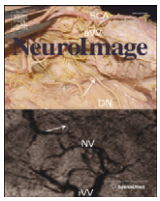
NeuroImage 60 (2012) 2294–2299

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg



Comments and Controversies

Confounding of norm-based and adaptation effects in brain responses

David Alexander Kahn, Geoffrey Karl Aguirre *

ARTICLE INFO

Article history:
Accepted 15 February 2012
Available online 25 February 2012

Keywords:
Prototype effect
Norm-based encoding
Representational similarity
Neural adaptation
Category learning

ABSTRACT

Separate neuroscience experiments have examined two properties of neural coding for perceptual stimuli. *Adaptation* studies seek a graded recovery from neural adaptation with ever greater dissimilarity between pairs of stimuli. Studies of *prototype* effects test for a larger absolute response to a stimulus which is distant from the center of a stimulus space. While intellectually distinct, these effects are confounded in measurement in standard neuroscience paradigms and can be mistaken for one another. Stimuli which are more distinctive are less subject to adaptation from perceptual neighbors. Therefore, a putative prototype effect may simply result from greater adaptation of prototypical stimuli by other stimuli in the experiment. Conversely, stimulus pairs which are the most perceptually distant from one another, and therefore expected to show the greatest recovery from adaptation, disproportionately draw from the extremes of the stimulus space. Thus, an apparent neural similarity effect may be created by an underlying prototype representation. We simulate BOLD fMRI results driven by each possible effect and demonstrate spurious results in support of the complementary effect. We then present an example fMRI experiment that demonstrates the confound and how it may be minimized. Finally, we discuss the implications of this intrinsic confound for studies of perceptual representation, neural coding, and category learning.

Introduction

A common target of neuroscience studies is the form of neural coding used to represent variation in stimulus properties. Very often, such studies use stimuli with linear variation along a single dimension. Examples of these "morphed" stimuli include facial image morphs of identity (Freeman et al., 2010; Jiang et al., 2006; Kahn et al., 2010) or emotional expression (Said et al., 2010a), mathematically defined abstract shapes (De Baene and Vogels, 2010; Panis et al., 2010), or auditory cues (Latinus et al., 2011).

Within the broad category of distributed neural encoding models (Barlow, 1972; Edelman, 1998), perceptual variation can be expected to have several neural correlates. Norm-based encoding models (Leopold et al., 2001; Rhodes and Jeffery, 2006) postulate that variation relative to a reference point in a stimulus space results in differential absolute responses to stimuli. These differences may take the form of a "prototype" effect (Valentine, 1991): a reduction in the neural response to a centrally-positioned prototype relative to those stimuli that are more extremely positioned. However, other distributed encoding models are possible, including those in which a stimulus space is represented using tuning functions that do not depend upon a particular point of space as a reference. In such a case, as in all distributed encoding models, perceptual variation could be indexed by the overlap in neural populations constituting two distributed representations. One manifestation of this form of neural representation is an "adaptation" effect (Grill-Spector and Malach, 2001; Henson and Rugg, 2003): a reduction in the neural response to a stimulus resulting from recent presentation of an identical or related stimulus. As defined, these two effects of encoding are intellectually distinct and based upon related and well-defined schema for neural representation.

Testing for these two effects of perceptual variation is possible via neuroimaging. Prototype effects, hypothesized to manifest as a larger bulk neural response to extreme stimuli, have been observed using functional magnetic resonance imaging (fMRI) in response to faces (Freeman et al., 2010; Loffler et al., 2005; Said et al., 2010a), face profile silhouettes (Davidenko et al., 2011) and abstract shapes (Panis et al., 2010). Similar findings have been demonstrated in monkey electrophysiological recordings (Leopold et al., 2006). Adaptation effects, a form of "carry-over" effect of one stimulus upon another (Aguirre, 2007), manifest as an increasing reduction in neural response for the latter stimulus in a sequentially-presented pair as a function of the pair's dissimilarity in fMRI (Drucker and Aguirre, 2009; Jiang et al., 2006) and ERP (Kahn et al., 2010). Graded neural adaptation related to stimulus similarity has been demonstrated in MEG (Furl et al., 2007) and in neuronal firing in monkey electrophysiology studies (De Baene and Vogels, 2010).

Despite being coherent and distinct predictions of neural models, we show here that these effects are confounded in measurement, and thus can be mistaken for one another. Importantly, while counter-balance (Aguirre et al., 2011) in the order of stimulus presentation is ultimately necessary to address this confound, it is not sufficient to remove it.

* Corresponding author at: Department of Neurology, University of Pennsylvania, 3 West Gates, 3400 Spruce Street, Philadelphia, PA 19104, United States. Fax: +1 215 349 5579.
E-mail address: aguirreg@mail.med.upenn.edu (G.K. Aguirre).

