Rewards are categories.

Erik J. Peterson
Dept. of Psychology
Colorado State University
Fort Collins, CO

Chapter 3 – fMRI analyses

An acquisition

Data Details. fMRI data was acquired at the Intermountain Neuroimaging Consortium (INC) facility located at the University of Colorado at Boulder on a Siemens Allegra 3T (whole body) scanner. All 18 right-handed participants were pre-screened for the typical fMRI exclusion factors (e.g. metal implants, mental disorders, etc). High resolution anatomical data was acquired as a T1-weighted structural image, MPRAGE sequence, at 1x1x1 mm, (256 x 156 x 192) with a TR of 2530 ms, and TE of 1.64 ms, with a flip angle of 7°. All functional (i.e. BOLD) data was acquired with T2-weighted echo-planar imaging (EPI), at 2.29 x 2.29 x 4.00 mm (96 x 96 x 26), with a TR of 1500 ms, a TE a 25 ms, a flip angle of 75° and a FOV of 220 mm.

Four sets of functional data were acquired. The first was of the "refresher" for part 1 of the behavioral training (p**??**), spanning 241 volumes. The second and

third spanned part 2 of the stimulus-responses learning task, divided into 2 (nearly) even sets lasting 390 and 394 volumes respectively (again see p**??**). The fourth scan featured repeated presentation of gratings from both reward categories, in a random order. The intent of this scan was to isolate rewarding activity outside the primary task. This localizer was not in the end useful (discussed on p5).

Preprocessed (model) food. Following DICOM to nifiti-1 conversion using dicom2nii (`http://www.mccauslandcenter.sc.edu/mricro/mricron/dcm2nii.html`), each dataset was subjected to the following preprocessing pipeline carried out in SPM8's batch mode (`http://www.fil.ion.ucl.ac.uk/spm/software/spm8/`). For complete code see, `https://github.com/andsoandso/fmri/tree/master/catreward/spm_m`. Anatomical data was first segmented into white and grey matter regions (Collignon et al., 1995). Based on these segments the parameters necessary for normalization into a standard reference space (T1 MNI-352, at 1 mm, MNI space or short) were calculated. Normalization had two steps. The first was a Bayesian 12-parameter affine transformation (Ashburner, Neelin, Collins, Evans, & Friston, 1997). The second was a set of nonlinear deformations, using a 1127 parameter discrete cosine transform (Ashburner & Friston, 1999). Anatomical data was then resampled from 1.27 to 1.00 mm^3 using fourth degree β-splines, and finally, using the parameters above, normalized into MNI space.

To correct for the slight head movements that often occuring during scanning, movement regressors for all volumes of the functional data were first calculated

(Ashburner & Friston, 1999). No participant moved more than 1.5 mm, so all data was retained. Functional data was then slice-time corrected, using slice 13 (the middle slice from the descending acquisition) as the reference, followed by coregisteration with the pre-processed (native-space) anatomical data, and resampling into 3 mm^3 voxels, again using fourth degree β-splines (Collignon et al., 1995). Functional data was then normalized into MNI space using the anatomically-derived parameters above. Finally, the functional data was spatially smoothed using a 6 mm FWHM Gaussian, though a copy of the unsmoothed data was retained for the ROI analyses (described on p5). Just prior to regression analysis, each voxel's time course was also low-pass filtered using finite impulse response model, with a cutoff at 0.008 Hz (Kruggel, Cramon, & Descombes, 1999). For all whole-brain analyses, the movement regressors were entered into the regression models as covariates, accounting for any head movement. Given the large spatial averages employed in the ROI analyses these weren't motion corrected (Poldrack, 2007).

The best of all possible signals. In fMRI, and in general time-series analysis, there is an intrinsic trade-off between detecting a signal in the presence of noise and estimating the shape of that signal (Dale, 1999; Birn, Cox, & Bandettini, 2002; Liu, 2004). One way to optimize over both these conflicting objectives is to manipulate the trial order in a rapid event-related design (Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000). One state-of-the-art method for optimizing the trial ordering process is a genetic algorithm which uses two (weighted) loss functions, one for signal detection and one for time-course estimation (Wager & Nichols, 2003). Kao, Mandal, Lazar, & Stufken, 2009, improved on Wager's (2003) intial design by adding in a

loss function for psychological considerations, greatly improving execution speed and documentation. As a result, Kao *et al's* (2009) method/code was used to optimize trial orders for part 1 and 2 of the behavioral task (p**??**), along with the reward category localizer scan (p1).

Mobs of Blobs

All statistical parametric maps (below) were derived from a Random Effects analysis (RFX, or "second-level" in SPM8 jargon), multiple comparison corrected assuming Gaussian Random Fields using the Family Wise Error Rate (FWE) at the $p < 0.05$ level, with a minimum cluster size of 4 voxels (Worsley et al., 1996).

Whole brain activity for the stimulus-response learning portion of the behavioral experiment (i.e. part 2, p**??**) was examined first by comparing all trials to the baseline (rest) condition. This data is presented in two ways. First is the statistical thresholded image. This contrast map showed significant bilateral activity in the cerebellum, insula and anterior cingulate ($t(15) = 6.59$, $p < 0.05$; Figure 1). Second is an overlay of the raw t-values, which allows for visual confirmation the observed significant effects were robust and widespread in their respective regions, but also allowed for the analysis of overall and subthreshold patterns of activity. These raw data suggested near threshold levels of activity in the head of the caudate, ventra medial, dorsallateral frontal cortices as well as (weaker) activity in the occipital lobe (Figure **??**). And indeed in a two-way ANOVA looking at that interaction between gains and losses, significant clusters were observed in head and body of caudate, insula, posterior and anterior cingulate with the posterior activation extending into the

precuneus, as well as in dorsallateral (i.e middle frontal) PFC, and in ventralmedial PFC (Figure 3; $F(1, 270) = 30.76$, $p < 0.05$). When gains and losses were examined separately, but again compared to rest, both had activity in the same areas as in the combined condition (not shown).

Regions and models

The right chunks. Following whole-brain analysis, regions of interest were selected using two seperate yet related methods. The first employed only regions from the Harvard-Oxford probabilistic anatomical atlas, using the 50% cutoff (Desikan et al., 2006). The second combined anatomical regions with functional clusters isolated using both the data collected during the second half of part 1 (i.e. the "refresher") and from the reward category localizer (p1). Comparisons between them showed the anatomically-limited functional clusters and the entire anatomical regions displayed very similar results. So to limit the complexity of later analyses, and to increase power, functional clusters were discarded in favor of the larger anatomical regions. Most anatomical regions of interest were selected *a priori* based on previous studies of reinforcement and category learning (see the *Introduction* for a review). Left and right subcortical regions of interest were the dorsal caudate, ventral striatum/nucleus accumbens, and putamen. Bilateral cortical areas were the middle frontal cortex (i.e. dorsallateral PFC), frontal medial cortex (which contains ventralmedial PFC), and orbital frontal cortex. Based on the whole-brain maps (p4), regions for the insula, anterior and posterior cingulate (ACC and PCC for short) were included as well. While I've no strong *a priori* hypothesis for the role of these regions may play or

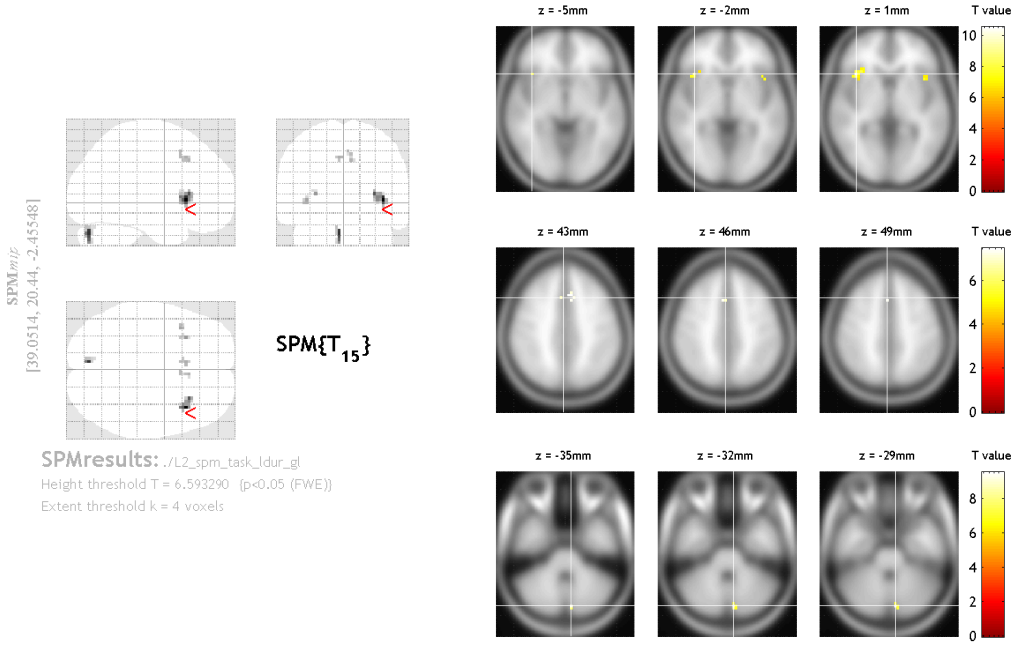

Figure 1. Statistical parametric map for all trials in the stimulus-response learning task (i.e. part 2, p**??**), compared to the rest period. *Left* is a glass brain, showing all significant clusters. *Right* is a set of axial slices highlighting strong areas of activity overlaid onto the T1 MNI-352 template. *Z* is the height of the axial slice in MNI space.

