

Appendix B: Model specification

The model used in this paper is a simplified version of the one described by Hayes et al (2019; see their Appendix B). In this appendix we present a formal specification of the model (Cooper & Guest 2014), along with some comments on our implementation in R. The motivation for the model is to consider how one might learn an arbitrary smooth function $\phi : \mathcal{X} \rightarrow [0, 1]$ that assigns to every possible stimulus $x \in \mathcal{X}$ some subjective belief that x possesses the unknown property. In the general case the stimulus space \mathcal{X} can be richly structured, but in the current setting we can consider stimuli that vary only on a single continuous dimension.

The prior distribution

We specify a prior distribution over smooth functions in two parts. First, it is convenient to define a logit-transformed version of the function $f : \mathbb{R} \rightarrow \mathbb{R}$, and define the prior on this transformed space:

$$f(x) = \log \frac{\phi(x)}{1 - \phi(x)}$$

After adopting this representation it is relatively simple to use the Gaussian process (see Rasmussen & Williams 2006) to specify a prior over possible functions f . In particular, although the Gaussian process is used to sample continuous functions f , it has the property that every finite dimensional marginalisation yields a multivariate Gaussian distribution over that set of values. What this means is that because we are interested in the values of f only at those stimulus locations that are used in the experiment, we can work directly with this marginal prior and sample values only at those n points (where $n = 6$ in this case). Letting $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)$ denote the mean vector at these points and letting $\boldsymbol{\Sigma} = [\sigma_{ij}]$ for $i, j \in 1, \dots, n$ define the covariance matrix, this marginal prior may be written:

$$f(x_i) \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

The covariance matrix is used to control the smoothness of the inferred function f and defined by the kernel function $\sigma_{ij} = K(x_i, x_j)$. Following the logic outlined by Hayes et

This note specifies the content of Appendix B to Wen et al (in preparation), and is intended to be read as part of that paper. It was written by Danielle Navarro (d.navarro@unsw.edu.au), and prepared using the papaja package (Aust & Barth 2020) for R markdown (Xie, Allaire & Golemund 2019).

al, we apply a radial basis kernel function kernel in which the correlation diminishes as a Gaussian function of the distance d_{ij} between items in psychological space:

$$K(x_i, x_j) = \tau^2 \exp(-\rho d_{ij}^2)$$

and

$$\sigma_{ij} = \begin{cases} K(x_i, x_j) & \text{if } i \neq j \\ K(x_i, x_i) + \sigma^2 & \text{if } i = j \end{cases}$$

In summary, the Gaussian process prior is described by four parameters, which have the following interpretation:

- μ is the prior mean/centrality parameter (for ϕ), where $\mu_i = \mu$ for all i
- τ is a smoothness parameter and controls the baseline level of smoothness
- ρ is a smoothness parameter and controls how quickly the correlation decays with psychological distance
- σ is a prior variance parameter used to specify the extent to which the function can deviate from μ

In practice, μ and *sigma* together play the role of “pulling” the function toward some value. The intended usage here is to impose a “regress toward zero” bias on the function: the real world is sparse, and if we sampled an arbitrary property and stimulus of psychological relevance it would be more likely that the stimulus doesn’t possess that property: e.g., sample a random fruit (apple, orange, banana, etc) and some random colours (red, orange, black, etc). There are many combinations of these values that lead to property-negative statements (apples are not black, oranges are not green, etc) and only a few that are true (oranges are orange). For more discussion of sparsity biases see Navarro and Perfors (2011).

By way of comparison, τ and ρ are the parameters defining the kernel function K , and together they impose a bias towards smoothness. Stimuli that are near each other in psychological space should have similar inductive properties (Shepard 1987) and these two parameters work together to impose this smoothness.