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2012.02.051

An example stimulus space

Consider a simple experiment that presents stimuli in a counterbalanced order from a set of five, evenly spaced morphed faces (Fig. 1a; morphs created using Photoshop CS5.5, Adobe; & JPsychoMorph). We may then ask if different face morphs have systematically different relationships to the set of stimuli as a whole.

Faces from the center of the space will, on average, be preceded and followed by faces which are more similar: on average, there will be a transition of 1.2 positions within the stimulus space from a center stimulus to the prior or next stimulus in the sequence (Fig. 1a). In contrast, faces from the ends of the stimulus space will have transition sizes of 2.0 positions from sequentially adjacent trials on average. Thus, the position of the stimulus within the space is related to the size of transitions in which it is involved. If different neural responses attended stimulus transitions of different sizes, this relationship would produce different average neural amplitudes to the different faces, *even if the neural responses to the faces themselves were identical*. This is a mechanism by which neural adaptation to stimulus similarity alone might be mistaken for a prototype effect.

This can be appreciated in the complementary analysis as well (Fig. 1b). Consider the sizes of transitions that are available between the faces in the experiment. Only stimuli from the ends of the space can be involved in the largest transitions. Conversely, small transitions disproportionately involve the faces from the center of the space. If the faces from the ends of the stimulus space evoked larger neural responses than faces from the center, this relationship would produce different average neural responses to the transitions of different sizes, *even if there was no effect of transition size itself upon neural response*. This is a mechanism by which prototype effects alone might be mistaken for neural adaptation to stimulus similarity.

We note that the first of these concerns has been recognized previously (Davidenko et al., 2011; Panis et al., 2010). We expand upon these prior observations by highlighting the reciprocal nature of this confound (which affects more than just studies of norm-based encoding), describing steps to mitigate the problem, and illustrating the explanatory potential when this complexity is embraced by experimental designs rather than eliminated.

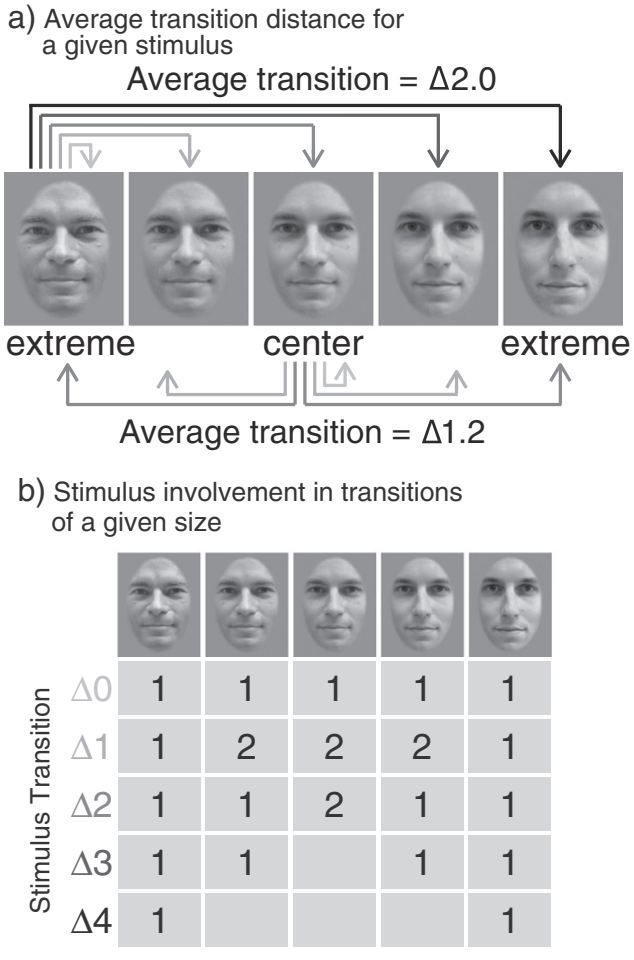


Fig. 1. Consequences of a counterbalanced experimental design with a 5-exemplar morph space. (a) An example stimulus set consisting of the two authors of this paper morphed in 5 equal steps. The average distance of all possible transitions from the central face is less than that from either extreme face. (b) Relative representation of each stimulus in every possible transition distance for a counterbalanced stimulus sequence. Transition distance is measured as the number of steps within the stimulus space between the preceding and current stimuli.

A simulated experiment

We conducted a simulation of an experiment that uses a linear morph space. Following the parameters of a recent study of prototype representation (Panis et al., 2010), we created a sequence for presentation of five stimuli (along with blank trials) using OptSeq2 (NMR Center; Massachusetts General Hospital, Boston, MA).[1] An inter-trial-interval of 2000 ms was assumed.

