Assessing inter-individual variability in brain-behavior relationship with functional neuroimaging

Maël Lebreton[1,2,*] & Stefano Palminteri[3,4]

[1]Amsterdam Brain and Cognition (ABC), Nieuwe Achtergracht 129, 1018 WS Amsterdam, the Netherlands.
[2]Amsterdam School of Economics (ASE), Faculty of Economics and Business (FEB), Roetersstraat 11, 1018 WB Amsterdam, the Netherlands.
[3]Institute of Cognitive Sciences (ICN), University College London, WC1N 3AR, London, United Kingdom.
[4]Laboratoire de Neurosciences Cognitives (LNC), INSERM U960, École Normale Supérieure, 75005 Paris, France.
[*] To whom correspondence should be addressed (m.p.lebreton@uva.nl)

Investigating inter-individual differences in brain-behavior relationships is fundamental to decipher the neural substrate of cognition, and to realize the full potential of neuroimaging applications. In this context, accurately assessing the statistical dependencies between inter-individual differences in behavior and inter-individual differences in neural activity is essential. In the present perspective we consider two hypotheses: 1) BOLD signal scales linearly with behavioral variables across individuals and 2) BOLD signal encodes behavioral variables on a similar scale across individuals. We formally show that these two hypotheses produce opposite brain-behavior correlational results in group-level analyses. We empirically explore these hypotheses in four fMRI studies, and find that, regarding the representation of values in the prefrontal cortex, the normalization hypothesis dominates. Independently from the generalizability of these findings, our results illustrate the importance of explicitly testing the scaling law between brain signals and behavioral variables before engaging in the study of functional inter-individual differences.

Introduction. "*There is very little difference between one man and another; but what little there is, is very important.*"[i] They are two main distinguishable and complementary ultimate goals in cognitive neuroscience: understanding the average – *typical* - brain and linking its structure and functions with cognition and behavior, and understanding how individuals differ from each-others from the normal to the pathological ranges. With respect to these two quests, inter-individual differences in brain-behavior relationships are fundamentally important either because they constitute a statistical challenge to understand the typical brain, or because they represent the very object of interest (Braver et al., 2010; Gabrieli et al., 2015).

Functional neuroimaging constitutes a powerful tool to investigate the neurobiological underpinnings of cognition. A large fraction of the cognitive neuroscience literature combines the use of the fMRI modality and of a behavioral task to uncover the neurobiological bases of behavior. Individuals might be differentiated according to external heterogeneity factors (such as the diagnostic criterion for some pathology, psycho-socio-economic measures: "traits"), or to behavioral measures recorded during the experiment ("task performance") (**Figure.1.A**).

Investigations of inter-individual variability usually attempt to link individual difference in brain activation with these sources of heterogeneity (**Figure.1.B**). These links are then interpreted in term of neural resource mobilized to complete the task – sometimes with opposite post-hoc rationalization. It is paradigmatic in the example of executive control literature, where positive associations between the activation of the frontal regions and the performance can be interpreted as an effective increase in cognitive control mobilization, whereas negative associations between the activation, in the same regions, and performance can also be interpreted as an increase in neural efficiency (Poldrack, 2015; Yarkoni and Braver, 2010).

These inconsistencies are possible in the literature because the hypotheses concerning the inter-individual relative scaling of brain activation and behavior are not explicitly stated and tested. In the present article, we remind important statistical properties of standard fMRI data analysis and how they affect the result and the interpretation of neural inter-individual variability results. More precisely we will show how inter-individual variability results are crucially dependent of the underlying scaling law between the brain (Blood Oxygen Level Dependent, BOLD) signal and the behavior. We also show that two simple - and equally realistic - hypotheses about the inter-individual scaling of brain activations and behavior can lead to opposite results and conclusions. We then explicitly test

[i] William James (1897). The Importance of Individuals, In *The Will to Believe and Other Essays in Popular Philosophy*

these two hypotheses against several imaging datasets, investigating value-based decision-making (Lebreton et al., 2009, 2012, 2015; Palminteri et al., 2015).

fMRI analysis background. To investigate the neural correlates of cognitive functions, the classical fMRI analysis strategies rely on the general linear model (GLM) framework (Friston et al., 1994), and follow a *multi-level summary statistics* approach to approximate mixed-effects designs (Beckmann et al., 2003; Friston et al., 2005; Holmes and Friston, 1998; Woolrich et al., 2004; Worsley et al., 2002).

In a first step, the linear relation between the time series of BOLD signal within a specific brain region or voxel and the time series of the different experimental design variables, behavioral measures and confounds are assessed at the individual (first-) level. Thus, for each individual k, the following GLM is estimated $\mathrm{Y}_k = \beta_k.\mathbf{x}_k + u_k$ (1), where Y_k is the BOLD times-series of individual *k*, $\mathbf{x}_k$ is the time-series of one experimental factor of interest of individual *k*, and u_k is a Gaussian noise, accounting for auto-correlation, and β_k is then the vector of the unstandardized linear coefficient of regressions to be estimated.

First-level summary statistics, i.e. estimated individual betas $\widehat{\beta_k}$ or contrasts of betas, are then used in a population (second-) level analysis, to address the research question at stake. This corresponds to a second-level GLM estimation $\beta = \beta_G.\mathbf{x}_G + u_G$ (2), where β is the vector or matrix of first-level betas $\widehat{\beta_k}$, β_G is the vector of final group-level parameters, $\mathbf{x}_G$ is the group design matrix and u_G specifies the group-level residuals. In the most typical case, $\mathbf{x}_G$ is a vector of ones, which corresponds to what is referred to as random effect – a "one-sample t-test". However, $\mathbf{x}_G$ can also contain categorical information about different conditions and groups of subject -in cases one wants e.g. to test differences between different pathological or non-pathological sub-populations with a "two-sample t-test" and/or ANOVAs (**Figure.1.B.**). Finally, $\mathbf{x}_G$ can also include individual parametric covariates, whose associations with brain activations translates into second-level "multiple regression".

Overall, this approach allows a great flexibility in the design of first and second-level analyses. This two-step summary statistics scheme has been also demonstrated to be a good approximation of full mixed-model analyses (Beckmann et al., 2003; Friston et al., 2005), to be robust in respect to violation of sphericity assumption –i.e. differences in first-level variance levels (Friston et al., 2005; Mumford and Nichols, 2009)-, and to produce very comparable results across the different implementations used by the existing software packages (Bennett and Miller, 2010; Gold et al., 1998; Morgan et al., 2007). Thanks to this practical simplicity and robustness, and notwithstanding notable exceptions (Haxby, 2012; Haxby et al., 2001; Haynes and Rees, 2006), the GLM approach has therefore become the dominant one in fMRI research

Leveraging the two steps summary statistics approach, investigations of inter-individual differences have commonly used first-level contrasts of parameter estimates in inter-individual correlations or multiple regressions across individuals (often referred-to as brain-behavior correlations in the literature), in order to account for a heterogeneity factor of interest (**Figure.1.B.**).

First-level statistics in multiple regressions regression. Of primary importance for this paper is the fact that the two steps summary statistics scheme described in the preceding paragraph relies on *unstandardized* first-level betas -$\widehat{\beta_i}$- in second level analyses. Using a very intuitively helpful notation (Cohen et al., 2013), we can write, for each independent variable $\mathbf{x}_i$, $\widehat{\beta_i} = \frac{\sigma(\mathrm{Y})}{\sigma(\mathbf{x}_i)}\rho(\mathbf{x}_i,\mathrm{Y})\sqrt{VIF_i}$ (4). Here, $\rho(\mathbf{x}_i,\mathrm{Y})$ is the semi-partial correlation between $\mathbf{x}_i$ and Y, and $VIF_i = \frac{1}{1-R^2{}_{\mathbf{x}_i\mathbf{x}_j}}$, (with $R^2{}_{\mathbf{x}_i\mathbf{x}_j}$ indexing the variance explained by a regression with $\mathbf{x}_i$ as a dependent variable, and all $\mathbf{x}_{j,j\neq i}$ as independent variables) is the Variance Inflation Factor, which quantifies $\widehat{\beta_i}$ over-estimation due to multicollinearity issues (i.e. due to the correlations between $\mathbf{x}_i$ and the other independent variables $\mathbf{x}_{j,j\neq i}$). Hence, a fundamental property of $\widehat{\beta_i}$ is that their value is proportional to linear dependency between the dependent and the independent variable – $\rho(\mathbf{x}_i,Y)$-, but also to the ratio of their standard deviation $\frac{\sigma(\mathrm{Y})}{\sigma(\mathbf{x}_i)}$ – i.e. to the scaling of those variables. $\widehat{\beta_i}$ then quantifies the change in Y (in Y unit) for an increase of 1 $\mathbf{x}_i$ (in $\mathbf{x}_i$ unit).

