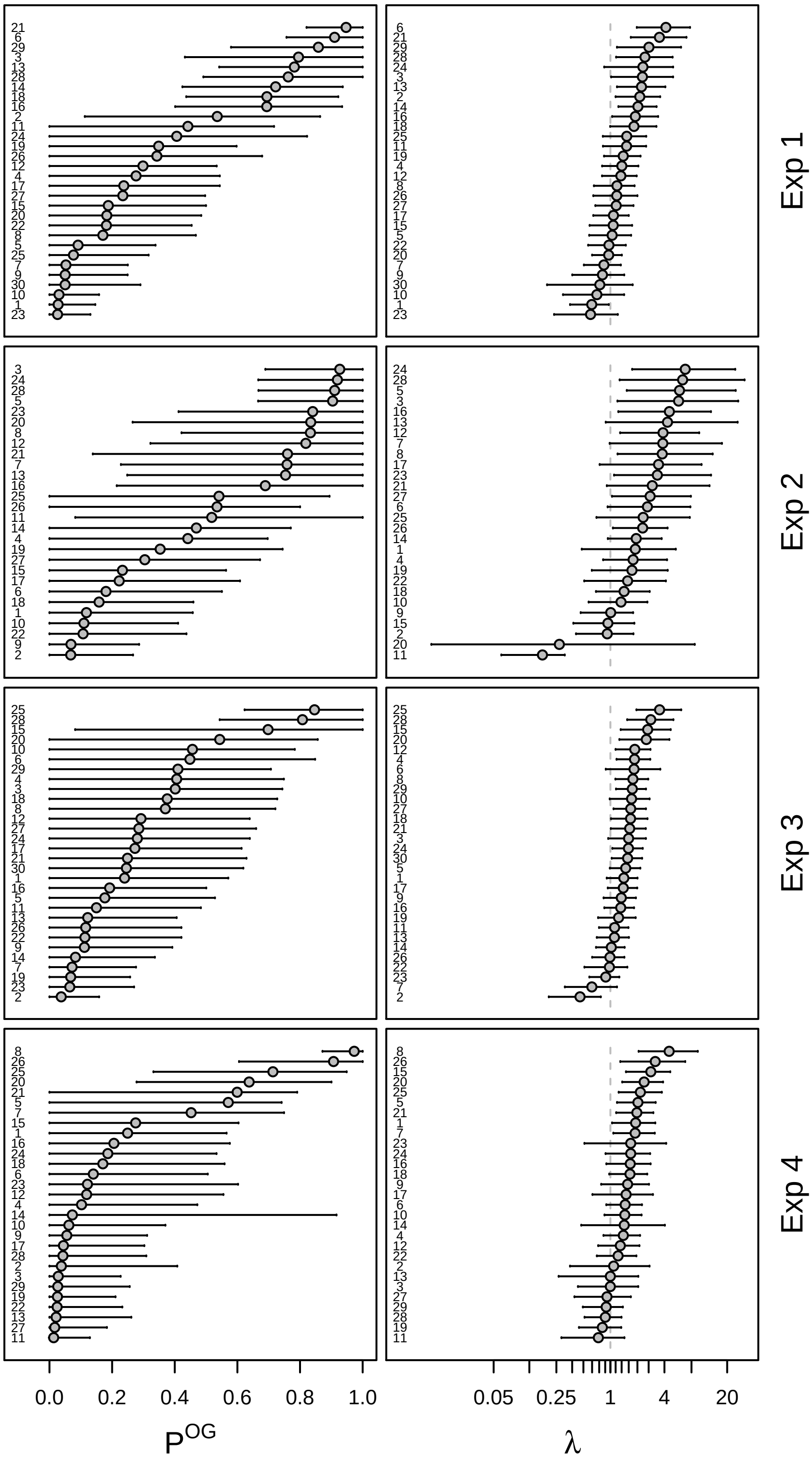

Figure 2: Dotplots (median and 95% HDI) of individual participant parameter estimates that determine the extent to which guessing is informed in the mixture and logistic rule models for Experiments 1–4. Participant parameters were sampled from hierarchical distributions as described above and in the main manuscript. See text for description of the parameters, λ is plotted on a log scale, numbers along the left refer to the participant in a given experiment with points ordered from lowest to highest.