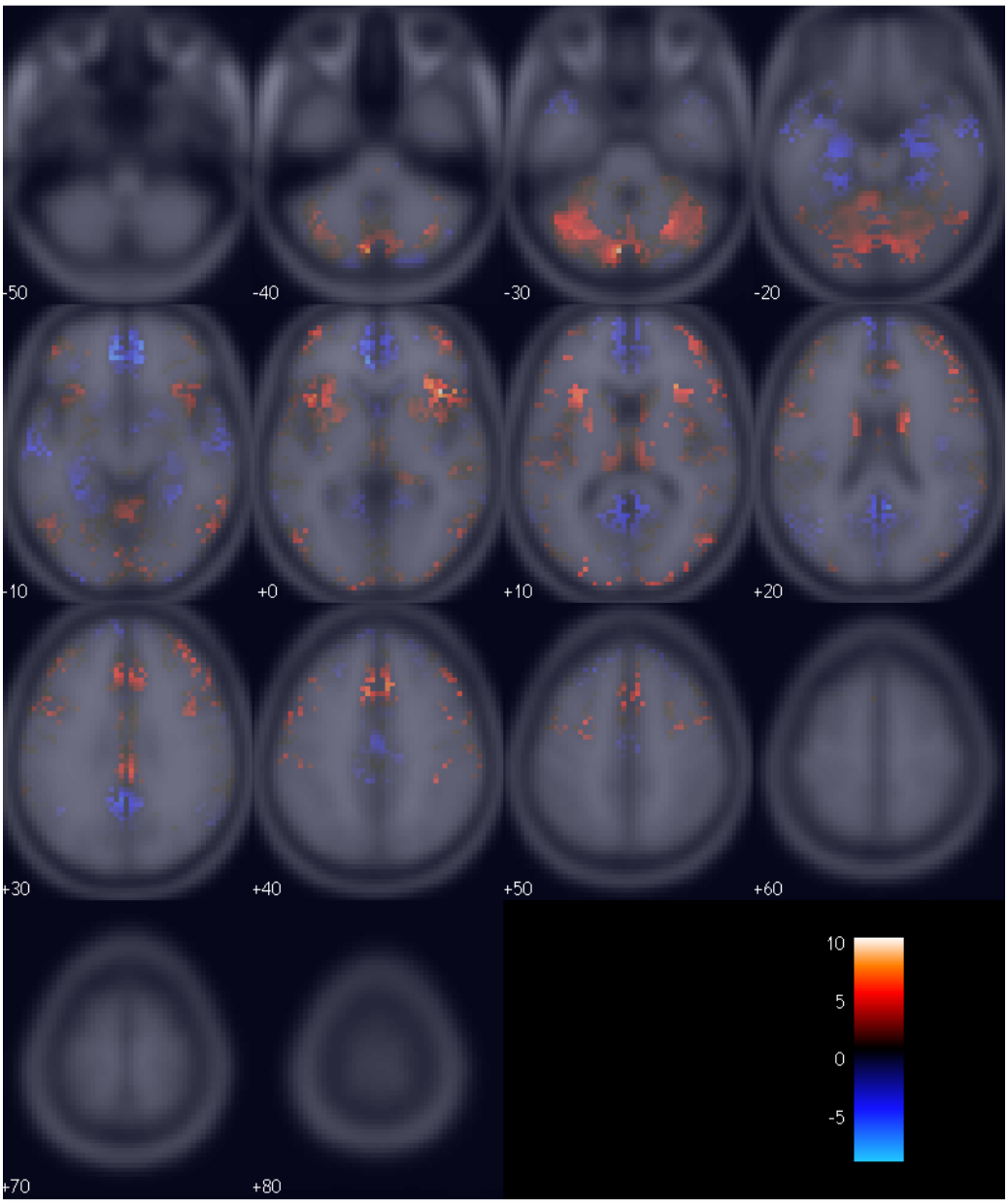

Figure 2. (Raw, that is unthresholded, t-values for all trials in the stimulus-response learning task (i.e. part 2), compared to the rest period, overlaid onto the T1 MNI-352 template. Each number is the height of the axial slice in MNI space.

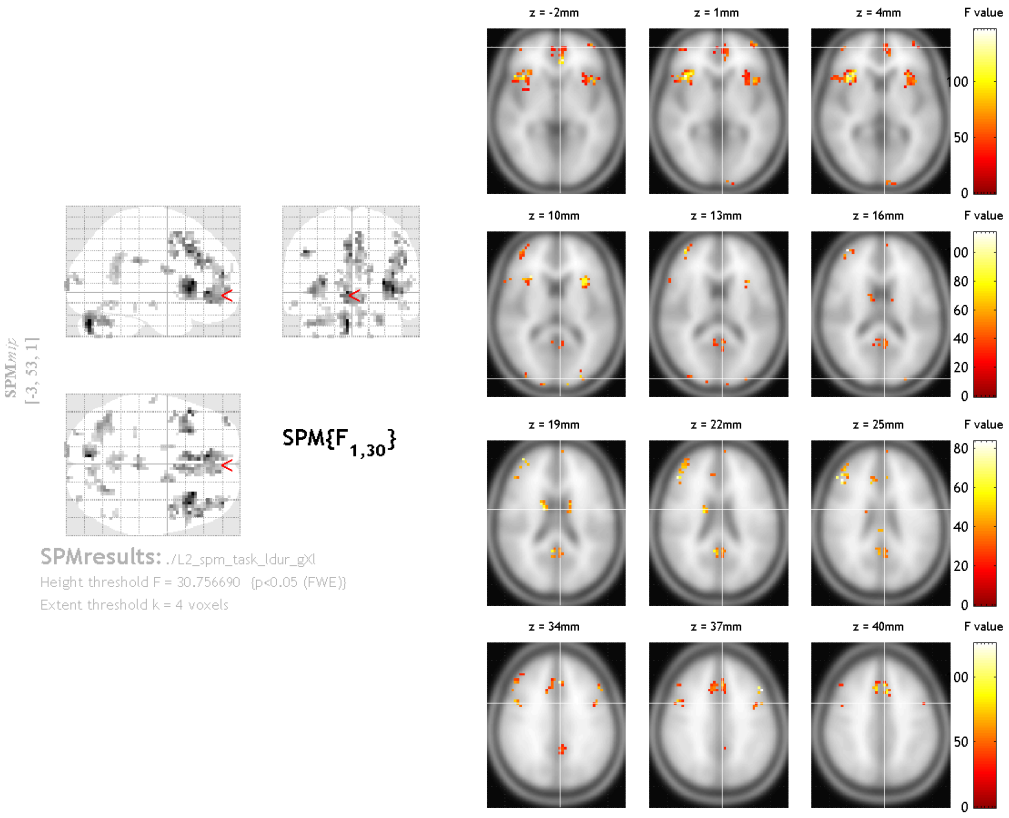

Figure 3. Statistical parametric map for all trials in the stimulus-response learning task (i.e. part 2) examining the interaction between gains and losses. *Left* is a glass brain, showing all significant clusters. *Right* is a set of axial slices highlighting strong areas of activity overlaid onto the T1 MNI-352 template. *Z* is the height of the axial slice in MNI space.

may not play, activity in each of these *post hoc* regions is common to human category learning experiments (Lopez-Paniagua & Seger, 2011; Seger, Peterson, Cincotta, Lopez-Paniagua, & Anderson, 2010; Cincotta & Seger, 2007; Seger & Cincotta, 2006, 2005).

A Way To(o) Many. There 6 models under evaluation: the three kinds of similarity adjustment ("none", "exp", and "gauss") mulitplied by the two possible reward codes ("acc" and "gl"), with the two terms of interest (i.e. value and the reward prediction error), that is 12 comparisons. The were also a number of *a priori* confounds to our signals of interest including the similarity metrics, the reward codes, and the grating parameters. Bringing the total to 23. As the models are not nested[1] and therefore not amenable to F-tests – the common statistical way to compare model fits. Further complicating the issue was the fact that each of the models is covariate, if not collinear, with the others. To top it off, none of the three similarity-adjustments are statistically independent; reinforcement learning can be viewed as a regression of the reward code onto behavioral choices. All these factors combined would make statistical testing difficult, to say the least. But fortunately finding *the* best model is not the goal.

The latest recordings of phasic (i.e. reward prediction) activity in the VTA/SNc suggests a complicated reward and prediction error coding scheme (see p**??**), wherein several separate sets of calculations may be carried out independently (Kim, Shimojo, & O'Doherty, 2006; Matsumoto & Hikosaka, 2009; Smith, Berridge, & Aldridge,

[1]Often defined by whether or not two models can be made identical by adding or subtracting parameters (Forster, 2000)

2011). The observed BOLD signal is then an aggregate of these many activities. It is possible, even likely, then that more than one of the models is correct making null hypothesis tests an incorrect choice. Model selection is the right choice.

Model selection is the process of finding a *family* of models that best predict a given dataset (Rao, Wu, Konishi, & Mukerjee, 2001). Most techniques try to wisely balance parsimony with increasing fit (i.e. solving the bias versus variance dilemma (Geman, Bienenstock, & Doursat, 1)). Unfortunately most model selection techniques require assumptions my models cannot meet (e.g. statistical independence). The few that can tend to be complex recent statistical inventions. Rather than navigate the those troubled and unproven waters, I took a simpler approach. Each model was independently examined and ranked, in an approach loosely similar to model averaging (Forster, 2000).

An AIC score (Akaike Information Criterion (Akaike, 1974)) was assigned to each of the models/codes for every participant and region of interest. The absolute AIC score across participants is not however meaningful. Only the relative values are of interest (Wagenmakers & Farrell, 2004). As a result, individual's scores were normalized and ranked by subtracting from each from the best (lowest) score (Anderson, Burnham, & Thompson, 2000). The normalized set was then transformed to Akaike Weights, a way to easily compare the conditional probabilities of each model being true (Wagenmakers & Farrell, 2004). The Akaike Weights were then averaged across participants for each model and region of interest.

Information on Information. AIC is a measure of loss; how much information is lost by substituting the model for the true distribution, i.e. the data. The lower the AIC score, the better the model. Unlike null hypothesis tests and Bayesian measures, AIC-based methods do not seek to find *a* truth, but instead serve to rank models. AIC offers then only relative insight, and is unable to make any claims about absolute significance. Significance is a separate question, one I'll return to later. Besides this limitation, AIC has some substantial advantages. Five are reviewed below.

One, unlike maximum-likelihood AIC is designed to be a parsimonious score. It penalizes for additional parameters. It may therefore choose a worse model (as measured by likelihood or mean squared error) over a better but more complex one. This is the essence of Occam's razor[2].

Two, it fits with the process of science. When designing an experiment it is rare that there are only two possible outcomes, instead typically there are several competing hypothesis, some of which may not be mutually exclusive. AIC's focus on relative differences and evidential weights meshes perfectly with the reality of multiple working hypotheses (Burnham, 2004).

Three, truth can remain elusive. A common alternative to AIC is BIC, the Bayesian Information Criterion. Like AIC, BIC is derived from the log-likelihood of a model, however its derivation requires a rather strict (and often unrealistic) assumption – that the true model is among the candidates (Forster, 2000). And while it may be philosophically debatable whether any mathematical model can *completely*

[2]Famously and pithily expressed as, "Entities are not to be multiplied beyond necessity".

describe reality, in this study I know my models are incomplete. As, one, the human reinforcement learning literature contains several recent theoretically unaccounted for findings and, two, there are theoretical developments I do not include here to keep the models tractable (see the *Introduction* for a review).

Four, AIC values are easily interpretable once they're transformed to Akaike Likelihoods or Weights[3]. The likelihood is, as you would expect, simply the likelihood the model is correct (based on the information loss associated with it), while the Akaike Weights are normalized likelihoods. As the Weights sum to one, the conditional likelihood of one model compared to another is just the ratio of their weights (Burnham, 2004). For example, the conditional likelihood of model A over model B is just w_A/w_B. That is, the likelihoods and Akaike Weights are intrinsically measures of effect size (Anderson et al., 2000; Forster, 2000). Despite the fact that it is often used to express the likelihood of correctly rejecting the null hypothesis, the p value is not a measure of effect, as p is contingent not just on effect size but on sample number.

Five, AIC has a history with models of categorization. McKinley & Nosofsky, 1996; Maddox & Bohil, 2001, among several others, used AIC to compare behavioral results to several alternative models of categorization.