We first simulated the case in which a neural population has norm-based (prototype) coding for the stimuli, but no neural adaptation takes place. Fig. 2a (top row) shows the "carry-over matrix" (Aguirre, 2007) which characterizes the neural response to a given stimulus as a function of the prior stimulus. As can be seen, the modeled neural response is entirely determined by the identity of the current stimulus ("direct" effects). The particular amplitudes of response used were taken from the measure of a behavioral prototype effect (Upper left panel of Fig. 4 of Panis et al., 2010).

Given the sequence of stimuli and the matrix of neural responses, a simulated BOLD fMRI signal was generated (grey line, top row, Fig. 2b) using an assumed hemodynamic response function (Aguirre et al., 1998).

We then analyzed the simulated data using a model that tested only for the presence of neural adaptation effects, and ignored the possibility of a prototype effect. In the model, covariates were generated to model transitions between the stimuli of different step sizes. As can be seen, the model of a non-existent neural effect fit a substantial portion of variance in the simulated BOLD data (red line, top row, Fig. 2b). A plot of the loading on the model covariates reveals a spurious effect that could easily be mistaken for a linear neural adaptation effect (top row, Fig. 2c). Therefore, in data that contain only "prototype" neural effects, a spurious neural adaptation effect might be found.

Next, we simulated the case in which a neural population scales the amplitude of response dependent upon the similarity of the prior stimulus in the sequence, but which has equal responses to all the stimuli in isolation (Fig. 2a, bottom row). Again, simulated BOLD fMRI data were generated. These data were then modeled assuming that only direct effects of the stimuli are present in the data, and ignoring any possible neural adaptation. Separate covariates were fit to the average neural response to each stimulus identity (Fig. 2b, bottom row). The result (Fig. 2c, bottom row) is a spurious "prototype" effect, in which larger amplitude neural responses are measured for the stimuli from the extremes of the stimulus range. Therefore, in data that contain only neural adaptation effects, a spurious "prototype" effect may be measured.

[1] While the optseq program offers "preoptimized first-order counterbalancing", it does not actually provide perfect counter-balance of the stimuli (Aguirre et al., 2011). This has no consequence for the didactic purpose of our simulation, but would complicate attempts to remedy the confound within a linear model.

