Bell et al. reply

Andrew H Bell[1,2,3*], Chris Summerfield[2*], Elyse L Morin[1], Nicholas J Malecek[1], Leslie G Ungerleider[1]

[1]Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD 20892, USA
[2]Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1, 3UD, United Kingdom
[3]MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, UK CB2 7EF, United Kingdom

* These authors contributed equally to this study.

Corresponding Author: Andrew H Bell, andrew.bell@psy.ox.ac.uk

Main Text

In their comment [1], Vinken and Vogels take issue with our claim [2] that "IT neurons encode long-term, latent probabilistic information about stimulus occurrence". They offer a biologically plausible model of our findings, that they argue is based on neuronal fatigue. However, like our account, their model includes latent variables that are modulated slowly with stimulus probability. Models without such latent processes, such as those based on temporally local fatigue effects, cannot explain our findings. Although we share their desire for more clarity about the mechanisms underlying visual expectation, and appreciate their thoughtful critique, we would argue that their comment mostly restates our findings with a more complex model and alternate terminology.

Vinken & Vogels (hereafter, V&V) are correct that our design risks conflating the encoding of a latent probabilistic variable with a fatigue effect, i.e. an activity-dependent spike rate adaptation that is local in time and intrinsic to the neuron. Response fatigue is well-known to be strongest following a noiseless, recent adapter [3] and so we took care to verify that our effect of "expectation" was not simply driven by exposure to previous probe stimuli which (unlike the cue) were fully visible and obligatorily fixated by the monkey for 500ms at the end of the trial. To this end, in [2] we used a regression-based approach to demonstrate that our effect of expectation was robust to the removal of variance associated with probe identity on the *last three trials* prior to each cue. We also found no correlation across the neuronal population between sensitivity to faces and the potential adaptive effect due to a face on the previous trial, ruling out a spike-frequency dependent fatigue mechanism. Moreover, we note that the suppressive effects occurred exclusively in the late, sustained portion of the neural response, but did not dampen the initial stimulus-evoked peak, as typically observed following local stimulus repetition [4]. Subsequently, V&V's comment prompted us to compare the relative modulatory influence of the long-term history accruing from cues and probes. We found that the suppressive effect accrued from the cue history (which allowed the formation of expectations) but not from the probe history (although probes were both more visible and more recent; **Fig. 1**). All of these findings are consistent with an effect of expectation, rather than a local fatigue-based mechanism.

V&V do not dispute these empirical observations. Rather, they have built a more elaborate model that explains them in a biolologically plausible fashion. In their model, neuronal responsivity is modulated by two latent variables which vary depending on the long-term history owing to the cue (F_t) and probe (A_t). Each variable is updated using a hand-picked rate parameter; for example, following each cue $F_{t+1} = F_t - \alpha r_t^{norm} + L$ (L is a relaxation term). This equation has a similar functional form to the delta-rule model we use in [2] and the parameter value they chose to simulate our effects qualitatively, $\alpha = 0.12$, is close to our best-fitting estimate of $\alpha = 0.05$. Note that this low value for α implies a modulation of IT responsivity by stimulus probability over many trials (i.e. long-term), a finding wholly inconsistent with temporally local fatigue effects, such as spike rate adaptation. In other words, V&V's simulations are perfectly in line with the notion that neurons encode a latent quantity that varies slowly with the probability of stimulus occurrence. Their model is somewhat overparameterised (~13 hand-picked parameters at last count) in part because it was repeatedly elaborated following an earlier exchange concerning our data[1]. Although it may be only one candidate model among many, we are grateful to V&V for their

[1] We would like to refer the interested reader to full details of our exchange prior to submission of these comments, which are posted at http://decisions.psy.ox.ac.uk/publications/vinkelvogels

efforts to simulate our data in a biologically plausible fashion – a natural extension to the work described in [2].

V&V also raise an important technical point concerning our work. They point out that the regressors in eq.4 are potentially colinear, and use their simulations to demonstrate that a spurious face prediction error effect can emerge in our regression anlaysis, even in the absence of adaptation. Indeed, when the colinearity is removed by eliminating *p(face)* from the predictor matrix in simulation, this spurious effect disappears (**Fig. 2a**). This provided us with the opportunity to test whether, in the monkey data, we continue to observe a significant face prediction error signal when *p(face)* is removed from the regression. Indeed we do (**Fig. 2b**). Thus, whilst we are grateful to V&V for highlighting that care must be taken when using regression to analyse neural timeseries, the modulatory effects we report in the paper can in no way be attributed to our regression approach.