F-Them. AIC ranks offer no information about significance, in the familiar null hypothesis sense, or about the absolute fit of the model. I addressed both of these in

[3]Likelihood for model k among K working hypotheses/models is given by $L_k = e^{-0.5(AIC_k - min_K(AIC))}$, which is then normalized, becoming an Akaike Weight by $w_k = L_k / \sum_{k=1}^{K} L_k$ (Burnham, 2004).

a series of F-tests run prior to AIC analysis. These (fixed-effect, across participant) ominbus tests asked whether the total set of regression parameters for each linear model (described below) could explain the BOLD time series better than chance (i.e could the null hypothesis (of 0) be rejected). Keeping with recommendations of Burnham, 2004; Forster, 2000, who argue that as AIC and significance tests are so dissimilar that direct comparison/interaction between them will be at best misleading, the models are not discarded based on significance. All models are retained, and later AIC ranked. The F-tests are a separate measure whose results are integrated during interpretation, not during model selection.

Code, BOLD, and Models.. A total of 23 models were compared for each of the 12 regions of interest for each of the 16 subjects, 4416 comparisons in total. Each of the models is described below (Table 1). In general, a time-series (e.g the reward prediction error for each trial or the similarity for that trial's outcome) was convolved with a "canonical" haemodynamic response function, a mixture of gamma functions that serves as a parsimonious estimate of the (instantaneous) BOLD response (Friston et al., 1998). The convolved series was then low-pass filtered, matching the treatment of the BOLD data (p2). Each convolved and filtered model was then regressed onto the BOLD response for each participant's region of interest, retaining all parameters and fit measures inside subject-level HDF5 files. The HDF5 format offers high performance read/write operations, and widespread support across several scientific programming languages (`http://www.hdfgroup.org/HDF5/`).

No available fMRI analysis package returns AIC scores (or measures that

could be converted to such) and none allow for the efficient (i.e programmatic) analysis of many competing computational models. So I created a region of interest focused fMRI analysis tool in Python (v2.7.1) to meet those two needs. This module, simply named "roi", has since been release under the BSD license and is available for download at `https://github.com/andsoandso/roi`. It relies on the nibabel library to read the nifiti-1 files (v1.2.0; `http://nipy.org/nibabel`), nitime for timeseries analysis, (v0.4; `http://nipy.sourceforge.net/nitime/`) Numpy for generic numerical work (v1.6.1; `http://numpy.scipy.org/`), with the GLS function from the scikits.statsmodels module handling the regerssions (v0.40; `http://statsmodels.sourceforge.net/`). Model-to-BOLD fit parameters, as well as other useful metadata, was then extracted and stored in text files suitable for importing into R (v2.15.1; `http://www.r-project.org/`). All plotting and model ranking (as well as the F-tests) were carried out in R. For complete BSD licensed code see, `https://github.com/andsoandso/fmri/tree/master/catreward/roi/results`.

Our Kinds of Models. To ease visualization and analysis each of the models was classified into one of 5 families. Family one, denoted "boxcar", was identical to that first used in the whole-brain analysis (p4) – all trials versus the rest condition. This is a univariate time-series that predicts no trial-specific effects; No matter the task the brain, thus the BOLD response, just flicks on then off. It serves as a useful standard against which to compare the model-based regressors. The next two families were controls (i.e. *a priori* covariates). The reward codes, both raw and similarity adjusted, were in one family ("control_reward") and in the other were the

similarity metrics and grating parameters ("control_similarity"). The fourth family contained all the reward prediction errors ("rpe"). The fifth contained all value estimates ("value").

Table 1:: All models, their designations (Codes), families, and descriptions.

Number	Code	Family	Description
1	0_1	boxcar	The simplest model, a univariate analysis of all conditions.
2	acc	control_reward	Behavioral accuracy.
3	acc_exp	control_reward	Behavioral accuracy, diminished by (exponential) similarity.
4	acc_gauss	control_reward	Behavioral accuracy, diminished by (Gaussian) similarity.
5	gl	control_reward	Gains and losses.
6	gl_exp	control_reward	Gains and losses, diminished by (exponential) similarity.
7	gl_gauss	control_reward	Gains and losses, diminished by (Gaussian) similarity.
8	rpe_acc	rpe	Reward prediction error - derived from accuracy.

9	rpe_acc_exp	rpe	Reward prediction error - derived from accuracy diminished by (exponential) similarity.
10	rpe_acc_gauss	rpe	Reward prediction error - derived from accuracy diminished by (Gaussian) similarity.
11	value_acc	value	Value - derived from accuracy.
12	value_acc_exp	value	Value - derived from accuracy diminished by (exponential) similarity.
13	value_acc_guass	value	Value - derived from accuracy diminished by (Gaussian) similarity.
14	rpe_gl	rpe	Reward prediction error - derived from gains and loses.
15	rpe_gl_exp	rpe	Reward prediction error - derived from gains and losses diminished by (exponential) similarity.
16	rpe_gl_gauss	rpe	Reward prediction error - derived from gains and losses diminished by (Gaussian) similarity.

17	value_gl	value	Value - derived from gains and losses.
18	value_gl_exp	value	Value - derived from gains and losses diminished by (exponential) similarity.
19	value_gl_gauss	value	Value - derived from gains and losses diminished by (Gaussian) similarity.
20	exp	control_similarity	Outcome similarity (exponential).
21	gauss	control_similarity	Outcome similarity (Gaussian).
22	angle	control_similarity	Grating angle parameter.
23	width	control_similarity	Grating width parameter.

Model Results

I'll work through the many results first by subcortical areas then move on to the cortical. The general analysis strategy was to first find the top family, indicated by the largest family-average Akaike Weight. I then examined the next highest scoring to family to see if it was close to the top (i.e. $\leq$1.5 times as likely). If it was both, families were included. I then examined the relative likelihood of each model

in the top family/families. Within-family models that were about $\geq$1.5 times more likely then their neighbor were dubbed "substantively more informative". Like the significance thresholded in null hypothesis tests this $\geq$1.5 is an arbitrary threshold. However in order discuss and interpret these results a line must be drawn between meaningful and not, and $\geq$1.5 is a good minimum cutoff (Anderson et al., 2000; Forster, 2000). As I stated though at the outset, more than one model may be right. Thus the threshold was treated as a loose cutoff. To get a sense of overall model quality, I also calculated the likelihood of the best model over the boxcar (i.e. the non-parametric standard). Finally I examined all models, not just the top family, for any outliers that may have scored well despite their families overall poor performance.

Still as this was the first attempt to AIC-rank models of fMRI data, and while I put much thought and research into the above scheme, it may be flawed. It is also arbitrary (beyond the $\geq$1.5 cutoff); Why not discuss the top 3, or 4 families, or even just include them all? To attempt then to minimize the effect of these arbitrary, but necessary, decisions the complete set of models (and F-tests) are included for every region of interest.

From up high. For eight of the twelve regions of interest the "rpe" family scored highest. Of these eight, five were best described by "rpe_acc_gauss". The next best family was "control_similarity" with 3 regions, followed by "boxcar" with 1. Notably, "value" was not the most informative model family for any region of interest, and indeed the one region (ACC) for which it was second, "rpe" was 1.8 times more

likely.

Under Cortical. In the dorsal caudate (Figure 4), only the "rpe" family offered a more informative fit that the "boxcar", being 2.61 times more likely in the left and 2.85 in the right (though I'll just abbreviate left/right likelihoods as 2.61/2.85 from here on). Bilaterally, and using the "acc" coding scheme, the Gaussian similarity-adjusted model (i.e. "rpe_acc_gauss") was substantively more informative than either unadjusted model ("rpe_acc" – 1.45/1.54 or "rpe_gl" 1.82/1.70). Surprisingly, given its similarity to the Gaussian adjustment, "rpe_acc_exp" scored no better than the unadjusted models (above). In what will become a reoccurring theme when examining the F-tests, all models were significant bilaterally in the dorsal caudate (Figure 5). And while the F-values themselves to some degree mimic the patterns of the Akaike Weights, it would not be possible to reliably disassociate them given the slight relative differences.

Compared to the "boxcar", the putamen was also best described by the "rpe" family (1.53/2.31). However compared to caudate the putamen displayed markedly different within-family activity (Figure 4 compared to 6). The "rpe_acc" model was more substantively more likely (1.67/1.66) than the next highest ranking similarity model (i.e. "rpe_acc_gauss"). However due to the marginal bilateral significance (Figure 7), this interesting reversal must be viewed cautiously. The bilateral consistency in the Akaike Weights does offer some room for optimism (Figure 6, specifically I refer to the consistency and relative strength of "rpe_acc").

The right and left halves of the nucleus accumbens, the ventral portion of the

striatum, were quite divergent in their fits (Figure 8). However both ranked the "control_similarity" family as the most informative, however this family was only about 1.10 times more likely than the next family ("rpe"), which itself was not substantively better then its neighbor, and so on. So while there is strong evidence that the top models, "angle" in left (3.07) and "rpe_gl" (3.06) in the right, are better than "boxcar", the overall bilateral heterogeneity, weak family effects, combined with by-and-large non-significant outcomes on the left half, and weak F-values on the right (Figure 9), suggest this region was not strongly activated by the task. Further analysis is therefore futile.

On that thinkin' sheet. The insula was the one region, both cortically and subcorticallly, to be nearly equally well described both by the "acc" and "gl" reward codes (Figure 10). In the top ranking "rpe" family (which was 1.37 times more likely than its neighbor "control_similarity", and 2.31 more likely than the "boxcar") within-family models showed divergent patterns based on the code. The "acc", "rpe_acc" and "rpe_acc_gauss" models dominated "rpe_acc_exp" (around 1.4). The "gl" code had the opposite effect, "rpe_gl_exp" dominated "rpe_gl" and "rpe_gl_gauss" by a somewhat similar amount (1.20). The strong overall significance of all models (Figure 11) suggests these relative rankings may reflect truth; like the dorsal caudate the F-values the have same patterns as Akaike Weights.