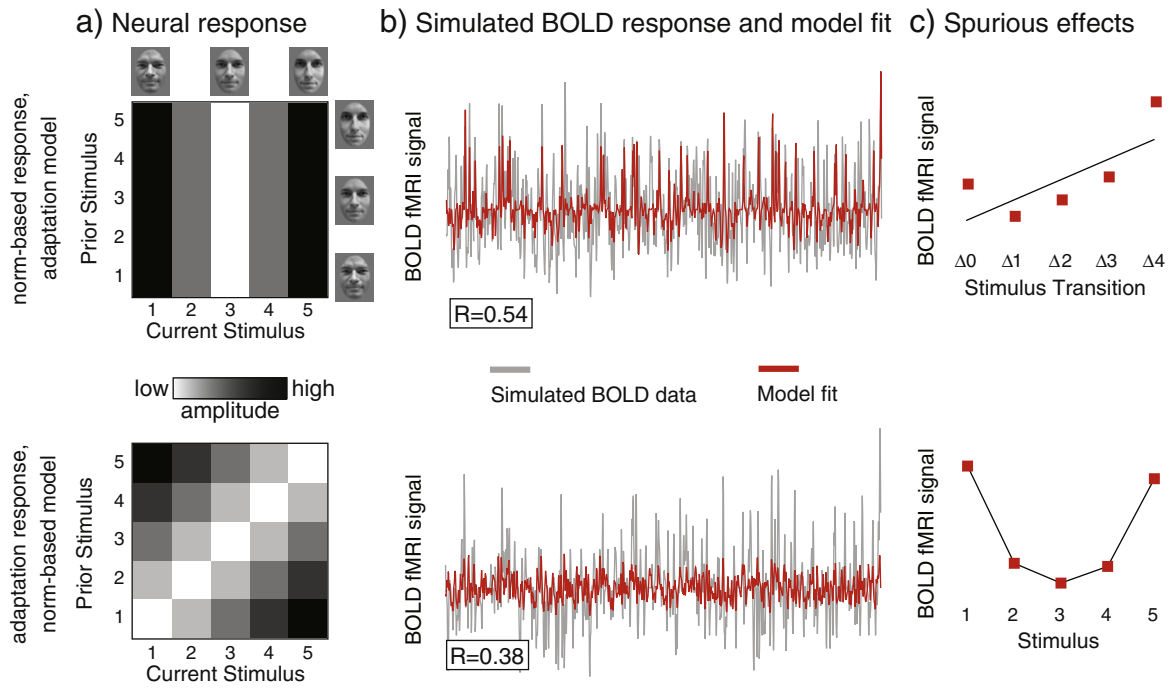


Fig. 2. A simulation of fMRI BOLD response demonstrating confounding of prototype and similarity effects. TOP ROW (a) A neural response model in which only prototype effects are postulated, with extreme stimuli resulting in a greater bulk response. The amplitude of neural response is driven entirely by the current stimulus with no modulation by preceding stimuli. (b) Simulated BOLD response for a counterbalanced stimulus presentation driven by prototype effects (grey). The model fit (red) represents a covariate modeling a linear adaptation effect for transition distance but not the prototype effect driving the data. (c) A spurious linear neural adaptive effect of similarity resulting from solely un-modeled prototype effects. BOTTOM ROW (a) A neural response model in which only stimulus similarity effects are present, with large transitions resulting in the greatest neural response (recovery-from-adaptation) and repetitions yielding the smallest. Individual responses are a function of the distance of the prior stimulus to the current stimulus (b) Simulated BOLD response for a counterbalanced stimulus presentation driven by similarity effects (grey). The model fit (red) represents a covariate modeling a prototype effect for stimulus identity but not the similarity effect driving the data. (c) A spurious effect of prototype resulting from solely un-modeled neural adaptive effects.

An empirical example

We next collected fMRI data from one participant (naive to the hypotheses of this study) to demonstrate these spurious effects in practice, and mitigation of the confound through concurrent modeling of both effects. Stimuli were 5 radial frequency contours (RFCs; Op de Beeck et al., 2001) created along a linear axis of varying RFC-phase and amplitude, and rendered with a pseudorandom black checkerboard texture on a gray background. The stimuli, subtending $5° \times 5°$ of visual angle, were back projected onto a screen and viewed by the subject via a head coil mounted mirror. The five stimuli, an additional target stimulus (an RFC orthogonally related to the morphed stimuli), and a blank trial were presented in sequences defined by second-order counterbalanced $k = 7$, $n = 3$, de Bruijn cycles (Aguirre et al., 2011). Each of 4 runs consisted of 343 continuous trials. Trials consisted of a 900 msec stimulus presentation followed by a 200 msec inter-stimulus interval of a grey blank screen (Fig. 3a). The duration of all blank trials was either doubled or tripled pseudorandomly to fit the 343 trials to 154 TRs. The subject was directed to monitor for the appearance of the target stimulus and respond with a button press. Echo-planar BOLD fMR images were collected (TR = 3 sec), with 3 mm isotropic voxels on a Siemens 3-T Trio with a 8-channel head coil. A functional localizer (Harris and Aguirre, 2008) consisting of faces, objects, buildings, and phase-scrambled images was also run for use in region-of-interest (ROI) definition. We defined an ROI corresponding to the left ventral lateral occipital complex (LOC) that had a significantly greater response to objects and faces (as compared to buildings and scrambled images) and a significant response to the average (main effect) of all shape stimuli in the primary experiment as compared to a blank screen (Fig. 3b). The response to the different RFC shapes and their transitions were obtained within this ROI.

The raw data were sinc-interpolated in time to correct for slice acquisition order and motion corrected using least squares minimization. The effects of adaptation and prototype in the data, both in isolation and concurrently, were analyzed using a modified general linear model (Worsley and Friston, 1995). After accounting for serial correlation in the residuals and the covariates used, the statistical tests we report below had approximately 110 effective degrees of freedom.

Our first model contained covariates only for adaptation, without differences in absolute response to individual stimuli modeled. Four covariates modeled the possible step sizes from the prior stimulus to the current stimulus (Δ1 through Δ4, as in Fig. 1; identical stimulus repetitions, Δ0, served as a reference condition for the entire model to avoid over-fitting of the model to the data). Additional covariates modeled the main effect of stimulus presentation versus the blank trials, targets, transitions from blanks to a stimulus, and the whole-brain global signal. The weightings on these covariates are presented in Fig. 3c, top-left panel. A significant effect of step size, (evaluated simply as a t-test for [Step 4 - Step 1]; $t = 3.46$, $p = 0.0004$) is present in this analysis, but is subject to potential confound of norm-based effects.

A complementary analysis modeled only the "direct" effect of each stimulus. Four covariates were fit the response to the presentation of each non-target stimulus, referenced to the central stimulus. The additional covariates included in the model were the same as that for the prior, "adaptation only" analysis. The covariate weights for this model are presented in Fig. 3c, bottom-left panel. The presence of a norm-based effect of prototype should manifest as a greater response for the extreme stimuli relative to the central stimulus. A t-test for [(stimulus 1 − stimulus 3) + (stimulus 5 − stimulus 3)] in this one subject showed an effect in this direction ($t = 1.18$, $p = 0.12$). Thus a norm-based effect of prototype could be present, but is similarly subject to confound due to un-modeled effects of adaptation.