Outside the neuroimaging community, it is common to also report and use *standardized* betas $\widehat{b_i}$. Those are computed with normalized –or Z-scored- dependent Y^Z and independent $\mathbf{x}_i{}^Z$ variables: $\mathbf{x}_i^Z = \frac{\mathbf{x}_i-\mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}$ and $\mathrm{Y}^Z = \frac{Y-\mu(Y)}{\sigma(Y)}$. This implies $\sigma(\mathbf{x}_k{}^Z) = \sigma(Y^Z) = 1$, and

therefore $\widehat{b_i} = \rho(\mathbf{x}_i, Y)\sqrt{VIF_i}$. $\widehat{b_i}$ quantifies the change in Y (in Y standard deviation) for an increase of 1 $\mathbf{x}_i$ (in $\mathbf{x}_i$ standard deviation).

Finally, one can compute $\widehat{t_i}$, the *Student t-statistic* of $\widehat{\beta_i}$, relative to the null hypothesis $\widehat{\beta_i} = 0$. Assuming a Gaussian noise u in (1), we have $\widehat{t_i} = \frac{\widehat{\beta_i}}{s(\widehat{\beta_i})}$. Here, $s(\widehat{\beta_i}) = \sqrt{\hat{\sigma}}\sqrt{VIF_i}\frac{\sigma(\mathrm{Y})}{\sigma(\mathbf{x}_i)}$ is the standard error of the estimate $\widehat{\beta_i}$, and $\hat{\sigma} = \frac{SSE}{n-p}$, (where *n* is the sample size, p is the number of coefficients in the model including intercept, and SSE is the sum of squared errors) estimates σ^2, the variance of the errors of the regression model. Hence, one can easily show that $\widehat{t_i} = \frac{\rho(\mathbf{x}_i, Y)}{\hat{\sigma}}$, or in other words, that $\hat{t}_i$ only depends on the linear dependency between the dependent and the independent variable, and on the overall quality of the regression model. $\widehat{t_i}$ follow a Student's t-distribution with $(n-p)$ degrees of freedom (n = sample size, p = number of coefficients in the model – including intercept), from which the P-value corresponding to the null hypothesis $\widehat{\beta_i} = 0$ can be computed.

Z-values of $\widehat{\beta_i}$, which are sometimes preferred to *t-values* because they are independent of the sample size (number of degrees-of-freedom), are typically re-computed from these P-values.

The goal of this section was to recall that standardized and unstandardized betas indexes different quantities, and bear different meanings. Besides, it intended to stress the importance of considering the dependence of unstandardized betas $\widehat{\beta_i}$ on $\frac{\sigma(\mathrm{Y})}{\sigma(\mathbf{x}_i)}$. In the following, we will refer to statistical relationships between $\sigma(\mathrm{Y})$ and $\sigma(\mathbf{x}_i)$ as *scaling laws*, and investigate how they impact assessments and interpretations of inter-individual differences in brain activity as indexed by unstandardized betas.

Scaling laws impact the assessment of inter-individual differences. In the following examples, we illustrate this crucial point by considering two opposite, though plausible scaling laws between a behavioral measure and the BOLD signal, which are neuro-biologically plausible and that we called the *proportional* and the *normalization* hypotheses. We consider an idealized situation, where the BOLD signal in a brain region/voxel (Y) encodes a behavioral parametric measure of interest ($\mathbf{x}_k$) whose distribution ($\sigma(\mathbf{x}_i)$) varies between individuals due to a heterogeneity source (see **Figure.1.A** and **B**). We also assumes that $\rho(\mathbf{x}_i, \mathrm{Y})$ and $\sqrt{VIF_i}$ are independent of $\sigma(\mathbf{x}_i)$ and $\sigma(\mathrm{Y})$, i.e. that the quality of this encoding is similar across subjects and does not depend on the individual brain activation or behavioral variable or range.

The *proportional* hypothesis represents the typical case on which are based random-effects analyses in neuroimaging: the brain-behavior linear relation captured by the $\widehat{\beta_k}$ is kept constant across individual, notwithstanding random variations, i.e. $\widehat{\beta_k} \propto \alpha$ (5). Note that in this case, by neglecting inter-individual differences in SR_i and VIF_i (4) and (5) then imply, $\sigma(\mathbf{x}_k) \propto \sigma(\mathrm{Y}_k)$ (6), i.e. the BOLD activation scale proportionally with the behavior. As a consequence, and somehow counterintuitively, no difference in "activations" -as indexed by $\widehat{\beta_k}$- accounts for the differences in behavior $\mathbf{x}_k$ (**Figure 1.C.a**). In other words, under the proportional hypothesis, *behavioral differences are reflected by underlying differences in brain activity, but not by differences in brain activations, as revealed by the standard fMRI analysis.*

Under the *normalization* hypothesis, the BOLD signal also encodes the variable of interest, but on an identical scale across individuals, independent of the behavioral output, i.e. $\sigma(\mathrm{Y}_k) \propto \alpha$ (7). Note that in this case, (4) and (7) imply $\widehat{\beta_k} \propto \frac{\alpha}{\sigma(\mathbf{x}_k)}$ (8), i.e. the activation summary statistics is inversely correlated with the standard deviation of the behavior. Then, inter-individual or between-group differences in "activations" -as measured by $\widehat{\beta_k}$-, logically derive from differences in behavior $\mathbf{x}_k$ (Figure 1.C.b). In other words, under the normalization hypothesis, *behavioral differences are not reflected by underlying differences in brain activity, but by differences in brain activations, as revealed by the standard fMRI analysis.*

Z-scoring the behavioral variables impacts the assessment of inter-individual differences. A classical procedure to account for inter-individual differences in behavioral scaling is to normalize –Z-transform- individually the measure of interest by subtracting its original mean $\mu_{X,k}$ and dividing the resulting centered variable by its original standard deviation $\sigma(\mathbf{x}_k)$, i.e. $\mathbf{x}_k^Z = \frac{\mathbf{x}_k - \mu(\mathbf{x}_k)}{\sigma(\mathbf{x}_k)}$. Let us assume that the two scaling hypothesis are still related to the original variables, i.e. (6) and (7) still hold, but that first level analysis are conducted with normalized variables, i.e. $\sigma(\mathbf{x}_k^Z) = 1$, hence $\widehat{\beta_k^Z} = \sigma(\mathrm{Y}_k)$(9).

Under the *proportional* hypothesis, (6) and (9) imply

$\widehat{\beta_k^Z} \propto \sigma(\mathbf{x}_k)$ (10). In that case, Z-scoring induces a monotonic relationship between the activation summary statistics and the standard deviation of the behavior, which was absent before Z-scoring (**Figure 1.C.c**).

Conversely, under the *normalization* hypothesis, (7) and (9) imply $\widehat{\beta_k^Z} \propto \alpha$ (11). In that case, the Z-scoring cancels the monotonic relationship between the "activation" –as measured by $\widehat{\beta_k^Z}$- and the standard deviation of the behavior, leaving no inter-individual or between-group differences (**Figure 1.C.d**).

These simple derivations and their consequences should carry an important message: Z-scoring fMRI regressors does not alleviate the concern about the relative scaling of behavior and brain-signal. Quite the contrary, we showed that typical inter-individual correlations significance largely depends on the interaction between the behavioral variable pre-processing and the scaling hypothesis. Therefore, one should be extremely careful when interpreting inter-individual (between-group) differences in unstandardized betas, and greater care should be put to the assessment of the underlying scaling hypothesis, and to the description of the experimental procedure, especially concerning the Z-scoring of the behavioral predictors.