Both the ACC and PCC displayed very similar rankings of their Akaike Weights (compare Figures 12 to 14), so I'll discuss them as one. Again the "rpe" family dominated (respectively 2.07 and 1.77 times more likely compared to the nearest

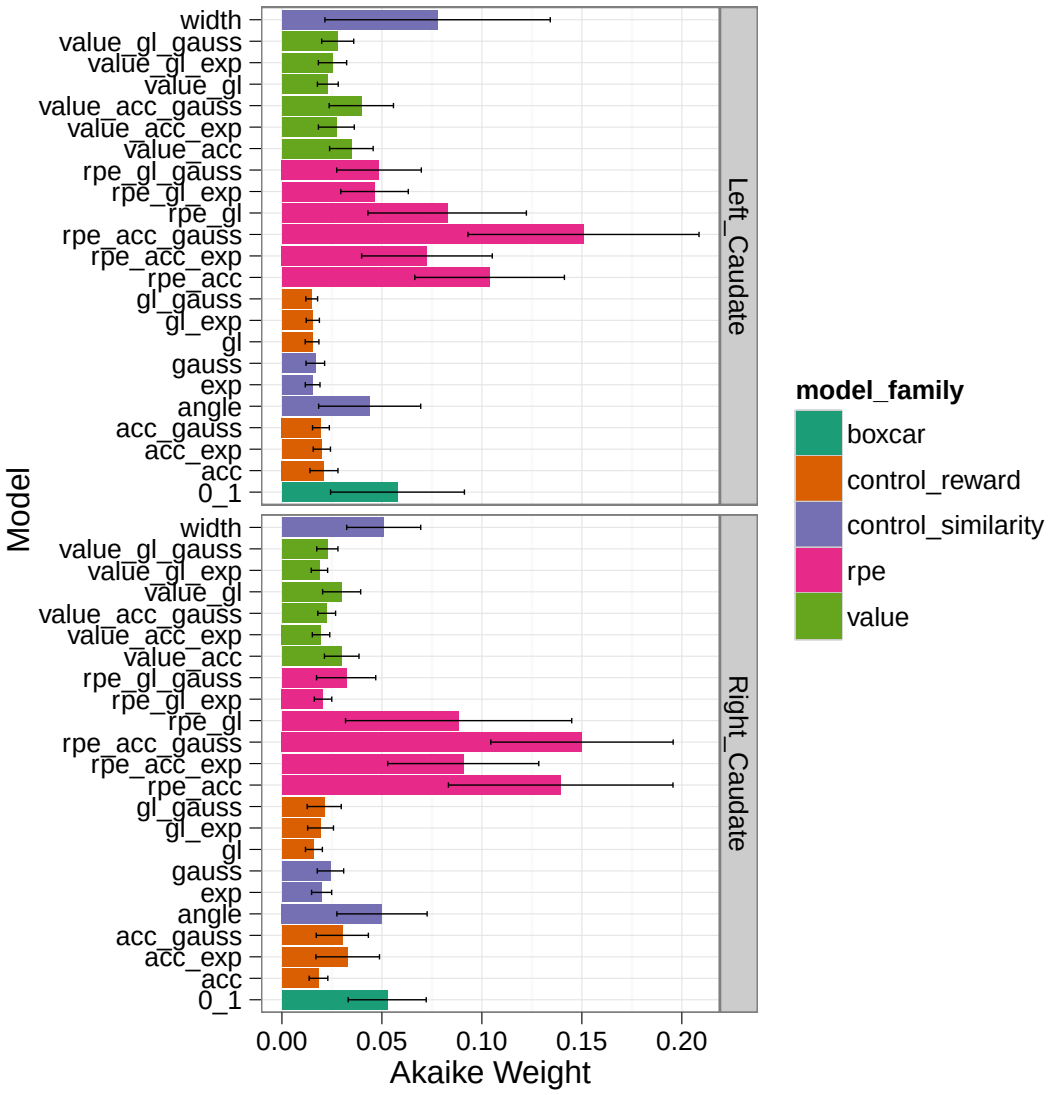

Figure 4. Dorsal caudate (left and right) – Akaike Weights for all models. Colors indicate model family (see p13 for details). Bars represent standard errors.

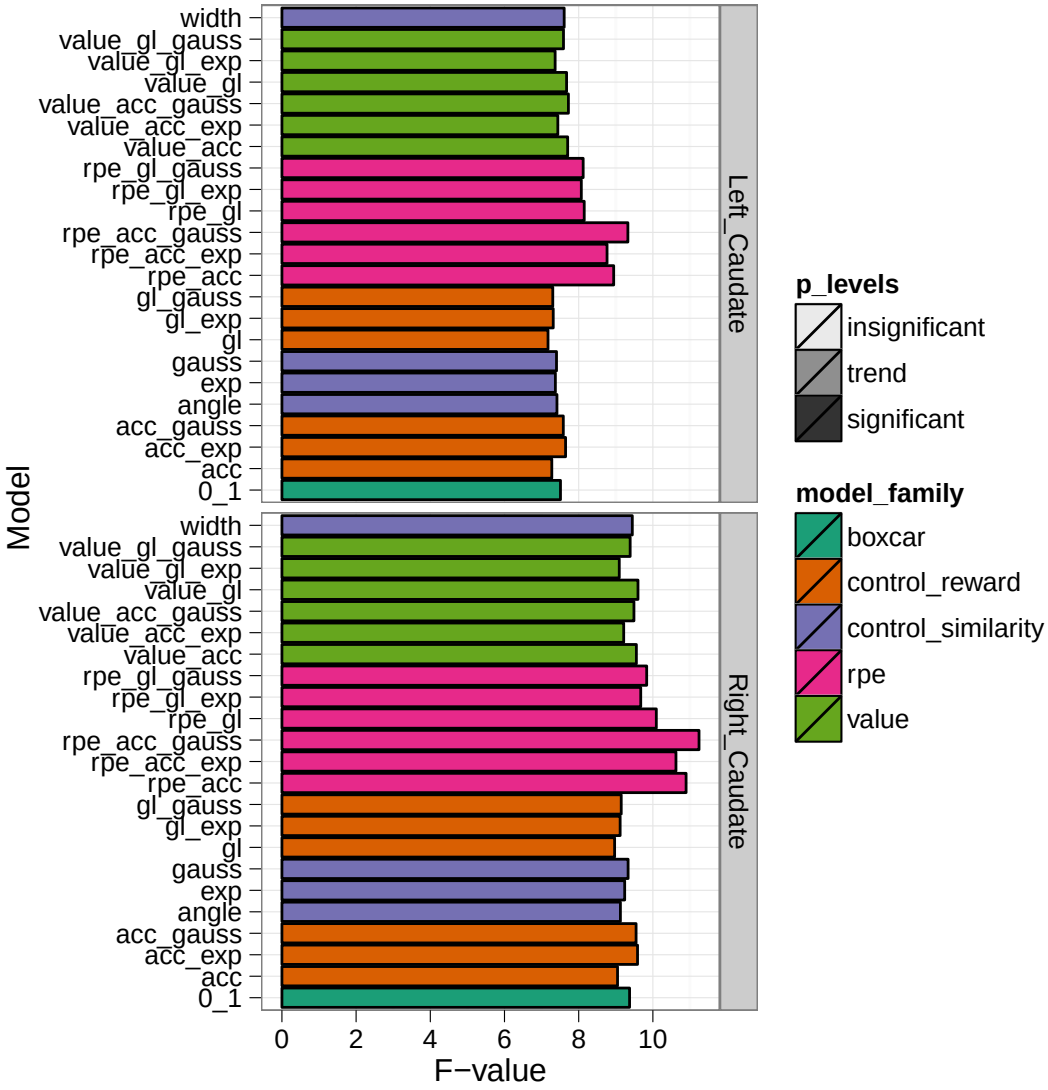

Figure 5. Dorsal caudate (left and right) – F-values for all models. Significance-level is denoted by the saturation, where the $p < 0.05$ level is significant, and trend is between $p < 0.05$ and 0.10. Colors indicate model family (see p13 for details).

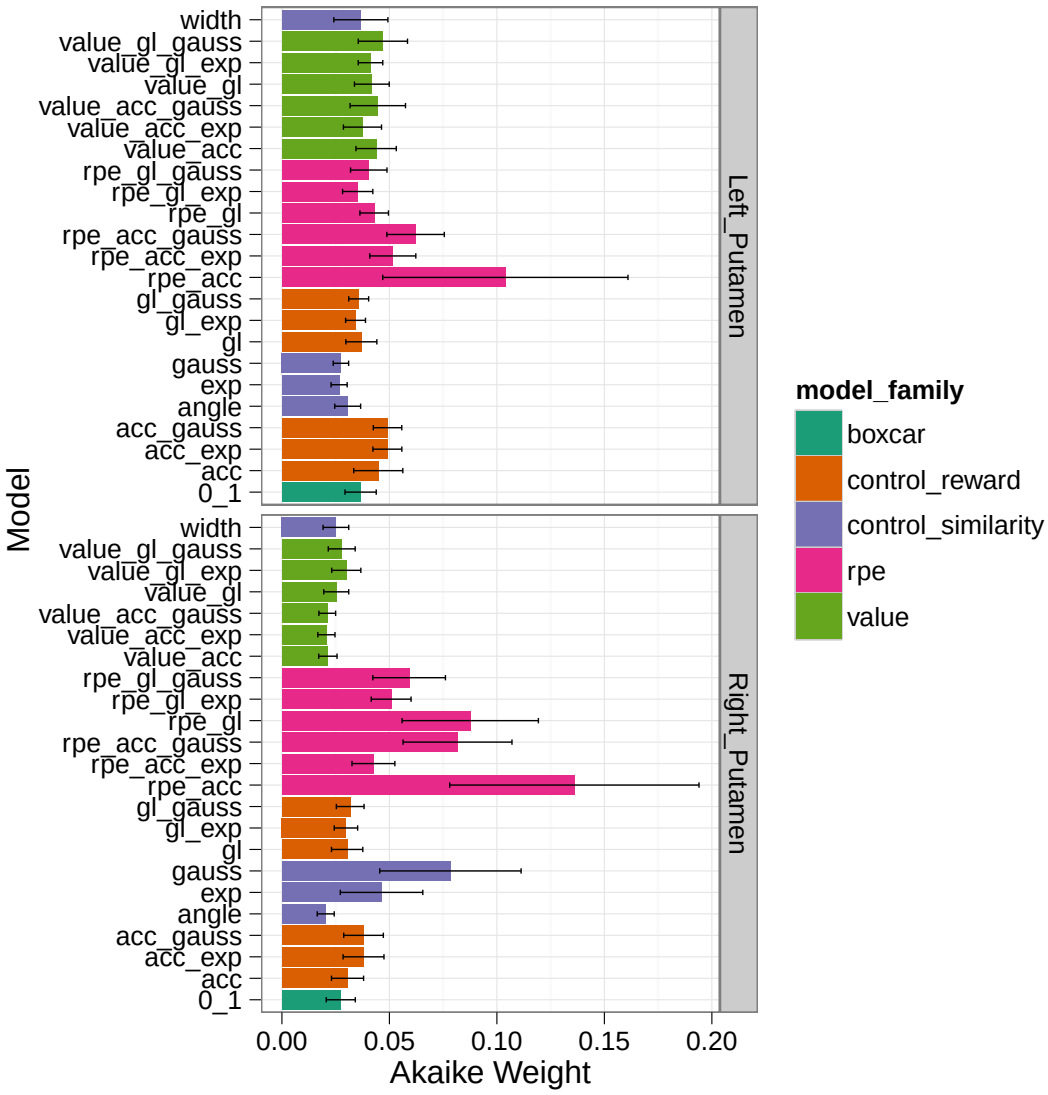

Figure 6. Putamen (left and right) – Akaike Weights for all models. Colors indicate model family (see p13 for details). Bars represent standard errors.