Table 1: Table of parameter estimates (posterior medians and 95% HDIs) of population level mean parameters for the extra informed guessing models for Experiments 1-4. p_{DIC} and p_{WAIC} are estimates of effective number of parameters for DIC and WAIC, respectively.

Exp	Model	k	a	$u_{0.3}$	$u_{0.5}$	$u_{0.7}$	P^{OG} or λ	DIC (p_{DIC})	WAIC (p_{WAIC})
1	Mixture	3.57 [3.35, 3.78]	0.91 [0.88, 0.95]	0.35 [0.30, 0.41]	0.47 [0.41, 0.53]	0.55 [0.50, 0.60]	0.29 [0.04, 0.52]	2241 (162)	2199 (102)
	Logistic Rule	3.53 [3.27, 3.77]	0.91 [0.88, 0.95]	0.33 [0.27, 0.39]	0.42 [0.36, 0.48]	0.50 [0.45, 0.55]	1.41 [1.02, 1.88]	2242 (155)	2221 (112)
2	Mixture	3.02 [2.71, 3.32]	0.89 [0.84, 0.93]	-	0.45 [0.38, 0.52]	-	0.49 [0.13, 0.77]	2007 (92)	2026 (95)
	Logistic Rule	2.85 [2.53, 3.18]	0.90 [0.86, 0.93]	-	0.47 [0.41, 0.52]	-	2.29 [1.21, 3.95]	1991 (116)	1985 (92)
3	Mixture	3.69 [3.43, 3.96]	0.93 [0.91, 0.96]	0.57 [0.49, 0.65]	0.65 [0.57, 0.72]	0.69 [0.62, 0.76]	0.23 [0.00, 0.44]	2181 (155)	2157 (110)
	Logistic Rule	3.64 [3.35, 3.91]	0.93 [0.91, 0.96]	0.56 [0.51, 0.62]	0.62 [0.56, 0.68]	0.66 [0.59, 0.72]	1.46 [1.08, 1.93]	2213 (181)	2155 (105)
4	Mixture	3.15 [2.73, 3.52]	0.93 [0.87, 0.96]	0.48 [0.41, 0.54]	0.61 [0.54, 0.68]	0.66 [0.60, 0.72]	0.11 [0.00, 0.34]	2237 (145)	2252 (132)
	Logistic Rule	3.15 [2.74, 3.49]	0.92 [0.87, 0.96]	0.49 [0.44, 0.54]	0.59 [0.53, 0.65]	0.63 [0.57, 0.70]	1.53 [1.09, 2.06]	2237 (158)	2231 (126)

2 Conventional Estimation of k

Pashler (1988) derived a principled formula for k in the whole display task. The underlying rationale is outlined in the main manuscript for the extended versions of this model, but the general logic is briefly reiterated here. The observer correctly detects a change when the changed item mis-matches an item in memory or, alternatively, participants may guess, giving a probability of making a hit; $h = d + (1 - d)g$, where the nature of g is left undefined. Pashler (1988) used a large array, well beyond the assumed scope of working memory, therefore if a participant makes a false-alarm it must be because they guessed so $f = g$. If $k \leq N$, however, $f = 0$ (Rouder, Morey, Morey, & Cowan, 2011). This logic results in a closed form estimator for k;

$$\hat{k} = N\left(\frac{\hat{h} - \hat{f}}{1 - \hat{f}}\right).$$

With a single probe item presented in an unoccupied study location, Cowan, Blume, and Saults (2013) note that the rationale underlying the Pashler formula is reversed. That is, participants don't detect change, but rather a match if the relevant information is in working memory; $1 - f = d + (1 - d)g$. However, a miss can only arise due to guessing, so; $1 - h = 1 - g$ (for $k \leq N$). The formula for this task is,

$$\hat{k} = N\left(\frac{\hat{h} - \hat{f}}{\hat{h}}\right).$$

These simplified models are appealing as one enters the estimated hit and false-alarm rates and receives an estimate of the (average) number of items the observer could retain. However, they suffer a crucial misspecification in that they assume that the observer attends fully on 100% of trials (see Rouder et al., 2008). As the presence of any errors in responding suggests that $k < N$ this can lead to small capacity estimates from small set sizes when rare errors occur. As discussed in the main manuscript, a reasonable approach to accounting for this (after Rouder et al., 2008) is to add a 'lapse' parameter, which determines the probability that the observer pays attention on a given trial. The introduction of this lapse parameter is crucial as it then requires specification of what participants do when they are uncertain; that is, what they do when they guess. This is because the lapse parameter determines the extent to which observers can use information to inform guessing, given that, logically, guessing cannot be informed when the observer has lapsed.