The likelihood function

The Gaussian process prior supplies a mechanism for a Bayesian reasoner to impose two qualitatively-reasonable biases: smoothness and sparsity. What it does not provide is a learning method: how should the learner revise their prior beliefs when they observe stimuli and their properties? That role is played by the likelihood function, which describes the learners beliefs about how these observations were sampled. Consider a participant who has observed stimulus x that is property-positive, and who understands the statistical implications of the sampling frame manipulation used in the experiment, and adopts a likelihood function that derived from a consideration of this sampling process:

Under **property sampling** the learner assumes that this item was chosen uniformly at random from the set of property-positive stimuli. If ϕ_i denotes the probability (or proportion) of x_i items (e.g., rocks of this specific size) that are plaxium positive, and

62 (for simplicity) assume that the base rates of all stimulus types are the same, then the
 63 probability that a random plaxium-positive stimulus will turn out to have value x_i is

$$P(x_i|\phi) = \frac{\phi_i}{\sum_j \phi_j}$$

64 where the sum in the denominator is taken across all possible stimuli (for these designs
 65 that’s just the six test items).

66 Under **category sampling** the items are sampled uniformly at random from items
 67 that belong to the category (in this case, the category consists of 2 of the 6 test items only).
 68 If there are $|c|$ items in the category that have the same base rates, then the probability
 69 of the stimulus value x_i is $1/|c|$. The probability that the item is also property positive
 70 depends on the value of ϕ_i and so

$$P(x_i|\phi) = \frac{\phi_i}{|c|}$$

71 Accordingly, the model implies that participants will apply a different likelihood function
 72 depending on the sampling condition, though it should be noted that this is a model *only*
 73 of participants who are sensitive to the sampling manipulation. It is useful for the purposes
 74 of identifying what specific pattern we should expect from someone who is sensitive to
 75 this manipulation (rather than making a generic suggestion that “people do something
 76 different in the two conditions”), allowing us to look for the theoretically relevant pattern
 77 of behaviour rather than test arbitrary terms in a linear model.

78 Computing generalisation gradients

79 Suppose the learner has seen a set of stimulus items \mathbf{x} that are property-positive. To
 80 complete the model specification, we describe how the generalisation function emerges from
 81 the Bayesian model implied by a given prior $P(\phi_h)$ over possible functions, and a likelihood
 82 $P(\mathbf{x}|\phi_h)$ that describes the probability of the learner observing \mathbf{x} if ϕ_h is the true function.
 83 Applying Bayes’ rule gives the posterior distribution

$$P(\phi_h|\mathbf{x}) = \frac{P(\mathbf{x}|\phi_h)P(\phi_h)}{\int P(\mathbf{x}|\phi_h)P(\phi_h) d\phi_h'}$$

84 The underlying hypothesis space over possible functions $\phi \in \mathcal{H}$ is continuous so the denom-
 85 inator integrates over all possible functions. In our implementation, we approximate the
 86 full hypothesis space \mathcal{H} of possible functions by considering only a finite set of possibilities,
 87 so the inference is written:

$$P(\phi_h|\mathbf{x}) \approx \frac{P(\mathbf{x}|\phi_h)P(\phi_h)}{\sum_{h'} P(\mathbf{x}|\phi_h)P(\phi_h)}$$

88 This application of Bayes rule provides a posterior distribution over possible hypotheses
 89 as to the identity of the generalisation function, where the learner remains uncertain as
 90 to which function applies. Viewing this as a prediction problem, a natural way for the
 91 learner to estimate the generalisation probability $g(y_i|\mathbf{x})$ (i.e., the probability that new

92 item y_i possesses the property given that the items \mathbf{x} possess the property) is to compute
 93 the posterior expected value of ϕ_i . For a discrete hypothesis space:

$$g(y_i|\mathbf{x}) = \sum_{h'} \phi_{ih} P(\phi_h|\mathbf{x})$$

94 This completes the formal statement of the model.

95 Implementation details

96 For the purposes of the current paper, the model was implemented in R (R Core Team
 97 2020) with some reliance on tidyverse functions (Wickham et al 2019). All details and
 98 documentation about the model and the R environment used to run analyses are available
 99 at the following repository: <https://github.com/djnavarro/sampling-differences>. Details
 100 regarding the environment were generated using the renv package (Ushey 2021) to aid future
 101 reproducibility. In all modelling exercises, the finite approximate hypothesis space consisted
 102 of $5^6 = 15625$ functions whose values are specified at the six stimulus values corresponding to
 103 the stimuli used in the experiment, using five possible probability values $\phi_i \in \{.1, .3, .5, .7, .9\}$
 104 for all items.