Z-scoring behavioral variables impacts random effect statistics. The interaction between the scaling hypothesis and the Z-scoring of the behavioral variable also have a significant impact on the quality of typical second-level random-effect analyses (i.e. one sample t-test on the $\widehat{\beta_k}$): under the proportional scaling hypothesis, Z-scoring $\mathbf{x}_k$ introduces a systematic behavioral variance in the $\widehat{\beta_k^Z}$ due to the dependence of $\widehat{\beta_k^Z}$ to $\sigma(\mathbf{x}_k)$ -see (5) vs (10)-, which can decrease the significance of random-effect model (see **Figure 1.C.a** vs **1.C.c**). On the opposite, under the normalization scaling hypothesis, Z-scoring $\mathbf{x}_k$ erase the systematic behavioral variance initially due to the difference in $\sigma(\mathbf{x}_k)$ -see (7) vs (11)-, hence it may increase the significance of random-effect model (see **Figure 1.C.b** vs **1.C.d**). Systematic investigation of brain-behavior scaling laws, in different cognitive domains and brain regions, might contribute to provide precious information about how to best pre-process variables of interest to perform classical fMRI analyses.

The importance of the design in assessing inter-individual differences. Consider an investigation of the neural bases of inter-individual differences in cognitive performance. In a typical setting, participants perform a cognitive task in the fMRI scanner, where a trial-to-trial measure of a cognitive variable, like confidence, decision-value, or reaction time, is taken as a measure of behavior performance and is recorded to be used as an fMRI regressors $\mathbf{x}_k$. Participants are then differentiated according to their overall performance during the task, or screened for an external measure of heterogeneity (IQ, psychometric measure or socioeconomic status), to be linked with the neural correlates of the cognitive variable. If the task is easy, high-performing people might exhibit ceiling performances, (generating smaller $\sigma(\mathbf{x}_k)$), inducing a negative correlation between individual performance and $\sigma(\mathbf{x}_k)$. Symmetrically, if the task is hard, low-performing people might exhibit flooring performances, (generating smaller $\sigma(\mathbf{x}_k)$), inducing a positive correlation between individual performance and $\sigma(\mathbf{x}_k)$. Given the dependencies between $\widehat{\beta_k}$ and $\sigma(\mathbf{x}_k)$, this will lead to opposite correlations between "activations" ($\widehat{\beta_k}$), and "individual performances" (proxied by $\sigma(\mathbf{x}_k)$) (**Figure 2.A**).

Ignoring scaling laws can lead to wrong conclusions. First, the generic term "brain activation" does not accurately describe the statistical relations underlying inter-individual differences assessments. We suggest that it could appropriately be replaced by "unstandardized regression coefficients" "t-values" or "Z-values", depending on the statistical measure used. There is indeed a sharp difference in the interpretation of those measures: the former -unstandardized $\widehat{\beta_k}$- combines information about 1) the strength of the linear dependency between the experimental independent variable and the BOLD signal and 2) the scaling between those two measures, whereas the latter -"t-values" or "Z-values"- only inform about the strength of the linear dependency between the experimental independent variable and the BOLD signal. Taking into account this important distinction would help clarify some confusions existing in the literature about two different questions: is the brain region linearly coding the variable on a *different scale* in different subjects? Or: is the region linearly coding the variable with a different *reliability* in different subjects?

Second, the interpretations of inter-individual differences in $\widehat{\beta_k}$ should be made with caution. For instance, let's assume that the activity in a region of interest (ROI) Y_k is causally responsible for a behavioral measure $\mathbf{x}_k$, and that a heterogeneity factor causes changes in Y_k, inducing proportional changes in $\mathbf{x}_k$. In the proportional context, comparisons between unstandardized betas in the

ROI might be inconclusive, potentially misleading to false-negative conclusions about the role of the ROI in the observed inter-individual differences due to the heterogeneity factor (**Figure.2.B**).

Investigating BOLD-behavior scaling laws in parametric value-rating fMRI studies. In this section, we assess the practical impact of BOLD-behavior scaling laws on fMRI analysis –random effects, and inter-individuals correlations (see also **Supplementary Material**). Experimental data consists in three published fMRI datasets investigating "values" –the presumed determinant of decision-making (Camerer, 2008; Rangel et al., 2008). Functional neuroimaging measures were recorded while subjects were performing similar tasks (**Figure 3.A** and **Figure Supp. 1.A**): judging the pleasantness of pictures of paintings, houses and faces (Study 1), the desirability of objects depicted in short videos (Study 2) or the desirability of events described in sentences (Study 3), and reporting those evaluations on a rating scale (Lebreton et al., 2009, 2012, 2015). As previously and extensively reported (Bartra et al., 2013; Clithero and Rangel, 2014; Peters and Büchel, 2010; Sescousse et al., 2013), we found that random-effect analyses on the parametric independent variable "value rating" ($\mathbf{v}$) elicit strong activations in a large ventral prefrontal region, including ventromedial prefrontal cortex (VMPFC) and medial orbitofrontal cortex (MOFC) (**Figure 3.B** and **Figure Supp. 1.B**).

First, taking into account both the size and the p-values of the VMPFC cluster correlating with values, we found that the random-effect analysis is more significant when we use the Z-scored ($\mathbf{v}^Z$) rather than the native independent variable *values* ($\mathbf{v}$) (**Figure 3.C** and **Figure Supp. 2**). These results are consistent with the normalization hypothesis and provide important clues about how to pre-process the value variable for future fMRI studies.

Second, in order to formally assess the two scaling laws in this context, we next tested the correlations derived in the preceding sections -equations (5), (8) (10) and (11)- involving "activation", as measured by $\hat{\beta}$ extracted from an anatomical VMPFC ROI, and the standard deviation of the value $\sigma(\mathbf{v})$. The results all suggest that the inter-individual representation of values in the VMPFC, in such rating tasks, follows a normalization scaling rule (**Figure 3.C** and **Figure Supp. 1.C**). We also checked that the correlations do not hold when using the estimated $\hat{t}$ statistics, such that the identified differences in $\hat{\beta}$ are rather due to scaling, than to coding -i.e. differences in the linear dependencies between BOLD and behavior (**Figure Supp. 2**). Overall, all results provide support for a normalization scaling law in the VMPFC in value-rating tasks: i.e. despite individual differences in the range (variance) of the behavioral value ratings, individuals exhibit similar range of BOLD signal in the core of the brain valuation system. This fact can be given concurrent interpretations: 1) the "true" underlying value signal range is actually captured by the fMRI analysis and is similar across individuals despite individual differences in the behavior –mostly due to differences in the calibration on the experimental rating scale; or 2) the underlying "true" value signal range is actually different across individuals -following the differences in the range of ratings reported it on the experimental scale-, but there are experimental limitations [ii] which prevent the correct assessment of this inter-individual variability. This raise new questions –e.g.: can we infer whether an option is more valuable to an individual than to another? -, whose answers will determine our ability to fulfill some of the promises of fMRI applications.

Implications for model-based fMRI. Model-based fMRI typically use as dependent variables in first-level GLMs ($\mathbf{x}_k$) *latent variables* derived from individual choice patterns (**Figure.4.A**): a computational model is selected, its free-parameters are adjusted to account at best for behavioral data, and the model parameters are used to generate the *latent variables* of interest $\mathbf{x}_k$ (O'Doherty et al., 2007). Importantly, the model free-parameters can be either considered as fixed -i.e. shared across individuals- or random-effects -i.e. each individual's parameters are drawn from a common population distribution (Daw, 2011). In the case of random effects, model free-parameters often control the latent variable distribution parameters, i.e. $\sigma(\mathbf{x}_k)$. We intuitively illustrate this link for several models used in the value-based decision-making literature (**Figure.4.B**). It must be noted, however, that despite the apparent simplicity of the relationships depicted in **Figure.4.B**, these relations are not trivial and largely depend on the task setting and the stimuli space (**Figure.4.C**). This complex link between model-parameters and $\sigma_{X,k}$ have several implications with regard to inter-individual differences in brain-behavior relationships:

First, although treating model free-parameters as

[ii] These limitations can be of several natures: biological (e.g. difference in vascularization), physical (e.g. sensitivity of MRI gradient to inter-individual differences), and analytic strategies (e.g. pre-processing of MRI images). See also **Figure Supp. 5.**

random-effects often seem to provide the best account of individuals' behavior as assessed by rigorous model-comparisons, a common practice in the literature is to treat them as a fixed-effect –i.e. use a population parameter- to generate the latent variables for fMRI analysis (Daw et al., 2006; Gershman et al., 2009; Gläscher et al., 2009, 2010; O'Doherty et al., 2004; Palminteri et al., 2009; Pessiglione et al., 2008). This is justified by the fact that individual free-parameter estimates are "noisy" and using the data from the full population is an efficient way to regularize them. However, when individual parameters still provide a better account of the population behavioral data according to rigorous model-comparison procedures, one might argue that the variance modelled in the individual free-parameters actually captures an important individual variance in the cognitive process at stake, hence might contribute to give a better account of individual neurophysiological data. In the light of our findings concerning value signals, an alternative explanation to the preference of population parameters in fMRI can be formulated: the use of population free-parameters actually constrains $\sigma(\mathbf{x}_k)$ to a unique population value, provided that individuals are given the same input. Under the *normalization* scaling hypothesis, this can therefore substantially increases the statistical power of subsequent second-level random effects analyses. In this case, a better way to model brain activation (i.e. accounting for individual differences) would be to use individual model free-parameters, and Z-score the latent variables generated by these individual models. This discussion, again, raises the interest of better documenting scaling-laws in fMRI, so as to provide *a priori* principled rational to process independent variables of interest, in order to increase the sensitivity and replicability of model-based fMRI.