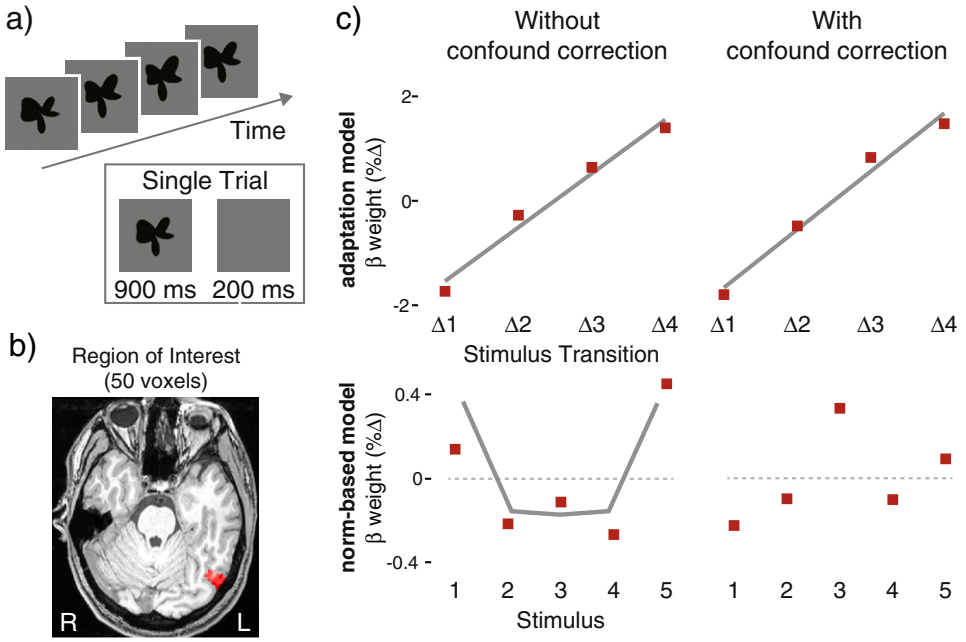


Fig. 3. An analysis of empirical fMRI data demonstrating both confounded and unconfounded measurement of adaptation and prototype effects. (a) Experimental design. Individual trials consisted of a stimulus presentation of 900 ms followed by a blank screen ISI of 200 ms. Trials proceeded continuously while the subject monitored for the appearance of an unrelated target shape (not shown). (b) The region of interest (ROI) used for statistical analysis, 50 voxels in ventral LOC, defined using an independent localizer comparison of [(Faces, Objects) − (Buildings & Scrambled Images), $t > 4$] crossed with a main effect of [Shapes, $t > 4$]. (c) TOP ROW Beta values from general linear models for covariates modeling the adaptive effects of transition distance. All values are mean-centered, One the left, the effects of stimulus identity are un-modeled in the GLM. On the right, the effects of both adaptation and identity are modeled concurrently. With concurrent modeling, the strength of the adaptive effect can be quantified without confound. BOTTOM ROW Beta values from a GLM for covariates modeling stimulus identity, indexed to the central stimulus. One the left, the effects of adaptation are un-modeled in the GLM. On the right, both effects are modeled concurrently. With concurrent modeling, the trend of a prototype effect is no longer present.

A third model contained both sets of covariates. The resulting beta values are presented in Fig. 3c, right panels. When controlling for the effect of prototype, the carry-over effects of adaptation observed in previous GLM remain in the larger model (evaluated as before, $t = 3.59$, $p = 0.0002$). However, when these carry-over effects are modeled in parallel, the suggestive trend of a norm-based effect of prototype, with the extreme stimuli yielding greater response than the central stimulus, is no longer present ($t = -1.17$, $p = 0.87$).

The ventral region we examined is close to that previously reported to exhibit proportional adaptation for two-dimensional closed contours (Drucker and Aguirre, 2009), and which is not thought to demonstrate significant norm-based effects of prototype (Panis et al., 2010).

With these data, we present an empirical example of the confound of adaptation effects and norm-based effects. We demonstrate that in the same data, incomplete modeling can lead to spurious effects for which complete modeling can account.

Implications for other studies

We simulated and measured this confound in a particular experimental design that presented the stimuli in a counter-balanced order, but it is present in other studies as well.

In an fMRI study, Jiang et al. (2006) argued in favor of a non-linear trend in recovery from adaptation as a function of stimulus dissimilarity. The authors presented a series of facial stimulus pairs of varying inter-stimulus distances drawn from a linear morph space, and observed the predicted recovery-from-adaptation. However, as the authors used only an extremely-positioned stimulus as the adapting stimulus, and an uneven selection of test stimuli, it is possible that their measures of recovery-from-adaptation were confounded by un-modeled effects of prototypicality.