Where we are more confused is by V&V's terminology. They argue that their simulations demonstrate that our findings can be accounted for by "adaptation" rather than "expectation" [1]. This point would be better made if it were accompanied by a mechanistic definition of these terms – not least because "adaptation" is a biophysical phenomenon, whereas "expectation" is a cognitive construct. We agree that our results would be rather trivial if it could be demonstrated that suppression of face-evoked IT firing rates was simply due to response fatigue. However, the analyses described above unequivocally rule out such an account. Rather, explaining our results requires a model in which neuronal responsiveness is modulated by a latent variable (such as F_t) that fluctuates on a timescale of tens of seconds (at least) – i.e. that encodes "long-term, latent probabilistic information about stimulus occurrence". Of course, one might take issue with our use of the term "expectation" on the grounds that the monkeys may not be subjectively "expecting" anything – of course, deprived of the opportunity to ask them, we cannot rule this out. However, we do know that the animals capitalised on the probabilistic task structure to bias responding towards the cue (face or fruit) that was more probable over the long term, in exactly the manner that expectant humans are observed to do. Thus, it seems reasonable to ascribe our neural findings to visual "expectation".

Of much greater interest than these terminological issues are the precise neural mechanisms involved. In [2], we noted that our data are consistent with a wider computational framework, known as predictive coding [5], which proposes that suppressive signals originate in hierarchically higher cortical stages. However, we fully acknowledge that other mechanisms that permit latent quantities to be encoded over the long term could similarly account for our findings. These might include synaptic mechanisms, local circuit interactions, or interactions of these processes with the message-passing theory we emphasised in [2]. We agree entirely that more work is required to address this important question, and we hope that future discussion can proceed with greater consensus on the best terms to employ.

FIGURES

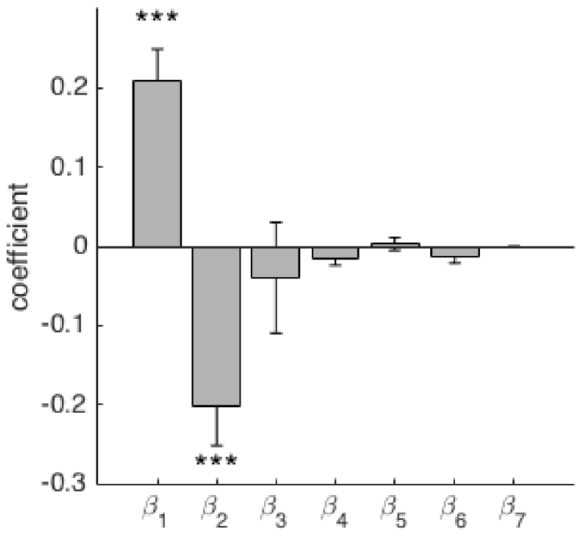

Fig. 1. Regression coefficients associated with cue identify (face vs. fruit; β_1), cue identity x cue probability (β_2), cue identity x response history (β_3), cue identity x previous trial (t-1 to t-3; $\beta_{4\text{-}6}$) and trial number (β_7) on IT firing rates, averaged over the post-stimulus period (0-500ms). *** indicated $p < 0.001$. Suppressive modulation is due to cue probability (β_2) not the history of fixated (fully visible) probes (β_3).

References

1. Vinken, K., and Vogels, R. (2017). Evidence for encoding of probabilistic information in macaque inferior temporal cortex explained by adaptation: a comment on Bell et al (2016). Current biology : CB.
2. Bell, A.H., Summerfield, C., Morin, E.L., Malecek, N.J., and Ungerleider, L.G. (2016). Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. Current biology : CB.
3. Vogels, R. (2016). Sources of adaptation of inferior temporal cortical responses. Cortex *80*, 185-195.
4. Verhoef, B.E., Kayaert, G., Franko, E., Vangeneugden, J., and Vogels, R. (2008). Stimulus similarity-contingent neural adaptation can be time and cortical area dependent. The Journal of neuroscience : the official journal of the Society for Neuroscience *28*, 10631-10640.
5. Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., and Friston, K.J. (2012). Canonical microcircuits for predictive coding. Neuron *76*, 695-711.