Second, inter-individual correlations between model free-parameters and activations, as measured with $\widehat{\beta_k}$, in an ROI encoding the latent variable $\mathbf{x}_k$ should be interpreted with much caution. Indeed, they may rely on simple mathematical dependencies between the free parameters, $\sigma(\mathbf{x}_k)$, and $\widehat{\beta_k}$, and they largely depend on interactions between the underlying brain-behavior scaling hypothesis and the processing (Z-scoring) of the latent variable $\mathbf{x}_k$. Besides, given that the link between model-parameters and $\sigma(\mathbf{x}_k)$ can reverse depending on the task design (**Figure.4.C**), one can anticipate reports of opposite inter-individual correlations (positive or negative) between model-parameters (e.g. a discount factor or a learning rate) and the "neural representation" of value, as measured by $\widehat{\beta_k}$.

Investigating BOLD-behavior scaling laws in a model-based learning fMRI study. In this section, we illustrate the points raised in the previous paragraph, using a fourth experimental fMRI dataset investigating value-based learning (Palminteri et al., 2015) (see also **Supplementary Material**). Participants were faced with repeated choices between abstract stimuli, which were probabilistically paired with different outcomes (neural, reward or punishment). The goal was to learn to select the stimuli, which maximize the occurrence of reward or minimize the occurrence of punishment (**Figure 5.A**). This task can be efficiently modelled with a variant of the Rescola-Wagner reinforcement-learning rule: this implies that participants learn, by trial and error, the value (Q-values) of the stimuli and make their choices by soft-maximizing expected value (see (Palminteri et al., 2015) and **Supplementary Material** for details). Two core free-parameters of the model capture the individuals' learning dynamics and choice variance: the temperature (i.e. the slope of the softmax decision rule, which controls the stochasticity of choices –or the tradeoff between exploration and exploitation), and the learning rate (which controls how much the information from a new feedback –in the form of a prediction-error- is incorporated in the stimulus Q-value). These free-parameters are typically set to maximize the likelihood of observed choices under the considered model. A brief analysis of different model-fitting, together with an exploration of the modelling outputs reveals two main results: 1) individual choices are better accounted for by fitting individual free-parameters than by fitting a single set of population free-parameters, even after accounting for the extra complexity engendered by this procedure (parsimony) (**Figure 5.B** and **Figure Supp. 4.A**) and 2) these individual parameters (temperature and learning rate) are very robustly associated with the individual standard deviation of the model-estimated latent variable (i.e. the Q-value of the chosen option – $\mathbf{q_c}$; **Figure 5.C**).

Turning to neuroimaging data, we report that random-effect analyses on the parametric independent variable ($\mathbf{q_c}$) elicit strong activations in the VMPFC (**Figure 5.D**). In order to assess the quality of individual-level brain-behavior correlations, we extracted $\widehat{\beta_k}$-corresponding $\hat{t}_k$ from an anatomical VMPFC ROI, and noticed that these t-statistics take significantly higher values when using Q-values generated with individual ($\mathbf{q_c}$) than population parameters ($\mathbf{q_c}^P$). Paralleling behavioral results, this reveals that, in the absence of scaling issues, the BOLD

signal is better accounted for by individually-fitted behavior (see **Supplementary Material**). Then, using the size of the VMPFC cluster in voxels, and the p-values of a t-test performed on $\widehat{\beta_k}$ extracted from the anatomical VMPFC ROI, we replicate the classical observation that population fMRI random-effects are apparently more significant using native Q-values computed with population ($\mathbf{q_c}^P$) than individual parameters (X). However, we additionally found that the best fMRI random-effect model is obtained using Q-values computed with individual parameters subsequently Z-scored per individuals ($\mathbf{q_c}^Z$) (**Figure 5.E**). These results have two implications: first, the fact that random effects are more significant using $\mathbf{q_c}^Z$ than $\mathbf{q_c}$ support the normalization hypothesis -this hypothesis is further supported by testing the correlations involving $\widehat{\beta_k}$ and $\sigma(\mathbf{q_c}_k)$ (see equations (5), (8) (10) and (11) and **Figure 5.E**). Second, they suggest that scaling issues might explain the apparent contradictory observations that fMRI random-effects are more significant using $\mathbf{q_c}^P$than $\mathbf{q_c}$ despite the superiority of individually-fitted models to account for individual behavioral choices. Overall the results of these analyses advocate for the use of individual parameters in value-related model-based fMRI, together with a Z-scoring of the model-estimated latent variable –*value*- to account at best for the inter-individual normalization effect occurring in the VMPFC.

We also explored the correlations between individual model parameters and activations. Because, in our case, model free-parameters positively correlate with $\sigma(\mathbf{q_c}_k)$ and that the Normalization scaling law implies that $\widehat{\beta_k}$ scale with $1/\sigma(\mathbf{q_c}_k)$, we expected individual learning-rates and softmax-temperatures to be inversely correlated to $\widehat{\beta_k}$ in the VMPFC. Experimental data support this prediction (**Figure Supp. 3.A** and **4.B**), which raises important questions about the interpretations of correlations between model-parameters and activations –as measured by $\widehat{\beta_k}$-, as they may be dependent on the statistical relationship between model-parameters and $\sigma(\mathbf{q_c}_k)$, which depends on the task design.

Conclusion

Researchers are increasingly interested in inter-individual variability in cognitive neurosciences, in the normal and pathological ranges. Importantly, the ability to assess and predict individual differences from neural measures -neuromarkers- constitute the cornerstone of the most promising application of fMRI in society (Gabrieli et al., 2015; Wang and Krystal, 2014) In this manuscript, we explored a specific type of neuromaker: task-dependent fMRI "activations", indexed by coefficients of regression between individual behavioral variables and BOLD signal. We recalled that unstandardized coefficients of regression $\widehat{\beta_k}$ depend on the ratio of the standard deviation –scale- of the dependent and independent variables. Therefore, task-dependent fMRI neuromarkers partly reflects *scaling laws* between the BOLD signal and the behavioral variable of interest, and documenting those scaling laws is paramount to correctly interpret assessments of inter-individual differences in cognitive neuroscience.

In this manuscript, we have proposed a new taxonomy -proportional/normalization- to qualify such inter-individual brain-behavior scaling relationship. Importantly, this taxonomy is based on a formalized description of the statistical dependency between the BOLD signal and the behavioral variable, rather than on a biological/functional over-interpretation of such statistical quantities –like in the current efficiency/activation taxonomy- (Poldrack, 2015). By doing so, it aims at providing a better account of fMRI data, hence helping the building of a cumulative cognitive science, based on the falsification of precise predictions. Although we acknowledge that the present paper does not cover the full range of potential link between brain activation and behavior (**Figure Supp. 6**), we think that this new perspective might contribute to reconcile previous contradictory findings, and foster a fruitful discussion on the way to interpret and assess investigations of individual-difference in neuroimaging.

We propose that a good practice before engaging in the study of fMRI inter-individual variability is to start documenting the statistical relationship between traits of interest (individual clinical scores, psycho-social measures, model free-parameters) and the standard deviation $\sigma(\mathbf{x}_k)$ of fMRI regressors. Ideally, researchers might explicitly test brain-behavior scaling laws for the cognitive function of interest, in the brain region of interest, using their specific task – indeed, one can expect that different cognitive processes, elicited with different tasks could follow different scaling law, in different brain regions. In order to improve the reproducibility of fMRI findings, it is paramount to formulate clear a priori hypothesis about inter-individual-differences (coding/scaling) and to use an appropriate operationalization.