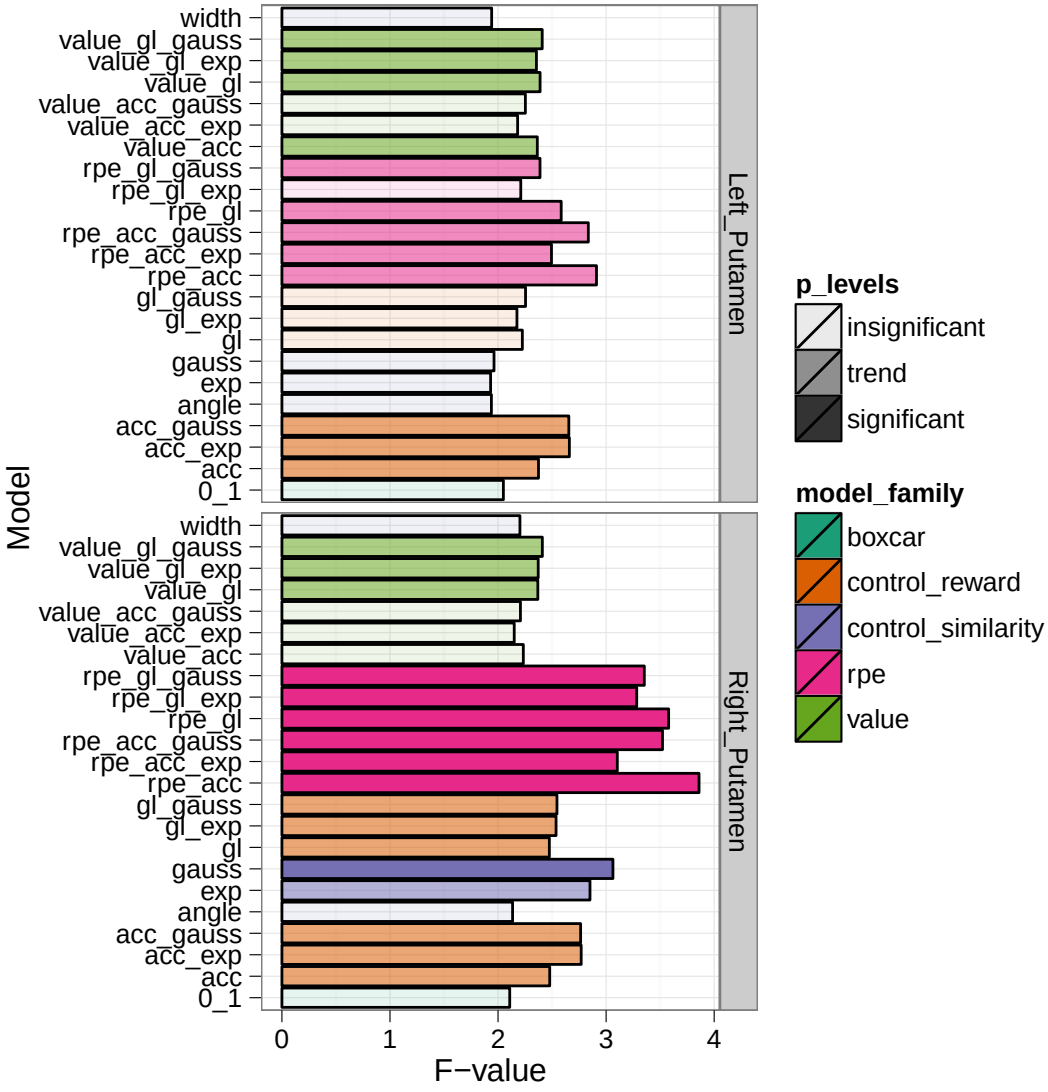

Figure 7. Putamen (left and right) – F-values for all models. Significance-level is denoted by the saturation, where the $p < 0.05$ level is significant, and trend is between $p < 0.05$ and 0.10. Colors indicate model family (see p13 for details).

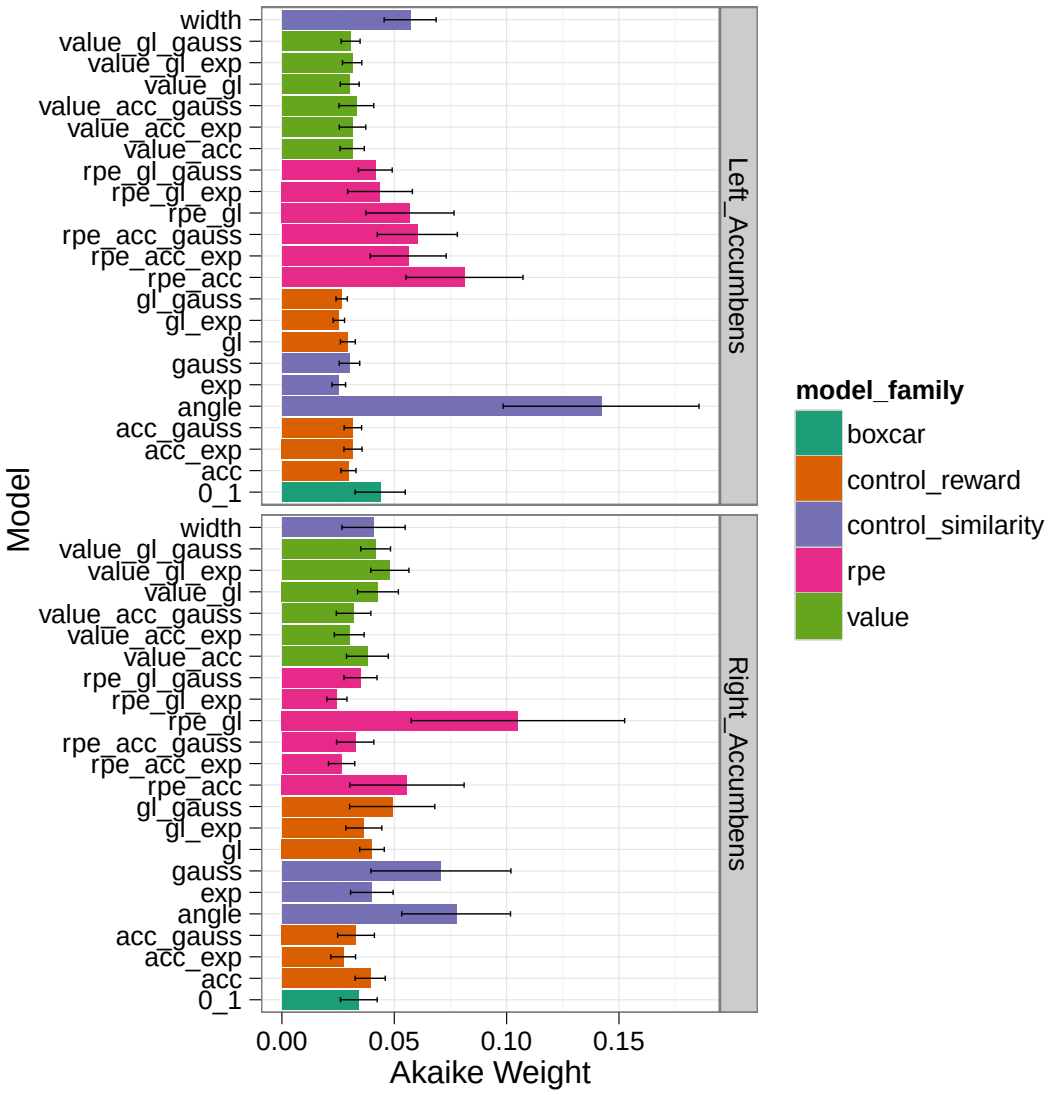

Figure 8. Nucleus Accumbens (left and right) – Akaike Weights for all models. Colors indicate model family (see p13 for details). Bars represent standard errors.

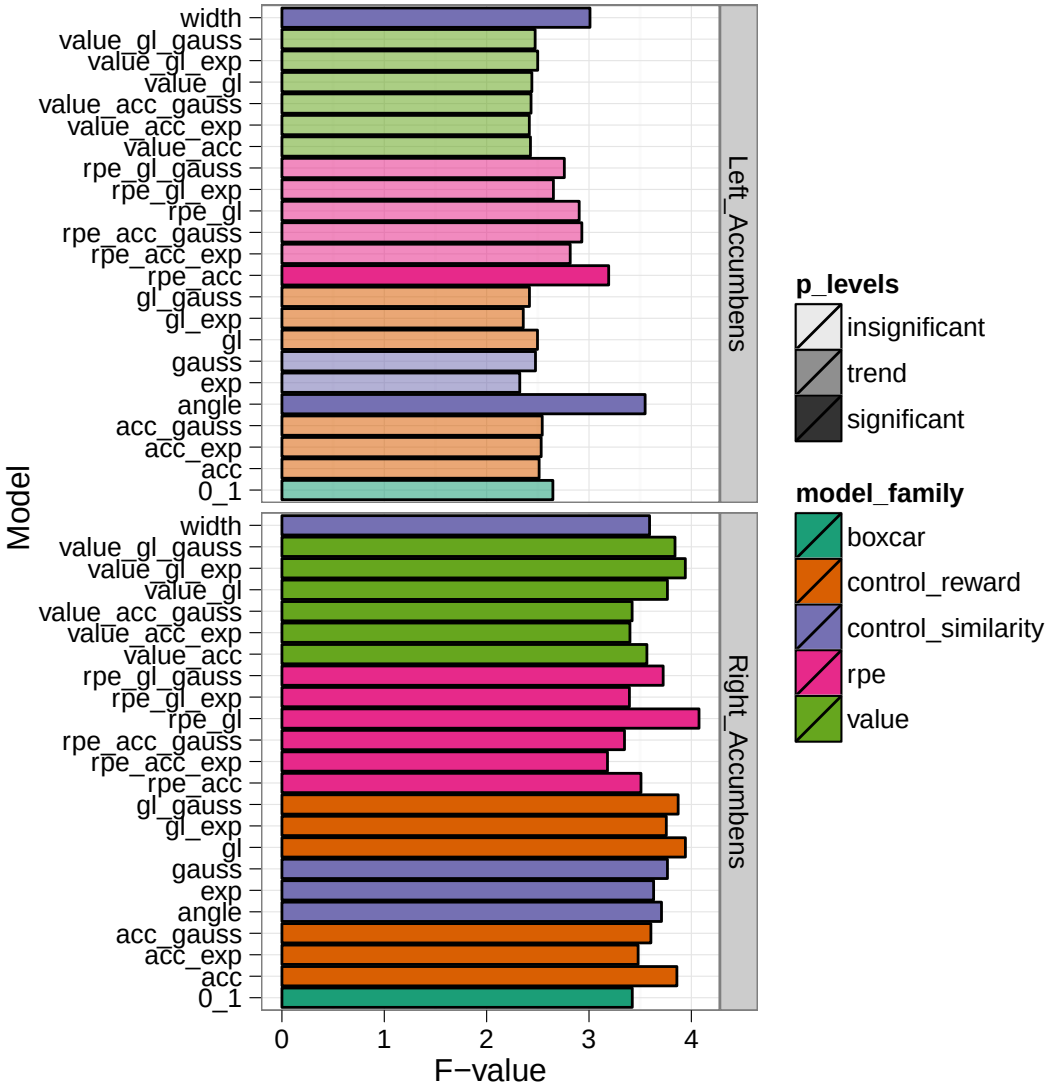

Figure 9. Nucleus accumbens (left and right) – F-values for all models. Significance-level is denoted by the saturation, where the $p < 0.05$ level is significant, and trend is between $p < 0.05$ and 0.10. Colors indicate model family (see p13 for details).

