We would like to thank all reviewers for their thoughtful and helpful feedback. In addition to incorporating helpful edits, we have rewritten sections 2.3 and 2.4.3 and updated sections 2.4.2 and 2.5 to clarify their key points. Below we highlight the essence of these changes.

Parameterizations and parametric assumptions (R1,R3,R4)

(**R1**) The model that we investigate splits into two pieces: an encoding model, $p(\boldsymbol{r}|\theta)$, and an optimized mapping from stimuli to parameters, $\theta(s)$. This split allows us to study features of optimized codes that are agnostic to the encoding. However, we can't connect these features to neural activity $\boldsymbol{r}$ without an encoding model. So, to apply our model to neural data, we must either specify or fit such an encoding model. (**R1**, **R3**) In the example in section 2.5, the encoding model is given by $p(\boldsymbol{r}|\boldsymbol{\eta}(\theta))$ where $p(\boldsymbol{r}|\boldsymbol{\eta})$ is an assumed exponential family distribution with parameters $\boldsymbol{\eta}$, and $\boldsymbol{\eta}(\theta)$ is a curve through the parameter space, which is fit, akin to generalized linear model regression. (**R3**) This allows us to separate the question of parameter fitting (i.e., finding $\boldsymbol{\eta}(\theta)$) from the question of adaptation (i.e., identifying how $\theta(s)$ changes with $p(s)$). The parameter space curve, $\boldsymbol{\eta}(\theta)$, can be fit to data from any stimulus distribution. This curve, $\boldsymbol{\eta}(\theta)$ is assumed to be fixed within the timescale of an experiment, while adaptation to different stimulus distributions, $p(s)$, determines $\theta(s)$. The timescale difference is essential for our model to be applicable, and can be checked in data by fitting the parameter curves $\boldsymbol{\eta}(\theta(s))$ for different stimulus distributions to ensure that they differ only in $\theta(s)$. This discussion has been added to sections 2.4.3 and 2.5.

(**R4**) The parameters, θ, flat Fisher space parameters, $\hat{\theta}$, and stimuli, s, are all sufficient statistics of s in the neural activity. So, what do we gain from changing $s \to \theta$ to $s \to \hat{\theta}$, and what is the role of the latter? The parameter θ is inherited from our parameterization of the encoding model $p(\boldsymbol{r}|\theta)$, but is not unique. In fact, we could choose post-hoc a new parameterization $\omega(\theta)$ so that the new parameters have any optimal distribution, $p_\omega(\omega)$, of our choice, with associated Fisher information $\sqrt{I_\omega} = \sqrt{I_\theta}\, d\theta/d\omega$. The flat-Fisher parameter $\hat{\theta}$ is the unique choice of ω for which this Fisher information becomes constant, or 'flat'. Furthermore, $p_{\hat{\theta}}$ is fully determined from measurable quantities, $p_{\hat{\theta}} = p_s/\sqrt{I_s}$. Thus the flat Fisher condition renders $p_{\hat{\theta}}$ independent of our initially chosen (potentially arbitrary) parameterization, and therefore reflects features of the code irrespective of this choice. (**R4**, **R1**) This highlights an important insight gained from working with general encoding models: the flat space distribution $p_{\hat{\theta}}$ reflects intrinsic features of the code adaptation, while other parameterizations can be confounded by features of their assumed encoding models. Previous approaches, e.g., Wei & Stocker (2015), have worked in a similar space, arguing from examples. Section 2.3 has been rewritten and updated to reflect this intuition.

Fixed point iteration approach (R1,R2,R3,R4)

(**R1**) The fixed point of neural adaptation has two key features. First, it depends only on the constraint, and has the form: $p_{\hat{\theta}} \propto \exp(\lambda C(\theta(\hat{\theta})))$. This closed-form expression for the fixed point guarantees that a fixed point will exist. Second, this fixed point can be experimentally identified by a *closed-loop experimental scheme*, where the neural activity measured in response to stimuli drawn from one stimulus distribution determines the stimulus distribution that will be presented next. Briefly, we allow the neural code to adapt to an initially uniform stimulus distribution, $p_s(s) \propto 1$, and then measure neural responses $\boldsymbol{r}_i$ to stimuli $s_i \sim p_s$. From these responses, we fit the flat space distribution, $p_{\hat{\theta}}$, which then becomes the new stimulus distribution. The process of re-adaptation, measurement, and updating the stimulus distribution repeats until there are no changes to the stimulus distribution. At this point, $p_{\hat{\theta}} = p_s$, and the adaptation fixed point is achieved. The fact that this fixed point distribution only depends on the constraint function, $C(\theta(\hat{\theta}))$, allow us to fit the constraint. (**R4**) Once we have recovered the constraint function from the last step of the iteration, we can fit the objective function from the first iteration step. For this step, the stimulus distribution is uniform, so the flat space parameter distribution has the form $p_{\hat{\theta}} \propto 1/\hat{f}^{-1}(\lambda C(\theta(\hat{\theta}) + Z)$, allowing the objective function to be fit using the constraint function that we already know. We have rewritten section 2.4.3 to clarify these points.

(**R2**, **R3**) Along with improving methods for fitting encoding models, application of these methods to neural data remains an high priority. The fixed-point procedure we described is a closed-loop experimental paradigm, and so is difficult to apply outside of this context. This doesn't preclude application of other parts of our theory to existing experimental data. Fairhall et al. (2001) and Brenner et al. (2000), for example, observed that the distribution of activity parameters (firing rate) was not impacted by (rapid) re-adaptation to different stimulus distributions. According to our log objective function test this implies a log Fisher information objective for the measured codes. This is reasonable for early sensory neurons, since the log objective function approximates mutual information. This was added as an example to section 2.4.2.

Revised statement of broader impact (R1,R3)

(**R1**, **R3**) We have revised the statement of broader impact to read: "This work builds toward improved understanding of the neural code, and thus, a better understanding of operation of the nervous systems and the behavior of humans and other animals. Such understanding is important for neural prostheses and may lead to novel treatment options of human brain diseases. One potential risk is that a better understanding of neural coding, particularly value coding, lends itself to abuse for human behavioral modification. This is not an immediate concern for the work presented here."