Finally we initiated this practice by documenting inter-individual normalization of values representation in the VMPFC using four datasets. This parallel recent findings

reporting within-individual range adaptation of value coding in the same area (Cox and Kable, 2014). This finding might contribute to improve our understanding of the valuation process, and provide principled rational to preprocess variables of interest and carry out model-based fMRI in the value-based decision-making community.

Acknowledgments:
We are thankful to Mathias Pessiglione (MP) and Giorgio Coricelli (GC) for granting us unlimited and unrestricted access to the fMRI datasets. These datasets were collected thanks to the European Research Council (ERC Starting Grant BioMotiv to MP and ERC Consolidator Grant 617629 to GC), a Research Grant from the Schlumberger Foundation to MP, and an Agence National de la Rercherche (ANR-11-EMCO-010) grant to GC. ML is supported by an EU Marie Sklodowska-Curie Individual Fellowship (IF-2015 Grant 657904), a Universiteit van Amsterdam – Amsterdam Brain and Cognition Talent Grant, and acknowledge the support of the Bettencourt-Schueller Foundation. SP is also supported by an EU Marie Sklodowska-Curie Individual Fellowship (PIEF-GA-2012 Grant 328822).

References

Bartra, O., McGuire, J.T., and Kable, J.W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. NeuroImage *76*, 412–427.

Beckmann, C.F., Jenkinson, M., and Smith, S.M. (2003). General multilevel linear modeling for group analysis in FMRI. Neuroimage *20*, 1052–1063.

Bennett, C.M., and Miller, M.B. (2010). How reliable are the results from functional magnetic resonance imaging? Ann. N. Y. Acad. Sci. *1191*, 133–155.

Braver, T.S., Cole, M.W., and Yarkoni, T. (2010). Vive les differences! Individual variation in neural mechanisms of executive control. Curr. Opin. Neurobiol. *20*, 242–250.

Camerer, C.F. (2008). Neuroeconomics: Opening the Gray Box. Neuron *60*, 416–419.

Clithero, J.A., and Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. Soc. Cogn. Affect. Neurosci. *9*, 1289–1302.

Cohen, J., Cohen, P., West, S.G., and Aiken, L.S. (2013). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (Routledge).

Cox, K.M., and Kable, J.W. (2014). BOLD Subjective Value Signals Exhibit Robust Range Adaptation. J. Neurosci. *34*, 16533–16543.

Daw, N.D. (2011). Trial-by-trial data analysis using computational models. Decis. Mak. Affect Learn. Atten. Perform. XXIII *23*, 3–38.

Daw, N.D., O'Doherty, J.P., Dayan, P., Seymour, B., and Dolan, R.J. (2006). Cortical substrates for exploratory decisions in humans. Nature *441*, 876–879.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., and Frackowiak, R.S.J. (1994). Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. *2*, 189–210.

Friston, K.J., Stephan, K.E., Lund, T.E., Morcom, A., and Kiebel, S. (2005). Mixed-effects and fMRI studies. Neuroimage *24*, 244–252.

Gabrieli, J.D.E., Ghosh, S.S., and Whitfield-Gabrieli, S. (2015). Prediction as a Humanitarian and Pragmatic Contribution from Human Cognitive Neuroscience. Neuron *85*, 11–26.

Gershman, S.J., Pesaran, B., and Daw, N.D. (2009). Human Reinforcement Learning Subdivides Structured Action Spaces by Learning Effector-Specific Values. J. Neurosci. *29*, 13524–13531.

Gläscher, J., Hampton, A.N., and O'Doherty, J.P. (2009). Determining a Role for Ventromedial Prefrontal Cortex in Encoding Action-Based Value Signals During Reward-Related Decision Making. Cereb. Cortex *19*, 483–495.

Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. Neuron *66*, 585–595.

Gold, S., Christian, B., Arndt, S., Zeien, G., Cizadlo, T., Johnson, D.L., Flaum, M., and Andreasen, N.C. (1998). Functional MRI statistical software packages: a comparative analysis. Hum. Brain Mapp. *6*, 73–84.

Haxby, J.V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. NeuroImage *62*, 852–855.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science *293*, 2425–2430.

Haynes, J.-D., and Rees, G. (2006). Decoding mental states from brain activity in humans. Nat. Rev. Neurosci. *7*, 523–534.

Holmes, A.P., and Friston, K.J. (1998). Generalisability, Random E ects & Population Inference. Neuroimage *7*, S754.

Lebreton, M., Jorge, S., Michel, V., Thirion, B., and Pessiglione, M. (2009). An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. Neuron *64*, 431–439.

Lebreton, M., Kawa, S., d'Arc, B.F., Daunizeau, J., and Pessiglione, M. (2012). Your Goal Is Mine: Unraveling Mimetic Desires in the Human Brain. J. Neurosci. *32*, 7146–7157.

Lebreton, M., Abitbol, R., Daunizeau, J., and Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. Nat. Neurosci. *18*, 1159–1167.

Morgan, V.L., Dawant, B.M., Li, Y., and Pickens, D.R. (2007). Comparison of fMRI statistical software packages and strategies for analysis of images containing random and stimulus-correlated motion. Comput. Med. Imaging Graph. *31*, 436–446.

Mumford, J.A., and Nichols, T. (2009). Simple group fMRI modeling and inference. Neuroimage *47*, 1469–1475.

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., and Dolan, R.J. (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. Science *304*, 452–454.

O'Doherty, J.P., Hampton, A., and Kim, H. (2007). Model-Based fMRI and Its Application to Reward Learning and Decision Making. Ann. N. Y. Acad. Sci. *1104*, 35–53.

Palminteri, S., Boraud, T., Lafargue, G., Dubois, B., and Pessiglione, M. (2009). Brain Hemispheres Selectively Track the Expected Value of Contralateral Options. J. Neurosci. *29*, 13465–13472.

Palminteri, S., Khamassi, M., Joffily, M., and Coricelli, G. (2015). Contextual modulation of value signals in reward and punishment learning. Nat. Commun. *6*.

Pessiglione, M., Petrovic, P., Daunizeau, J., Palminteri, S., Dolan, R.J., and Frith, C.D. (2008). Subliminal Instrumental Conditioning Demonstrated in the Human Brain. Neuron *59*, 561–567.

Peters, J., and Büchel, C. (2010). Neural representations of subjective reward value. Behav. Brain Res. *213*, 135–141.

Poldrack, R.A. (2015). Is "efficiency" a useful concept in cognitive neuroscience? Dev. Cogn. Neurosci. *11*, 12–17.

Rangel, A., Camerer, C., and Montague, P.R. (2008). A framework for studying the neurobiology of value-based decision making. Nat. Rev. Neurosci. *9*, 545–556.

Sescousse, G., Caldú, X., Segura, B., and Dreher, J.-C. (2013). Processing of primary and secondary rewards: a quantitative meta-analysis and review of human functional neuroimaging studies. Neurosci. Biobehav. Rev. *37*, 681–696.

Wang, X.-J., and Krystal, J.H. (2014). Computational Psychiatry. Neuron *84*, 638–654.

Woolrich, M.W., Behrens, T.E., Beckmann, C.F., Jenkinson, M., and Smith, S.M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. Neuroimage *21*, 1732–1747.

Worsley, K.J., Liao, C.H., Aston, J., Petre, V., Duncan, G.H., Morales, F., and Evans, A.C. (2002). A general statistical analysis for fMRI data. Neuroimage *15*, 1–15.

Yarkoni, T., and Braver, T.S. (2010). Cognitive Neuroscience Approaches to Individual Differences in Working Memory and Executive Control: Conceptual and Methodological Issues. In Handbook of Individual Differences in Cognition, A. Gruszka, G. Matthews, and B. Szymura, eds. (Springer New York), pp. 87–107.