In another fMRI study, Loffler et al. (2005) presented blocks of faces varying in distinctiveness from an average face. The authors demonstrated an increase in neural response for blocks of faces further from the average face, a finding which was presented in support of a mean-centric direct effect. While the prototype explanation is possible, the experimental design is confounded in that faces more similar to a prototype are geometrically less distinct. Blocks of more prototypical faces would thus be subject to greater neural adaptation, yielding a reduction in response amplitude driven by carry-over effects. Davidenko et al. (2011) subsequently used an elegant stimulus manipulation to call this result into question. Within blocks of stimuli, they manipulated the distinctiveness of parameterized face silhouettes while controlling the physical variability of the stimuli in the block. While this stimulus manipulation removes the confound, it does not provide a generalized solution to the joint examination of distributed and norm-based neural coding.

A similar stimulus set was used by Leopold et al. (2006) during electrophysiological recordings of inferotemporal cortex in monkeys. The authors demonstrated increased neural activity for faces further from the average face in support of a norm-based encoding model. While it is difficult to assess the potential for confound, (indeed, different experimental methods can minimize this potential, as we will discuss) this study demonstrates that stimulus sets vulnerable to confounding of prototype and adaptation are not limited to human neuroimaging.

The confound of prototype and adaptation effects will have a more subtle effect in multi-voxel pattern analysis (MVPA) studies that make use of one-dimensional stimulus sets (Said et al., 2010b). In this case, the uncertainty regards the form of neural representation that is used to decode the stimuli. MVPA studies make use of the pattern of direct-effects across voxels (the average response to a given stimulus across presentations). The ability of an MVPA analysis to classify the identity of a stimulus from the pattern of activity may not be the consequence of a difference in the neural response to the stimulus itself, but instead as a consequence of differences in relative neural adaptation.

Finally, a confound between norm-based and adaptation measures has implications for studies of category formation as well. A common hypothesis predicts enhanced recovery-from-adaptation for a stimulus transition of a given distance that crosses a perceptual category boundary, relative to a transition of the same distance that does not cross a category boundary (Goldstone, 1994). However, as category boundaries typically bisect a stimulus space (and large stimulus transitions crossing the boundary have no within-category analog), stimulus transitions crossing the category boundary will preferentially sample stimuli from the center of the space, while within-category transitions will involve more extremely positioned stimuli. In such a case, a smaller bulk response to centrally-oriented stimuli due to norm-based coding effects could negate or even *reverse* the predicted alteration in adaptation effects driven by a category boundary. Indeed, we are aware of results from our lab (and others) that demonstrate this reversal and have to date remained unpublished due to puzzlement regarding the cause and interpretation.

It is important to emphasize that the acknowledgment of this confound does not negate the presence of a neural effect in the studies we cite. Instead, this confound leads to uncertainty regarding the precise form of neural coding that produced a measured neural response.

Mitigation of confound

With an understanding of the potential of confound between adaptation and direct effects, we can consider several steps that may be taken to mitigate the problem. It should be noted that the study we used as a model for our simulation (Panis et al., 2010) is also a model example of awareness of this possible confound. The authors considered the possibility of adaptation effects in their data, and conducted an appropriate post-hoc test (effectively, a measure of carry-over for some stimulus pairings).

Similarly, Davidenko et al. (2011) anticipated the possibility of adaptation effects yielding spurious prototype effects in a block design study similar to Loffler et al. (2005). The authors mitigated the effect of stimulus variation upon their measurement by matching variability within block while varying prototypicality across block. For block-design studies, this method is a appropriate mitigation of the confound. We describe below additional, and more comprehensive, responses to this confound.

Principally, covariates for prototype and neural similarity effects (more generally, direct and carry-over effects) should be included in the same general linear model. As our empirical example demonstrates, the presence of either effect may then be measured after accounting for the confounding regularities that exist in the order of stimulus presentation (see also, e.g., Kahn et al., 2010). The use of a fully counter-balanced stimulus sequence (Aguirre et al., 2011) is crucial, as this allows the two types of effects to be estimated efficiently and without bias (DeCarlo and Cross, 1990).[2]

A limitation of this solution is that the degree of correlation between direct and carry-over effects can become substantial, particularly in fMRI studies which are affected by the temporal filtering properties of the hemodynamic response. Careful design of stimulus sequences can enhance power for measurement of carry-over effects (Aguirre et al., 2011), improving the ability to model carry-over effects for measurement or removal. More broadly, a model that includes both carry-over and direct effects may be assessed with an omnibus F-test without negative consequences of correlation within the covariates. This test would reveal that the neural signal does code for the stimuli or their relationship without determining the relative contribution of these effects.

One means of avoiding this issue is through the design of stimulus spaces. For example, stimuli drawn from a circle within a two-dimensional space are not subject to this confound, as every stimulus is equidistant from the (prototypical) center of the space. Of course, a downside to such a design is the inability to present a stimulus in the center of the space, thus precluding the study of norm-based coding.