105 References

- 106 • Aust, F. & Barth, M. (2020). *papaja: Prepare reproducible APA journal articles with*
 107 *R Markdown*. R package version 0.1.0.9942, retrieved from [https://github.com/crsh/](https://github.com/crsh/papaja)
 108 *papaja*
- 109 • Cooper, R. P. and Guest, O. (2014). Implementations are not specifications: Specifica-
 110 tion, replication and experimentation in computational cognitive modeling. *Cognitive*
 111 *Systems Research*, 27, 42-49. <https://doi.org/10.1016/j.cogsys.2013.05.001>
- 112 • Hayes, B. K., Banner, S., Forrester, S. and Navarro, D. J. (2019). Selective sampling
 113 and inductive inference: Drawing inferences based on observed and missing evidence.
 114 *Cognitive Psychology*, 113. <https://doi.org/10.1016/j.cogpsych.2019.05.003>
- 115 • Navarro, D. J., Dry, M. J. and Lee, M. D. (2012). Sampling assumptions in inductive
 116 generalization. *Cognitive Science*, 36, 187-223. [https://dx.doi.org/10.1111/j.1551-](https://dx.doi.org/10.1111/j.1551-6709.2011.01212.x)
 117 [6709.2011.01212.x](https://dx.doi.org/10.1111/j.1551-6709.2011.01212.x)
- 118 • Navarro, D. J. and Perfors, A. (2011). Hypothesis generation, the positive test strat-
 119 egy, and sparse categories. *Psychological Review*, 118, 120-34 [https://dx.doi.org/10.](https://dx.doi.org/10.1037/a0021110)
 120 [1037/a0021110](https://dx.doi.org/10.1037/a0021110)
- 121 • R Core Team (2020). *R: A language and environment for statistical computing*. R
 122 Foundation for Statistical Computing, Vienna, Austria. URL [https://www.R-project.](https://www.R-project.org/)
 123 [org/](https://www.R-project.org/).
- 124 • Rasmussen, C. E., and Williams, C. K. I. (2006). *Gaussian Processes for Machine*
 125 *Learning*
- 126 • Shepard, R. N. (1987). Toward a universal law of generalization for psychological
 127 science. *Science*, 237(4820), 1317-1323
- 128 • Tenenbaum, J. B., and Griffiths, T. L. (2001). Generalization, similarity, and Bayesian
 129 inference. *Behavioral and Brain Sciences*, 24(4), 629-64

- 130 • Ushey, K. (2021). *renv: Project Environments*. R package version 0.13.2. [https:](https://CRAN.R-project.org/package=renv)
131 [//CRAN.R-project.org/package=renv](https://CRAN.R-project.org/package=renv)
- 132 • Wen, Y. P., Desai, S. C., Navarro, D. J., & Hayes, B. K. (in preparation). *Who is*
133 *sensitive to selection biases in inductive reasoning?*
- 134 • Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*,
135 4(43), 1686, <https://doi.org/10.21105/joss.01686>
- 136 • Xie , Y., Allaire, J. J., and Golemund, G. (2018). *R Markdown: The Definitive*
137 *Guide*. Chapman and Hall/CRC. ISBN 9781138359338. URL [https://bookdown.org/](https://bookdown.org/yihui/rmarkdown)
138 [yihui/rmarkdown](https://bookdown.org/yihui/rmarkdown).