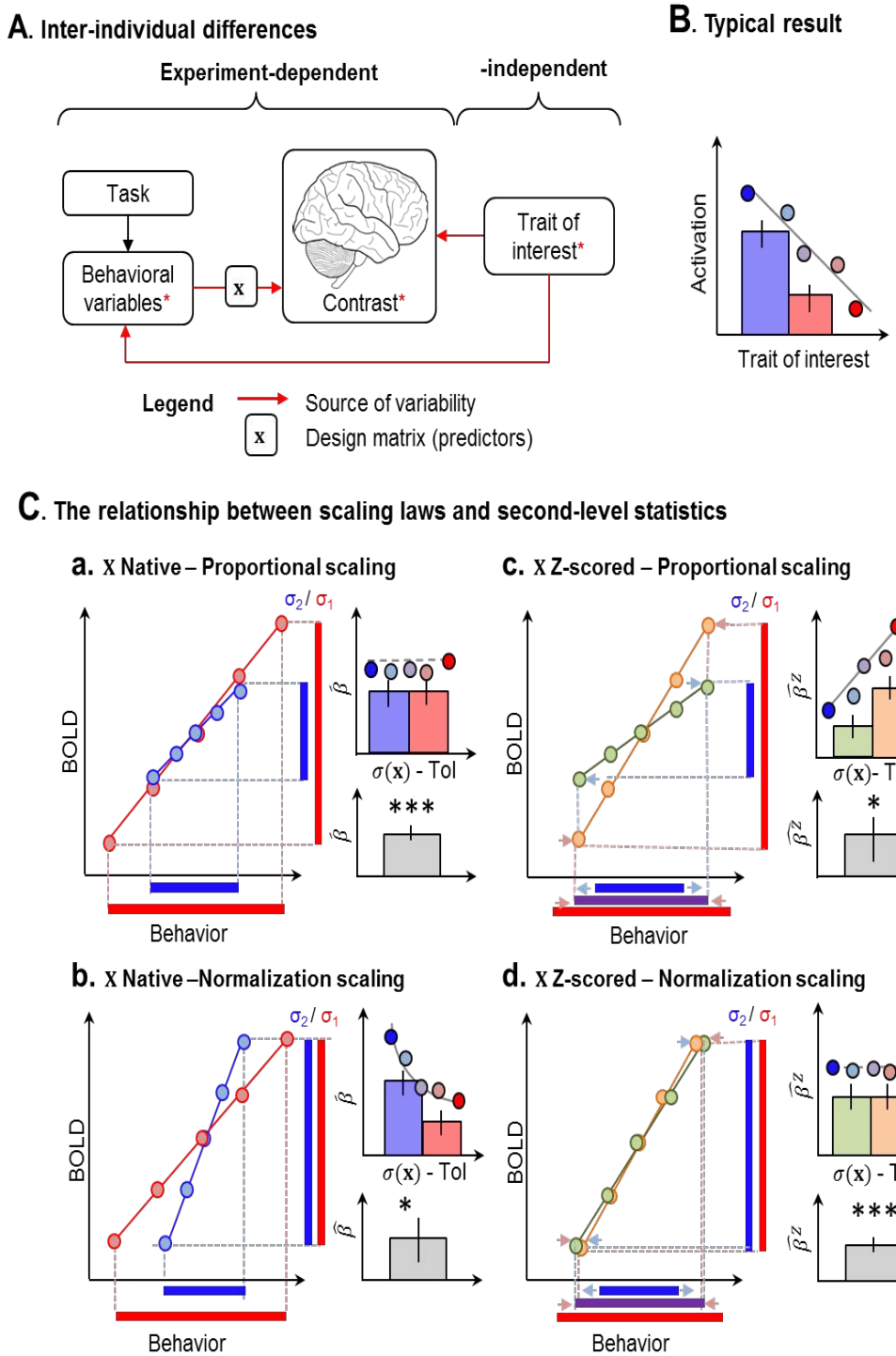


Figure 1: Inter-individual in brain-behavior relationships and scaling laws

A. Inter-individual differences in brain-behavior relationships are paramount in virtually all fMRI designs, regardless of whether they address these differences explicitly or not. They are typically assessed by linking individual fMRI contrast values with a trait of interest. However they also often leak in fMRI data analysis when behavioral variables are used as independent variables in the design matrix. Problems can arise when the trait of interest also generates differences in the variance of the behavioral variables. **B.** Typical results illustrating inter-individual differences in brain-behavior relationships, as they are reported in the literature, in the form of categorical contrast between group of subjects (bars±s.e.m) or group level linear correlation with a continuous trait (dots and regression line). **C.** The relationship between scaling laws and second-level statistics. The panels describe the impact of inter-individual differences in the standard deviation of the behavioral variable, under different preprocessing -native (a,b) or Z-scored variable (c,d)- and under different scaling-laws hypotheses -proportional (a,c) and normalization (a,d). Each sub panel contains three graphs. On the left we show an illustration of how the behavioral variable is related to the BOLD signal in two individuals with different initial standard deviations (blue vs. red or green vs. orange).

The individual unstandardized coefficients of regression (typically issued by the fMRI first-level analysis), corresponds the slope of the individual lines. On the upper right corner we show an illustration of the inter-individual statistical relations between brain activations and the Trait of Interest (Tol) – as a between-group analysis (histograms) or continuous inter-individual correlation (dots). On the bottom-right corner we show an illustration of the consequences for the significance of second-level random-effect analysis (*: lower significance, vs. ***: higher significance). $\hat{\beta}$ and $\widehat{\beta^z}$ respectively refer to fMRI unstandardized coefficients of regressions computed with a native scaling or an individual Z-scoring of the parametric regressor $\mathbf{x}$. $\sigma(\mathbf{x})$refers to the standard deviation of the parametric regressor $\mathbf{x}$.

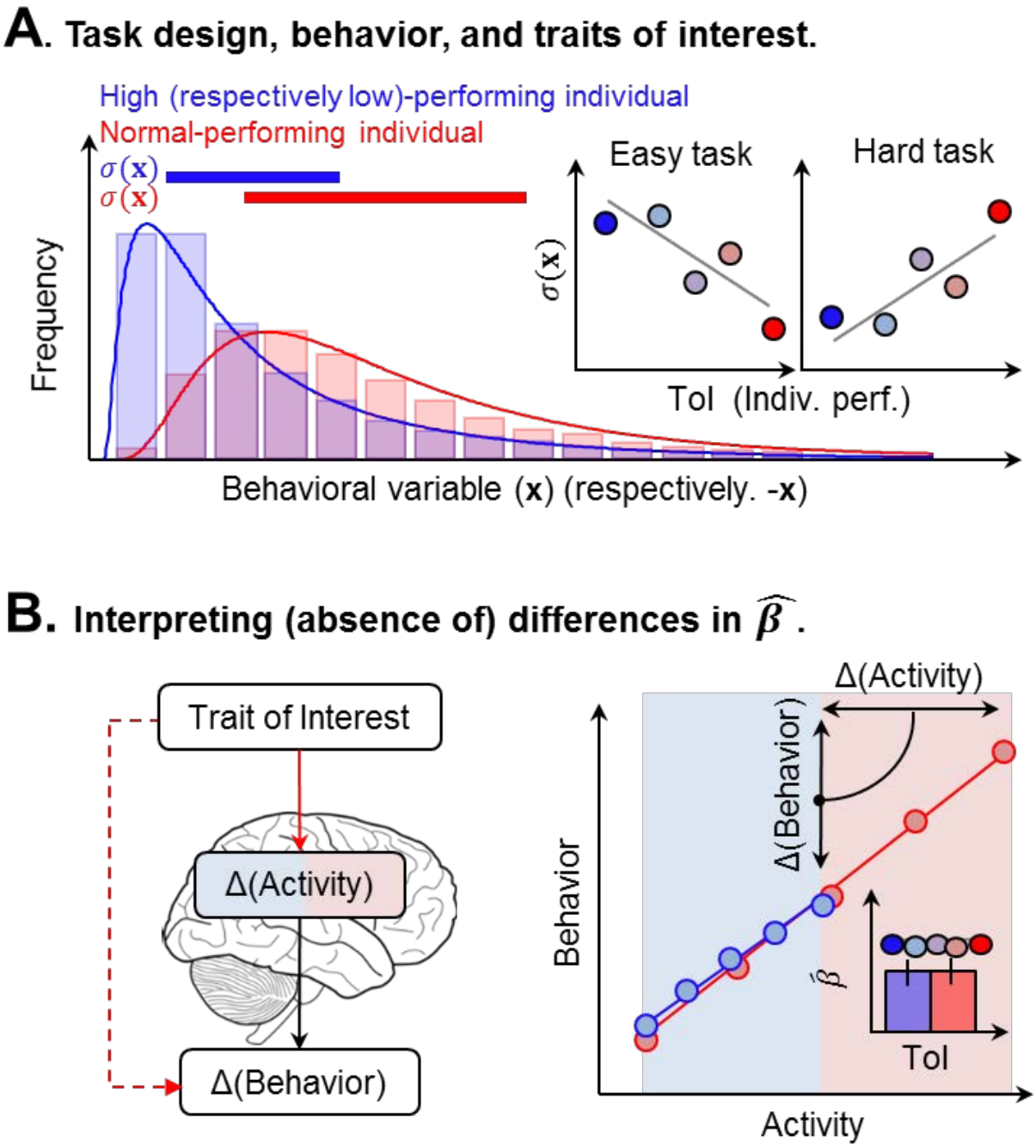
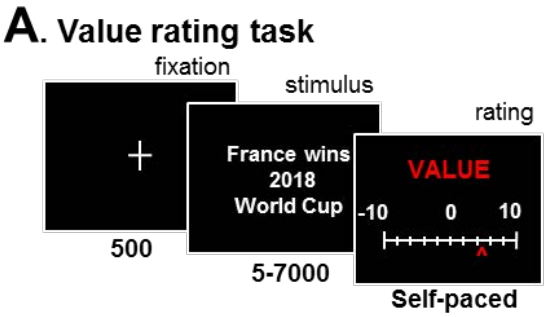
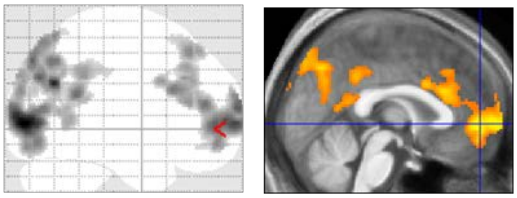