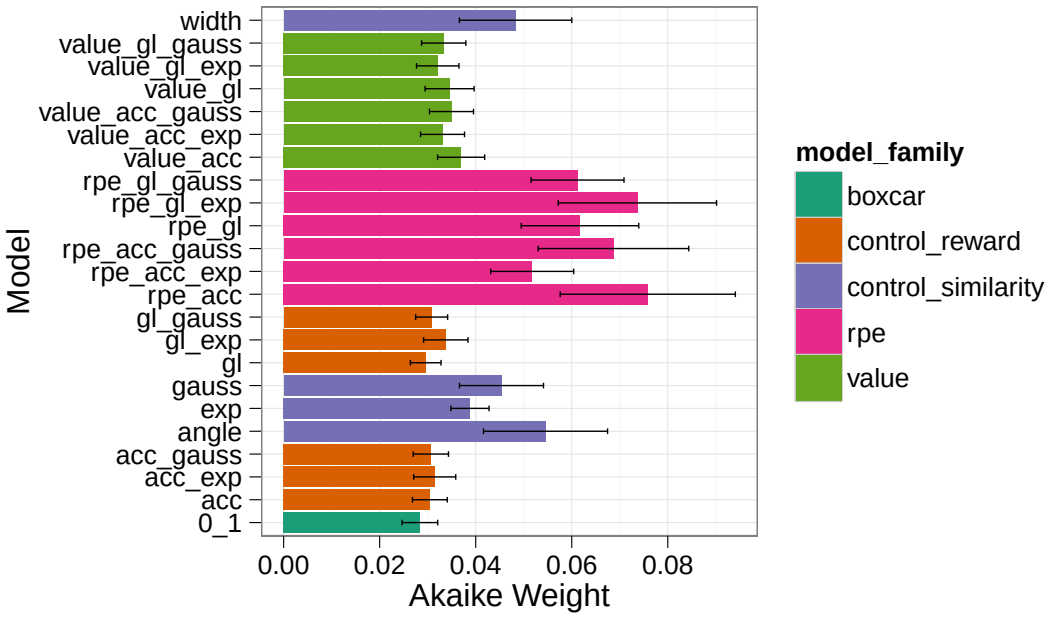

Figure 10. Insula – Akaike Weights for all models. Colors indicate model family (see p13 for details). Bars represent standard errors.

neighbor, 2.86 and 2.88 times more likely then the "boxcar"). Unlike the caudate and insula, the 2 similarity-adjustment models ("rpe_acc_gauss" and "rpe_acc_exp") were consistently more informative than the reward unadjusted ("rpe_acc"). Looking at the F-tests, both regions, especially in the "rpe" family were, were reasonably significant (Figure 13 and 15).

TODO – The remaining three cortical bits.... Some of them are weird.

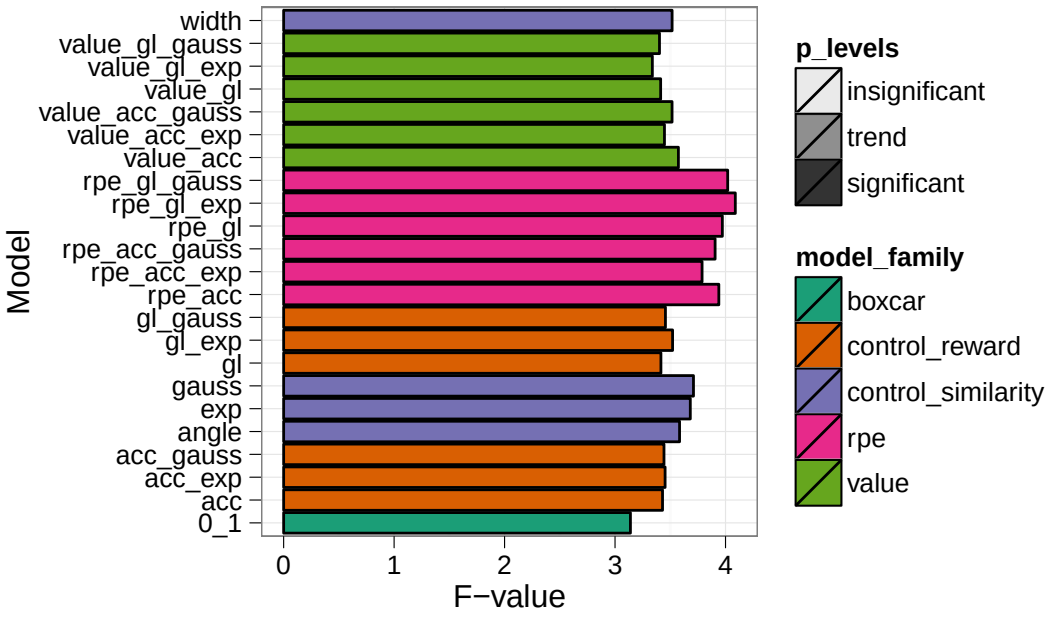

Figure 11. Insula – F-values for all models. Significance-level is denoted by the saturation, where the $p < 0.05$ level is significant, and trend is between $p < 0.05$ and 0.10. Colors indicate model family (see p13 for details).

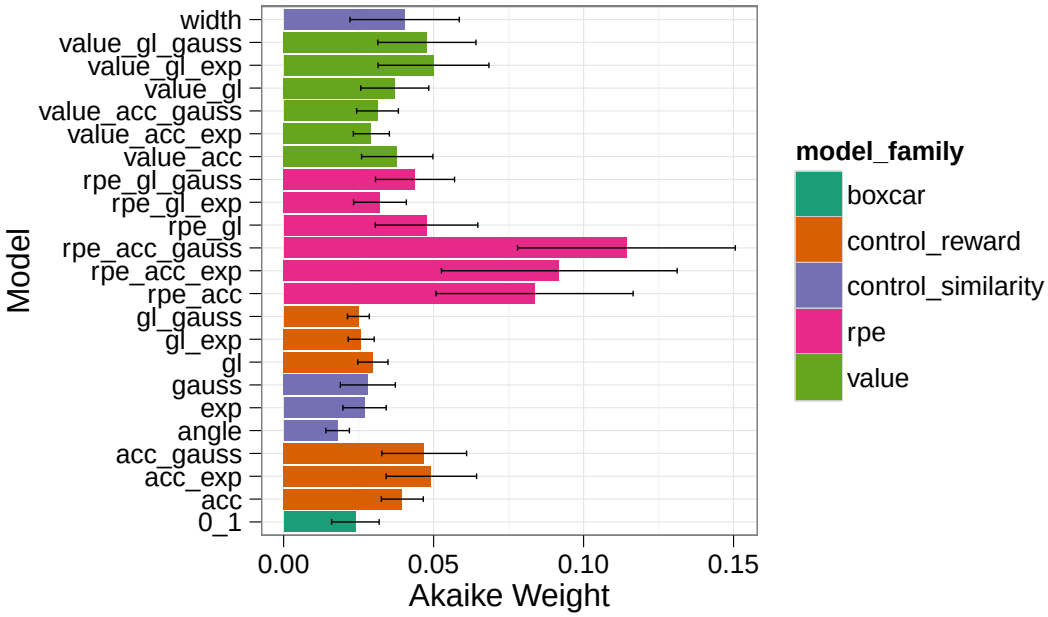

Figure 12. ACC – Akaike Weights for all models. Colors indicate model family (see p13 for details). Bars represent standard errors.

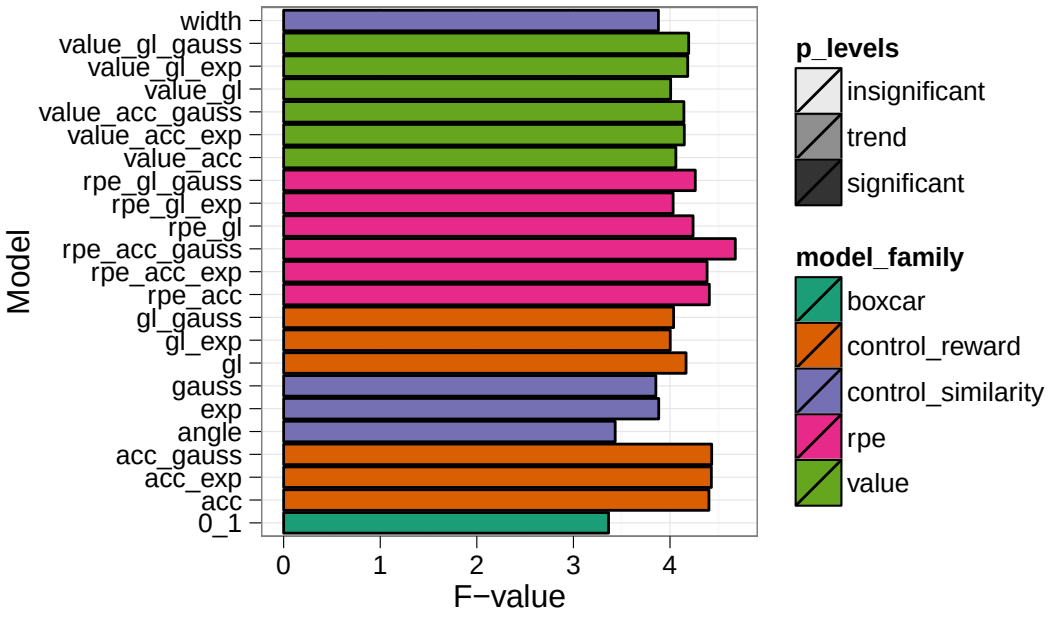

Figure 13. ACC – F-values for all models. Significance-level is denoted by the saturation, where the $p < 0.05$ level is significant, and trend is between $p < 0.05$ and 0.10. Colors indicate model family (see p13 for details).

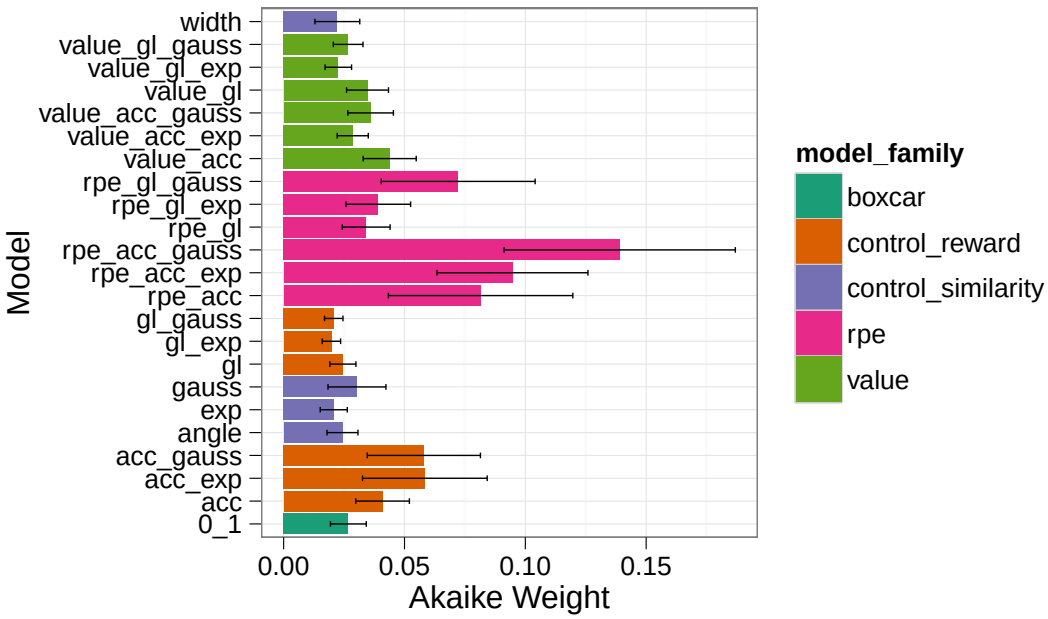

Figure 14. PCC – Akaike Weights for all models. Colors indicate model family (see p13 for details). Bars represent standard errors.