Given the evidence presented in the main manuscript (and in the supplementary models above)—that observers appear to use knowledge of the number of items in memory to update their base-expectation of a change occurring—we may expect

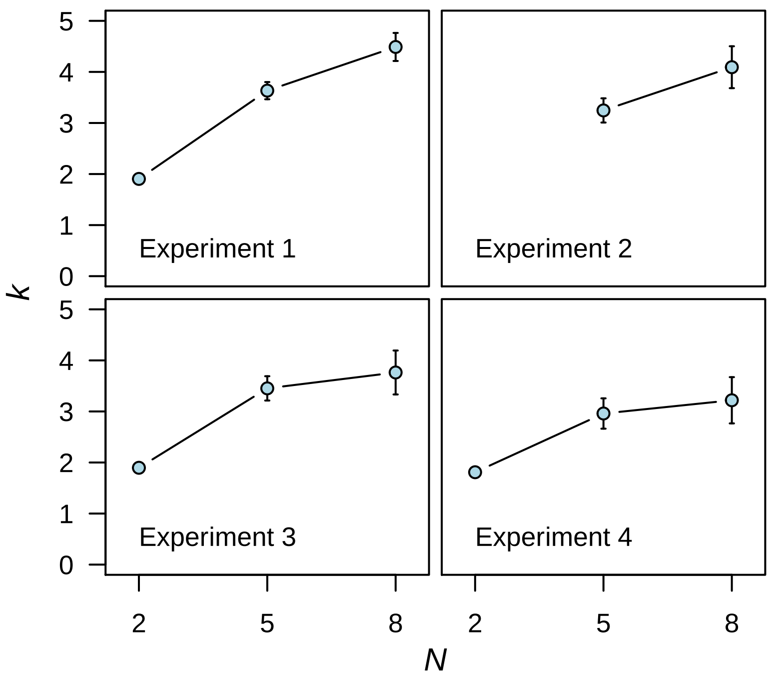

Figure 3: Estimates of k from simple processing models of Pashler (1988) and Cowan et al. (2013). Estimates for Experiment 2 are from trials with 1 possible change. Error bars are $\pm$ 2 $\times$ standard error.

the estimators above to give misleading answers regarding experimental effects. We applied them to aggregated hit and false-alarm rates from Experiments 1–4. The estimated ks are presented in Figure 3. The Pashler formula assumes that change trials introduce a single change, so for Experiment 2 we only used this formula for the $C = 1$ condition. In addition, rather than solve for each number of possible changes used in Experiment 2, we used the `optim` function in `R`, started at different random values, to search for maximum likelihood estimates of k with the full data set and this basic model along with the formula for d given in the main manuscript (see Appendix).

In all cases ANOVA resulted in significant main effects of set size on k. In the set size 2 condition, estimates of k are limited to 2 so it's possible that including this condition drives the main effect. However, ignoring the set size 2 condition, and focusing on conditions where a good deal of responses are assumed to be based on guessing, the main effect p-values remain under 0.05 for Experiments 1 and 3 ($p = 0.06$ in Experiment 4), indicating that the main effect does not solely rely on the small set size condition for these experiments. Along with the evidence presented in the main manuscript—that variable k versions of the uninformed guessing model provide better fit than the fixed k version, whereas for the informed model the fixed k version outperforms the variable k version—this suggests that failure to correctly account for the use of information in the change detection task can lead to misleading results.

Data and code to reproduce all of the analyses described above can be found here: `https://github.com/stephenrho/Guessing`

References

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 731–747.

Friedman, D., & Massaro, D. W. (1998). Understanding variability in binary and continuous choice. *Psychonomic Bulletin & Review*, *5*(3), 370–389.

Hardman, K. O., & Cowan, N. (2016). Reasoning and memory: People make varied use of the information available in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(5), 700–722.

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics*, *44*(4), 369–378.

Rouder, J. N., Morey, R. D., Cowan, N., Zwilling, C. E., Morey, C. C., & Pratte, M. S. (2008). An assessment of fixed-capacity models of visual working memory. *Proceedings of the National Academy of Sciences*, *105*(16), 5975–5979.

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin and Review*, *18*, 324–330.

Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233–250.