Figure 2: The importance and impact of scaling laws in investigations of inter-individual differences
A. Linking task design, behavior, and traits of interest. The red/blue histograms depict the distribution of a behavioral variable (reaction time, decision value, confidence) in two individual with variable performance. If the task is easy (respectively difficult), the high (respectively low)-performing individual can exhibit a ceiling (respectively floor) effect on performance. This creates statistical dependencies between the trait of interest (performance) and the standard deviation ($\sigma(\mathbf{x})$) of the behavioral performance variable (see graphical insets). Given the dependencies between fMRI $\hat{\beta}$ and $\sigma(\mathbf{x})$, this can lead to opposite correlations between "activations" –as measured by β-, and "individual performances" (proxied by $\sigma(\mathbf{x})$). **B**. Interpreting differences in unstandardized $\hat{\beta}$. Consider a brain region, which causally and proportionally causes a behavior (i.e. the more activation, the higher the behavioral variable, within and across subjects). In case a trait of interest (e.g. pathology) directly impacts the range of activation of this region (e.g. due to degeneration), this cannot be assessed/detected by differences in $\hat{\beta}$ (right, inset). Misunderstanding the signification of $\hat{\beta}$ and ignoring scaling laws can lead to erroneous negative conclusions.





C. Assessing scaling laws – Summary

	Random effects		Correlations			Hyp.
	VMPFC *k*	ROI -log(P)	$\widehat{\beta^{Z}} \propto \sigma(\mathbf{v})$	$\hat{\beta} \propto 1/\sigma(\mathbf{v})$	$\hat{t} \propto 1/\sigma(\mathbf{v})$	
Lebreton, et al. 2009	$\mathbf{v}^{Z} > \mathbf{v}$	$\mathbf{v}^{Z} > \mathbf{v}$	ns	+ (**)	ns	**Norm.**
Lebreton, et al. 2012	$\mathbf{v}^{Z} > \mathbf{v}$	$\mathbf{v}^{Z} > \mathbf{v}$	- (*)	+ (***)	ns	**Norm.**
Lebreton, et al. 2015	$\mathbf{v}^{Z} > \mathbf{v}$	$\mathbf{v}^{Z} > \mathbf{v}$	ns	+ (*)	ns	**Norm.**

Figure 3: Assessing BOLD-Behavior scaling relationships in value rating tasks

A. Example of Study 3 rating task (Lebreton et al., 2015). Successive screens displayed in one trial are shown from left to right, with duration in milliseconds. **B.** Example of group-level neural correlates of values in Study 3. The color code on glass brains (left maps) and sagittal slices (right) indicate the statistical significance of clusters that survived the whole-brain family-wise error (FWE) correction for multiple comparisons, computed at the cluster level ($P_{FWE\text{-}clu}$<.05, with a voxel-wise cluster-generating threshold P_{UNCORR}<.001). **C.** Main results assessing the scaling laws in three valuation studies. $\mathbf{v}$ and $\mathbf{v}^{Z}$ respectively indicate that the fMRI GLMs are designed with a native scaling or an individual Z-scoring of the parametric regressor *value*. VMPFC *k* refers to the size of the VMPFC cluster (in voxels, cluster-generating voxel threshold P_{UNCORR}<.001); ROI –log(P) refers to the negative logarithm of the P-value of a random effect analysis performed on the individual averaged coefficient of regression extracted from an anatomical VMPFC ROI. $\hat{\beta}$ and $\widehat{\beta^{Z}}$ respectively refer to fMRI unstandardized coefficients of regressions computed with a native scaling or an individual Z-scoring of the parametric regressor *value*. $\hat{t}$ refers to fMRI t-statistics derived from $\hat{\beta}$; $\sigma(\mathbf{v})$ refers to the standard deviation of the native value-rating measure. ns: non significant; +/-: positive or negative correlation; *: P<.05; **: P<.01; ***: P<.001;