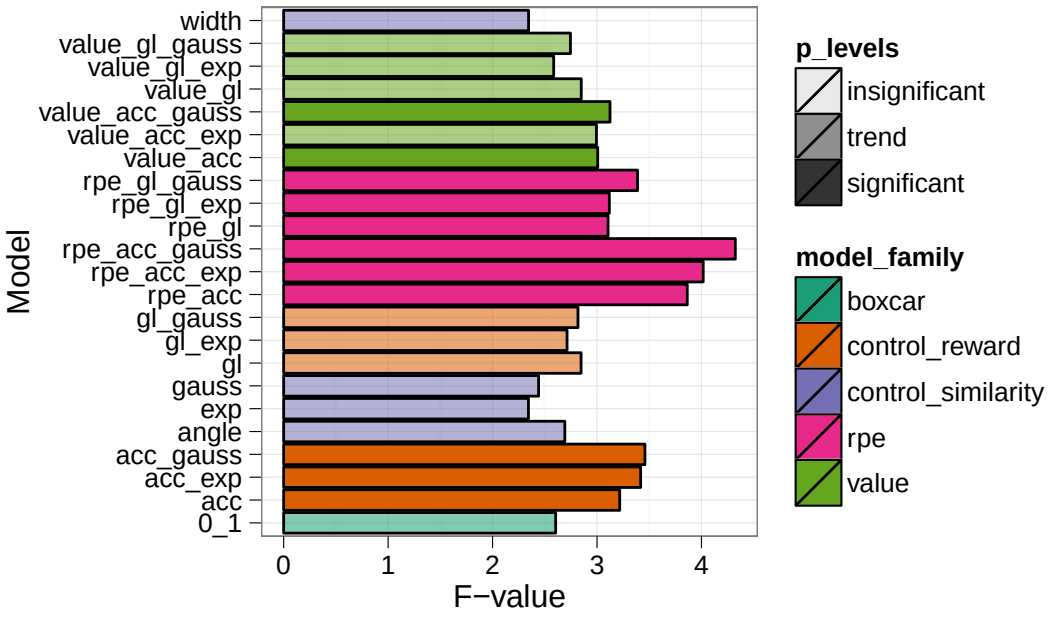

Figure 15. PCC – F-values for all models. Significance-level is denoted by the saturation, where the $p < 0.05$ level is significant, and trend is between $p < 0.05$ and 0.10. Colors indicate model family (see p13 for details).

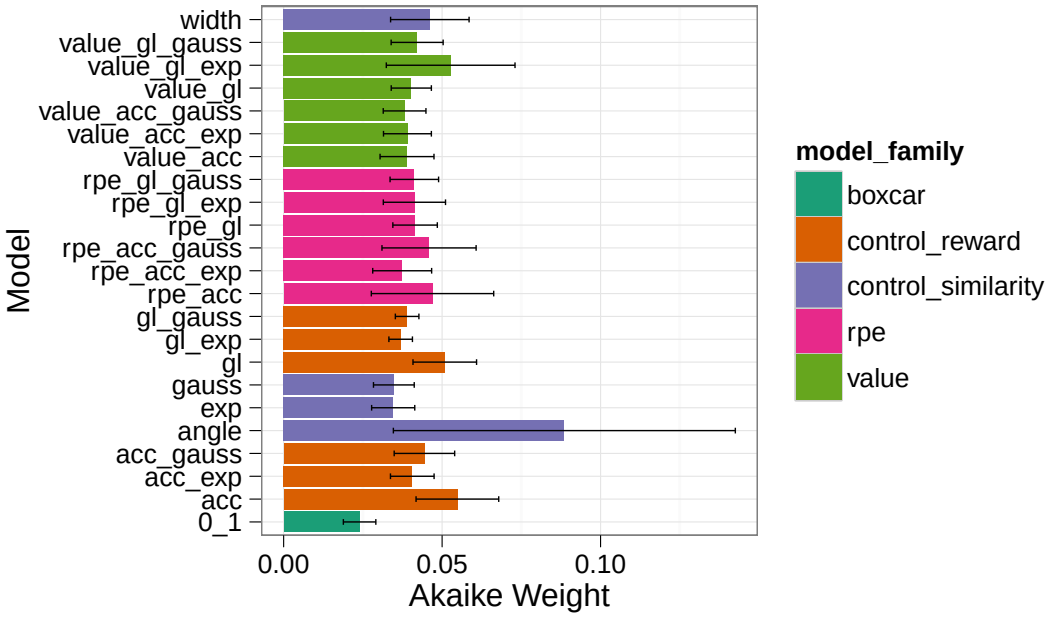

Figure 16. Frontal (ventral) medial PFC – Akaike Weights for all models. Colors indicate model family (see p13 for details). Bars represent standard errors.

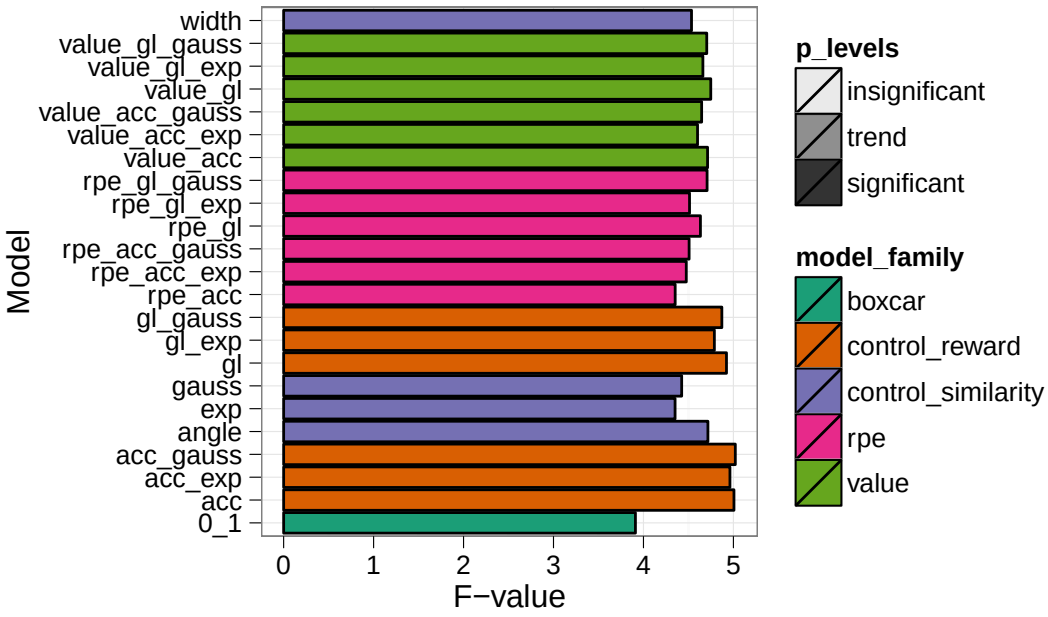

Figure 17. Frontal (ventral) medial PFC – F-values for all models. Significance-level is denoted by the saturation, where the $p < 0.05$ level is significant, and trend is between $p < 0.05$ and 0.10. Colors indicate model family (see p13 for details).

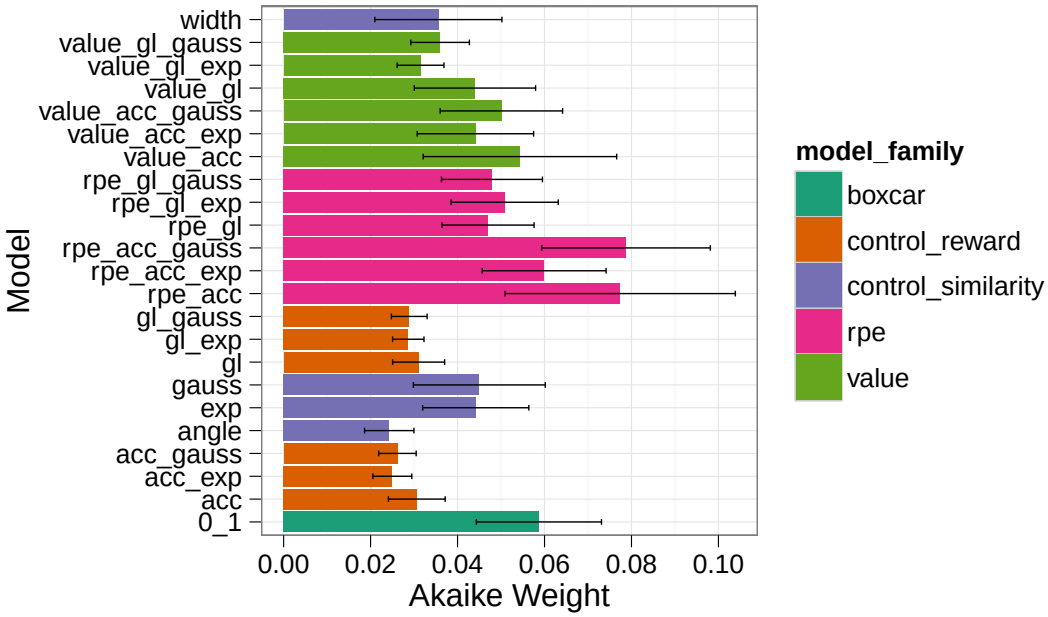

Figure 18. Orbital frontal cortex – Akaike Weights for all models. Colors indicate model family (see p13 for details). Bars represent standard errors.