Finally, experimental design may be used to minimize the presence of carry-over effects within the data. For example, a sparse fMRI design with long inter-stimulus intervals (e.g., greater than 6 seconds) would both plausibly reduce adaptation effects and reduce the degree of confound. Stimulus masking may be used to similar effect. The effectiveness of these measures for the reduction of carry-over effects would be an empirical question, with measurement of the effect subject to the same confounds discussed.

Discussion

We have explored a particular confound in the study of perceptual variation and stimulus representations. While individual neuroimaging studies have sought evidence for either prototype or adaptation effects in relation to neural encoding schemes, we find that these have the potential to be confused. We further demonstrate this confound empirically, and find that spurious effects can arise with incomplete modeling of fMRI data.

While this confound does not negate the existence of claimed neural effects, it may call into question their interpretation. As we have recommended, researchers interested in solely norm-based or adaptation effects have several avenues toward isolation of their effect of interest. We would argue, however, that instead of striving to solely mitigate one effect or the other, a more holistic perspective toward neural coding effects and their interaction could be useful.

As we have discussed, the concept of a prototype effect is related to a differential absolute response to a central stimulus relative to an extreme one. We can classify these prototype effects as a type of *first-order* effect of representation — one related to the unique neural response to a stimulus. Adaptation effects pertain to the overlap between distributed neural responses and arise from comparisons of stimuli — a type of *second-order* effect of representation.

The preservation of both the uniqueness of each stimulus, and the relationships of stimuli to each other is thought to be a primary goal of perceptual representation. In psychology, these two features are captured in the concept of a *stimulus space*: a representation of both stimulus identity and stimulus relationships. Stimulus spaces posit a unique location for each exemplar, with distances within space as an index of stimulus similarity. In relating this conceptual space to *neural* representations, it is possible to reframe prototype and adaptation effects. *First-order* effects of prototypicality speak to the position within a stimulus space, and *second-order* effects of similarity (e.g. adaptation) speak to distance.

The concept of a stimulus space goes further than these simple features; such spaces can have *topology*. In his classic psychological observation, Amos Tversky (1977) highlighted that similarity relationships were prone to asymmetry, particularly in comparisons involving prototypes. For example, most observers judge an ellipse to be more like a circle, than a circle to be like an ellipse. Such asymmetries can be understood as slopes in the surfaces of a stimulus space, biasing an otherwise equal metric relationship toward one direction. These slopes create a surface topology oriented about the prototype that has both *second-order* and *first-order* consequences.

On the second-order, surface topology would result in biases of *adaptation* effects in the direction of the prototype. Just as metric similarity has a neural adaptive effect, so does this asymmetric bias. As demonstrated by Kahn et al. (2010) in an ERP study, differential

[2] Our paper which introduced the notion of simultaneous modeling of direct and carry-over effects in neuroimaging (Aguirre, 2007) erroneously states that "direct and carry-over effects are orthogonal when the order of presentation of stimuli is serially first-order balanced." While this is true for the forms of neural response considered in that paper, the current work demonstrates that it is not a true statement generally, as confounds do exist for some forms of response.

adaptation in the N170 and N250 evoked potentials occurs for comparisons of prototypical and extreme faces dependent upon the order of comparison - a *second-order* effect of prototype. Notably, this finding would be impossible without simultaneous modeling of *first-order* prototype effects and *second-order* similarity effects.

On the first-order, surface topology would result in differential elevation of two points in the stimulus space. As the topological slopes are oriented toward prototypical stimuli, the elevation, and therefore absolute neural response to prototypical stimuli would be reduced relative to extreme stimuli — exactly the prediction of norm-based encoding models. Thus thinking in terms of stimulus space *topology* highlights several of the dominant effects of representation currently researched.

The cohesion of the topological stimulus space model lends itself to one more avenue of investigation, namely the *dynamics* of such spaces, and the interactions which might drive changes in topology. Importantly, Panis et al. (2010) offered evidence in favor of a dynamic prototype effect. Were the dual effects of prototype and stimulus similarity modeled in parallel, it might have been possible to disentangle both the first- and second-order effects of stimuli, as well as the interaction of the two in driving the dynamics of the other.

We believe this is a promising area of investigation. It is possible that *second-order* effects of neural adaptation are instrumental in the molding of prototype effects in the short and long term. One prediction is that the pattern of neural adaptation effects across the course of an experiment changes in concert with the emergence of the dynamic prototype, as found by Panis et al. (2010).

The conclusions of this paper are therefore two-fold. In regard recent neuroimaging studies, we highlight a confound of stimulus effects which draw into question existing interpretations. We also suggest a more cohesive approach to investigating neural stimulus spaces that enables study of the dynamics of perceptual representation.

References

Aguirre, G.K., 2007. Continuous carry-over designs for fMRI. Neuroimage 35, 1480–1494.

Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998. The variability of human, BOLD hemodynamic responses. Neuroimage 8, 360–369.

Aguirre, G.K., Mattar, M.G., Magis-Weinberg, L., 2011. de Bruijn cycles for neural decoding. Neuroimage 56, 1293–1300.

Barlow, H.B., 1972. Single units and sensation: A neuron doctrine for perceptual psychology? Perception 1, 371–394.

Davidenko, N., Remus, D.A., Grill-Spector, K., 2011. Face likeness and image variability drive responses in human face-selective ventral regions. Hum. Brain Mapp. doi:10.1002/hbm.21367 (EPub:).

De Baene, W., Vogels, R., 2010. Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. Cereb. Cortex 20, 2145–2165.

DeCarlo, L.T., Cross, D.V., 1990. Sequential effects in magnitude scaling: models and theory. J. Exp. Psychol. Gen. 119, 375–396.

Drucker, D.M., Aguirre, G., 2009. Different spatial scales of shape similarity representation in lateral and ventral LOC. Cereb. Cortex 19, 2269–2280.

Edelman, S., 1998. Representation is representation of similarities. Behav. Brain Sci. 21 (4), 449–467 (discussion 467–98).

Freeman, J.B., Rule, N.O., Adams, R.B., Ambady, N., 2010. The neural basis of categorical face perception: graded representations of face gender in fusiform and orbitofrontal cortices. Cereb. Cortex 20, 1314–1322.

Furl, N., van Rijsbergen, N.J., Treves, A., Friston, K.J., Dolan, R.J., 2007. Experience-dependent coding of facial expression in superior temporal sulcus. Proc. Natl. Acad. Sci. U.S.A. 104, 13485–13489.

Goldstone, R., 1994. Influences of categorization on perceptual discrimination. J. Exp. Psychol. Gen. 123, 178–200.

Grill-Spector, K., Malach, R., 2001. fMR-adaptation: A tool for studying the functional properties of human cortical neurons. Acta Psychol. 107, 293–321.

Harris, A., Aguirre, G.K., 2008. The representation of parts and wholes in face-selective cortex. J. Cogn. Neurosci. 20, 863–878.

Henson, R.N.A., Rugg, M.D., 2003. Neural response suppression, haemodynamic repetition effects, and behavioural priming. Neuropsychologia 41, 263–270.

Jiang, X., Rosen, E., Zeffiro, T., Vanmeter, J., Blanz, V., Riesenhuber, M., 2006. Evaluation of a shape-based model of human face discrimination using FMRI and behavioral techniques. Neuron 50, 159–172.

Kahn, D.A., Harris, A.M., Wolk, D.A., Aguirre, G.K., 2010. Temporally distinct neural coding of perceptual similarity and prototype bias. J. Vis. 10, 12.

Latinus, M., Crabbe, F., Belin, P., 2011. Learning-Induced Changes in the Cerebral Processing of Voice Identity. Cereb. Cortex doi:10.1093/cercor/bhr077 (EPub:).

Leopold, D.A., O'Toole, A.J., Vetter, T., Blanz, V., 2001. Prototype-referenced shape encoding revealed by high-level aftereffects. Nat. Neurosci. 4, 89–94.

Leopold, D.A., Bondar, I.V., Giese, M.A., 2006. Norm-based face encoding by single neurons in the monkey inferotemporal cortex. Nature 442, 572–575.

Loffler, G., Yourganov, G., Wilkinson, F., Wilson, H.R., 2005. fMRI evidence for the neural representation of faces. Nat. Neurosci. 8, 1386–1390.

Op de Beeck, H., Wagemans, J., Vogels, R., 2001. Inferotemporal neurons represent low-dimensional configurations of parameterized shapes. Nat. Neurosci. 4, 1244–1252.

Panis, S., Wagemans, J., Op de Beeck, H.P., 2010. Dynamic Norm-based Encoding for Unfamiliar Shapes in Human Visual Cortex. J. Cogn. Neurosci. 23, 1829–1843.

Rhodes, G., Jeffery, L., 2006. Adaptive norm-based coding of facial identity. Vision Res. 46, 2977–2987.

Said, C.P., Dotsch, R., Todorov, A., 2010a. The amygdala and FFA track both social and non-social face dimensions. Neuropsychologia 48, 3596–3605.

Said, C.P., Moore, C.D., Norman, K.A., Haxby, J.V., Todorov, A., 2010b. Graded representations of emotional expressions in the left superior temporal sulcus. Front. Syst. Neurosci. 4, 6.

Tversky, A., 1977. Features of similarity. Psychol. Rev. 84, 327–352.

Valentine, T., 1991. A unified account of the effects of distinctiveness, inversion, and race in face recognition. Q. J. Exp. Psychol. A 43, 161–204.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI Time-Series Revisited - Again. Neuroimage 2, 173–181.