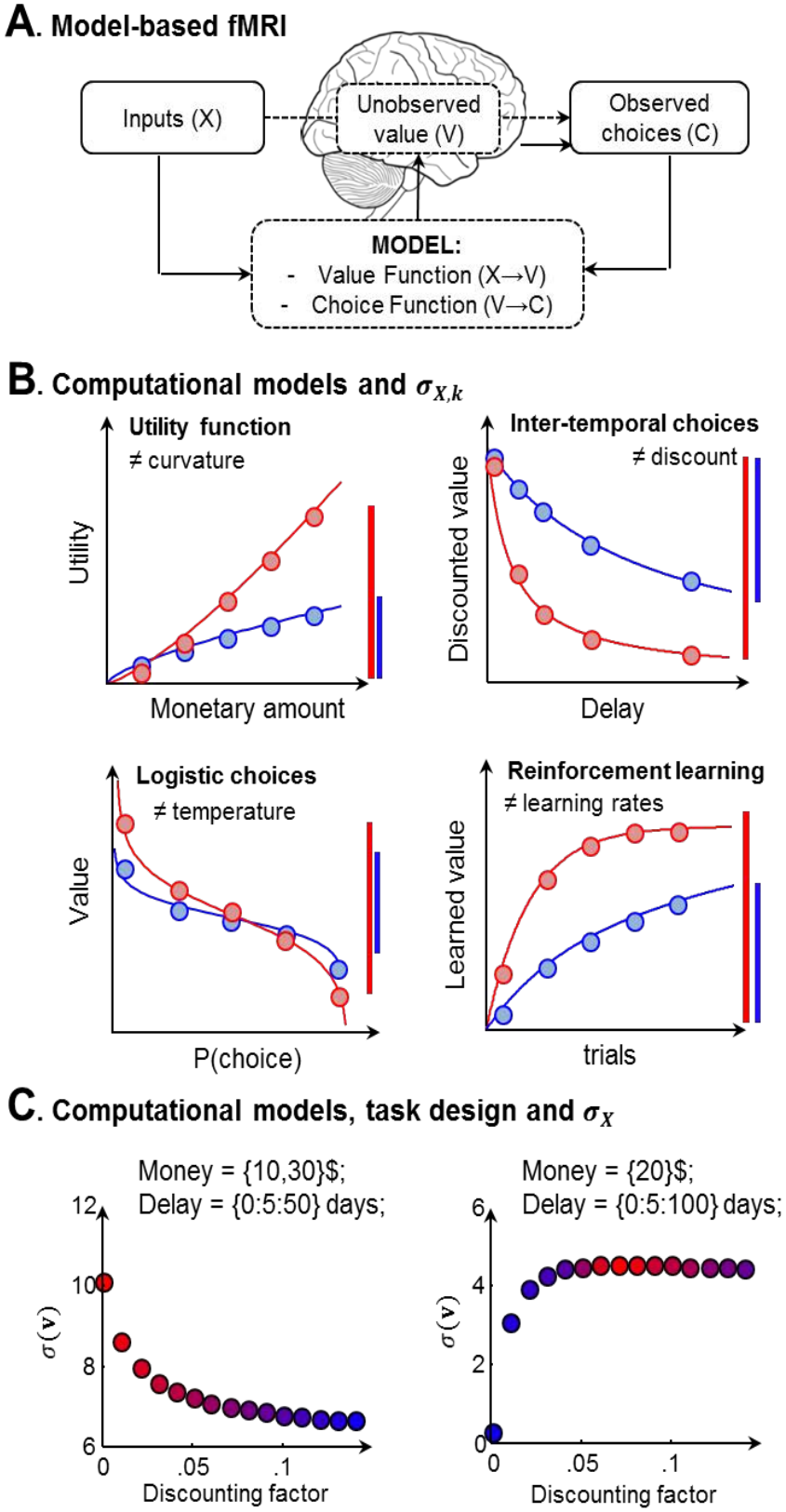
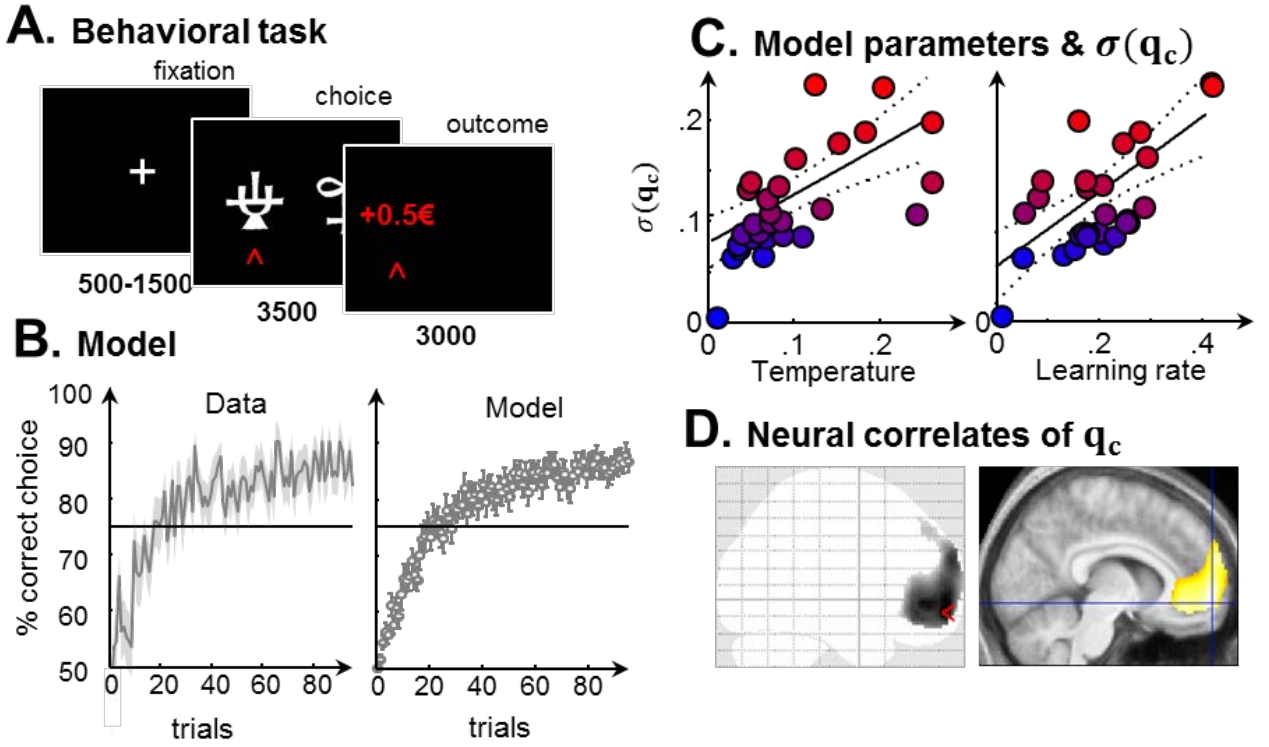


Figure 4: Model-based fMRI and the impact of model-parameters on inter-individual differences
A. Typical model-based framework used in the value-based decision-making community. **B.** Schematic illustrations of the impact of model parameters, on the standard deviation of the value variable in different context. Upper-left corner: impact of the utility curvature parameter on value; upper-right corner: impact of the discount factor on discounted value in inter-temporal choice tasks; lower-left corner: impact of the logistic choice temperature on the decision value; lower-right corner: impact of the learning rate on Q-values in reinforcement learning tasks. **C.** Computer simulations illustrating the impact of task designs on the relation between the model-free parameters and the standard deviation of the value variable. We considered a delay-discounting task with two stimuli space. Left: the stimulus space was composed of two monetary amounts (10 and 30$) combined with 11 delays (from 0 to 50 days with a 5 day incremental step). Right: the stimulus space was composed of tone monetary amounts (20$) combined with 22 delays (from 0 to 100 days with a 5 day incremental step). Discounted values where computed with a hyperbolic discount function $\mathbf{v} = \frac{A}{1+k\times D}$, where A is the monetary amount, D the delay, and k the individual discount factor (a free-parameter). We computed the values for all the stimuli in the stimulus space and estimated the standard deviation of this value distribution $\sigma(\mathbf{v})$ with 14 different discount factors, starting from .0005, with a .01 incremental step. Simulations show that in the first condition, $\sigma(\mathbf{v})$ *decreases* monotonically with individual discount factors, whereas in the second condition, $\sigma(\mathbf{v})$ *increases* monotonically with individual discount factors, before reaching a plateau. The dots color codes the relative $\sigma(\mathbf{v})$ value, from low (blue) to high (red).



E. Scaling law

Random effects		Correlations: R (p)			Hyp
VMPFC *k*	ROI -log(P)	$\widehat{\beta^Z} \propto \sigma(\mathbf{q_c})$	$\hat{\beta} \propto 1/\sigma(\mathbf{q_c})$	$\hat{t} \propto 1/\sigma(\mathbf{q_c})$	
$\mathbf{q_c}^Z > \mathbf{q_c}^P > \mathbf{q_c}$	$\mathbf{q_c}^Z > \mathbf{q_c}^P > \mathbf{q_c}$	- (*)	+ (***)	ns	**Norm**

Figure 5: Assessing BOLD-Behavior scaling relationships in a learning tasks

A. Behavioral task. Successive screens displayed in one trial are shown from left to right, with duration in milliseconds (Palminteri et al., 2015). **B.** Average behavior (left) and model fit (right), displayed as the correct choice rate. Shaded area (left) and error bars (right) correspond to mean±sem. **C.** Statistical relations between model free-parameters - temperature (left) and learning rate (right)- and the standard deviation of the chosen Q-values $\sigma(\mathbf{q_c})$. Solid line indicate the best linear fit, and dotted line the 95% confidence interval. The dots color codes the relative $\sigma(\mathbf{q_c})$ value, from low (blue) to high (red). D**.** Group-level neural correlates of *chosen Q-values*. The color code on glass brains (left maps) and sagittal slices (right) indicate the statistical significance of clusters that survived the whole-brain family-wise error (FWE) correction for multiple comparisons, computed at the cluster level (P_{FWE} <.05, with a voxel-wise cluster-generating threshold P_{UNCORR}<.001) **E**. Main results assessing the scaling laws. $\mathbf{q_c}^P$, $\mathbf{q_c}$ and $\mathbf{q_c}^Z$indicate that the fMRI GLMs are designed with the variable chosen Q-values generated with population ($\mathbf{q_c}^P$) or individual ($\mathbf{q_c}$ and $\mathbf{q_c}^Z$) model free parameters, and using a native scaling ($\mathbf{q_c}^P$ and $\mathbf{q_c}$) or an individual Z-scoring ($\mathbf{q_c}^Z$) of the variable. VMPFC *k* refers to the size of the VMPFC cluster (in voxels, cluster-generating voxel threshold P_{UNCORR}<.001); ROI – log(P) refers to the negative logarithm of the P-value of a random effect analysis performed on the individual averaged coefficient of regression extracted from an anatomical VMPFC ROI. $\hat{\beta}$ and $\widehat{\beta^Z}$ respectively refer to fMRI unstandardized coefficients of regressions computed with a native scaling or an individual Z-scoring of the parametric regressor *Qc*. $\hat{t}$ refers to fMRI t-statistics derived from $\hat{\beta}$; $\sigma(\mathbf{q_c})$ refers to the standard deviation of the native *Qc*. ns: non-significant; +/-: positive or negative correlation; *: P<.05; ***: P<.001;