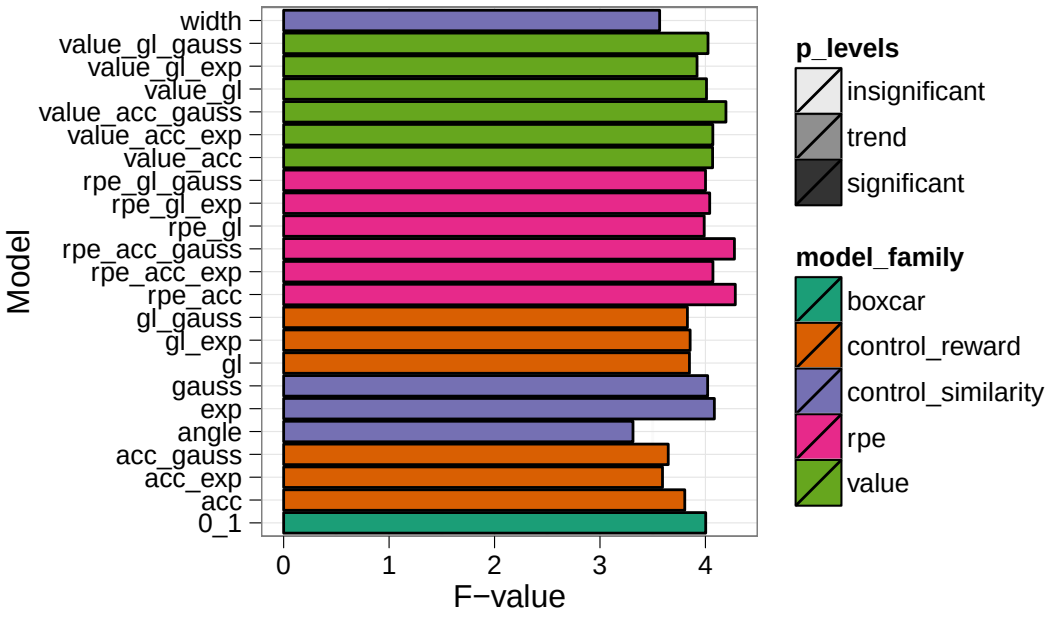

Figure 19. Orbital frontal cortex – F-values for all models. Significance-level is denoted by the saturation, where the $p < 0.05$ level is significant, and trend is between $p < 0.05$ and 0.10. Colors indicate model family (see p13 for details).

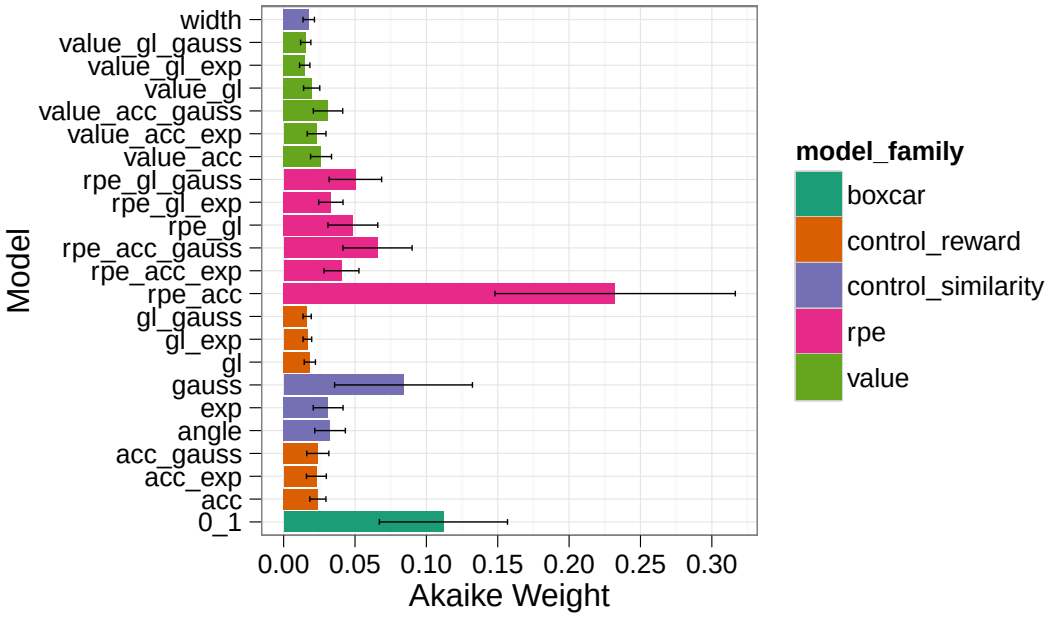

Figure 20. Middle frontal (dorsal-lateral) PFC – Akaike Weights for all models. Colors indicate model family (see p13 for details). Bars represent standard errors.

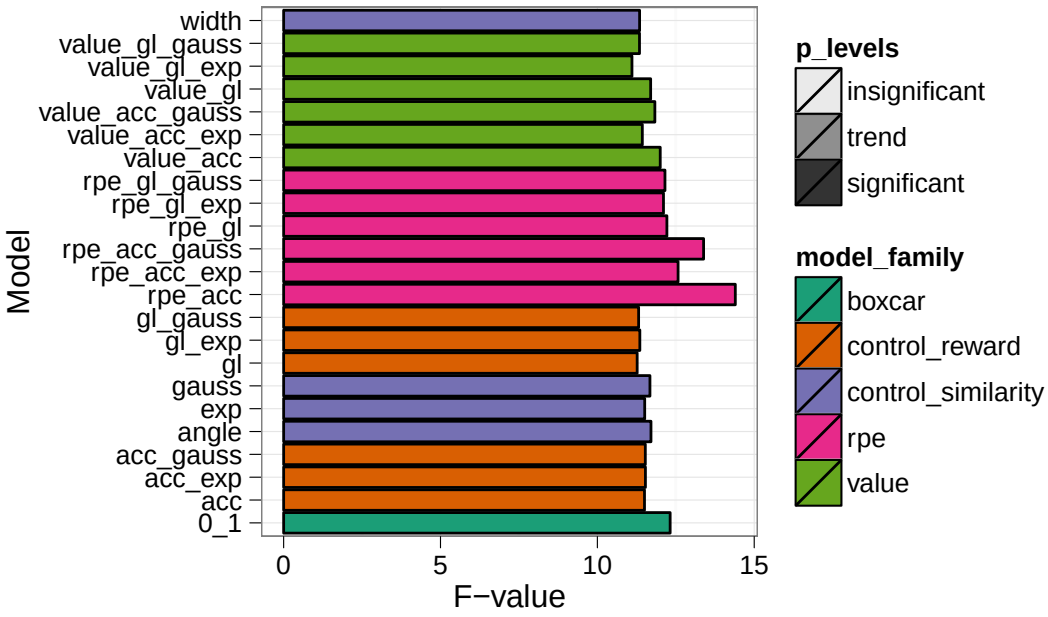

Figure 21. Middle frontal (dorsal-lateral) PFC – F-values for all models. Significance-level is denoted by the saturation, where the $p < 0.05$ level is significant, and trend is between $p < 0.05$ and 0.10. Colors indicate model family (see p13 for details).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(63), 716–723.

Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *The Journal of Wildlife Management*, *64*(4), 912–923.

Ashburner, J., & Friston, K. (1999, Jan). Nonlinear spatial normalization using basis functions. *HUMAN BRAIN MAPPING*, *7*, 254—266.

Ashburner, J., Neelin, P., Collins, D., Evans, A., & Friston, K. (1997). Incorporating prior knowledge into image registration. *Neuroimage*, *6*(4), 344–352.

Birn, R. M., Cox, R. W., & Bandettini, P. A. (2002, Jan). Detection versus estimation in event-related fmri: choosing the optimal stimulus timing. *Neuroimage*, *15*(1), 252–64.

Burnham, K. P. (2004, Nov). Multimodel inference: Understanding aic and bic in model selection. *Sociological Methods & Research*, *33*(2), 261–304.

Cincotta, C. M., & Seger, C. A. (2007, Feb). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of cognitive neuroscience*, *19*(2), 249–65.

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., & Marchal, G. (1995, Jan). Automated multi-modality image registration based on information theory. *Information Processing in Medical Imaging*, 263–274.

Dale, A. M. (1999, Jan). Optimal experimental design for event-related fmri. *Hum Brain Mapp*, *8*(2-3), 109–14.

Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et

al. (2006, Jul). An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, *31*(3), 968–80.

Forster, M. (2000, Mar). Key concepts in model selection: Performance and generalizability. *Journal of Mathematical Psychology*, *44*(1), 205–231.

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998, Jan). Event-related fmri: characterizing differential responses. *Neuroimage*, *7*(1), 30–40.

Geman, S., Bienenstock, E., & Doursat, R. (1, Jan). Neural networks and the bias/variance dilemma. *Neural Computation*, *4*(1), 1–58.

Kao, M.-H., Mandal, A., Lazar, N., & Stufken, J. (2009, Feb). Multi-objective optimal experimental designs for event-related fmri studies. *Neuroimage*, *44*(3), 849–56.

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006, Jul). Is avoiding an aversive outcome rewarding? neural substrates of avoidance learning in the human brain. *PLoS Biology*, *4*(8), e233.

Kruggel, F., Cramon, D. Y. V., & Descombes, X. (1999). Comparison of filtering methods for fmri datasets. *Neuroimage*, *10*(5), 530–543.

Liu, T. T. (2004, Jan). Efficiency, power, and entropy in event-related fmri with multiple trial types. part ii: design of experiments. *Neuroimage*, *21*(1), 401–13.

Lopez-Paniagua, D., & Seger, C. A. (2011). Interactions within and between corticostriatal loops during component processes of category learning. *Journal of cognitive neuroscience*, *23*(10), 3068–3083.

Maddox, W. T., & Bohil, C. J. (2001, Jun). Feedback effects on cost-benefit learning in perceptual categorization. *Mem Cognit*, *29*(4), 598–615.

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837–841.

McKinley, S. C., & Nosofsky, R. M. (1996, Apr). Selective attention and the formation of linear decision boundaries. *J Exp Psychol Hum Percept Perform*, *22*(2), 294–317.

Miezin, F. M., Maccotta, L., Ollinger, J. M., Petersen, S. E., & Buckner, R. L. (2000, Jun). Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage*, *11*(6 Pt 1), 735–59.

Poldrack, R. A. (2007, Mar). Region of interest analysis for fmri. *Social Cognitive and Affective Neuroscience*, *2*(1), 67–70.

Rao, C. R., Wu, Y., Konishi, S., & Mukerjee, R. (2001). On model selection. *Lecture Notes-Monograph Series, Model Selection*, *38*, 1–64.

Seger, C. A., & Cincotta, C. (2005). The roles of the caudate nucleus in human classification learning. *J Neurosci*, *25*(11), 2941–2951.

Seger, C. A., & Cincotta, C. M. (2006, Nov). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cereb Cortex*, *16*(11), 1546–55.

Seger, C. A., Peterson, E. J., Cincotta, C. M., Lopez-Paniagua, D., & Anderson, C. W. (2010, Apr). Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and granger causality modeling. *Neuroimage*, *50*(2), 644–56.

Smith, K. S., Berridge, K. C., & Aldridge, J. W. (2011, Jul). Disentangling pleasure from incentive salience and learning signals in brain reward circuitry. *Proc Natl Acad Sci USA*, *108*(27), E255–64.

Wagenmakers, E.-J., & Farrell, S. (2004, Feb). Aic model selection using akaike weights. *Psychon Bull Rev*, *11*(1), 192–6.

Wager, T. D., & Nichols, T. E. (2003, Feb). Optimization of experimental design in fmri: a general framework using a genetic algorithm. *Neuroimage*, *18*(2), 293–309.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996, Jan). A unified statistical approach for determining significant signals in images of cerebral activation. *Hum Brain Mapp*, *4*(1), 58–73.