



HAL
open science

Bridging across functional models: The OFC as a value-making neural network.

Mathias Pessiglione, Jean Daunizeau

► **To cite this version:**

Mathias Pessiglione, Jean Daunizeau. Bridging across functional models: The OFC as a value-making neural network.. Behavioral Neuroscience, 2021, 135 (2), pp.277-290. <10.1037/bne0000464>. <hal-03395604>

HAL Id: hal-03395604

<https://hal.sorbonne-universite.fr/hal-03395604v1>

Submitted on 22 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

***** special issue: The Magical Orbitofrontal Cortex *****

Title:

Bridging across functional models: the OFC as a value-making neural network

Authors:

Mathias Pessiglione^{1,2,*} & Jean Daunizeau^{1,2}

¹Motivation, Brain & Behavior lab (MBB), Paris Brain Institute (ICM), Pitié-Salpêtrière Hospital, Paris, France

²INSERM Unit 1127, CNRS Unit 7225, Sorbonne-Universités (SU), Paris, France

* Correspondence should be addressed to: mathias.pessiglione@gmail.com

The authors declare no conflict of interest.

Abstract:

Many functions have been attributed to the orbitofrontal cortex (OFC) - some classical roles, such as signaling the value of action outcomes, being challenged by more recent ones, such as signaling the position of a trial within a task space. In this paper, we propose a unifying neural network architecture, whose function is to generate a value from a set of attributes attached to a particular object. Our model reverses the logic of perceptual choice models, by considering values as outputs of (and not inputs to) the neural network. In doing so, the model explains why univariate value signals have been observed in both likeability rating and economic choice tasks, while the features associated to a particular task trial can be decoded using multivariate analysis. Moreover, simulations show that a globally positive correlation with subjective value at the population level can coexist with a variety of correlation coefficients at the single-unit level, bridging typical observations made in human neuroimaging and monkey electrophysiology studies of OFC activity. To better explain binary choice, we equipped the neural network with recurrent feedback connections that enable simultaneous coding of values associated to currently-attended and previously-considered objects. Simulations of this augmented model show that virtual lesions produce systematically intransitive preferences, as observed in patients with damage to the OFC. Thus, our neural network model is sufficiently general and flexible to account for a core set of observations and make specific predictions about both OFC activity during value judgment and behavioral consequence of OFC damage.

Keywords:

Orbitofrontal cortex; artificial neural network; subjective value; likeability rating; economic choice; electrophysiology; functional neuroimaging

Main manuscript:

The problem of value construction

One of the most common signals observed in recordings of OFC activity is related to subjective value. The notion of subjective value is borrowed from economic decision theories assuming that options in a choice set are positioned on a common scale, which assigns them both rank order and distance metric. Subjective value may be seen as a generalization of the expected utility concept (a function of objective attributes like magnitude and probability of monetary gains), to the case where there are no explicit attributes (like when one chooses between an apple and a banana, or between a between football game and a theater play). Qualification as ‘subjective’ simply designates the fact that different individuals might assign different values to the same option, which would be captured by different parameters in utility functions. In economics, expected utilities are typically considered as abstract quantities that describe a pattern of choice behavior, without a need for actual implementation in the mind / brain. Somewhat ironically, neuroscience studies in the last two decades have repeatedly reported observations of value signals in brain activity, primarily in the OFC (Fig. 1), on top of several other brain regions. Our aim here is to provide a mechanistic account for how the OFC can generate value signals in most typical situations.

Value signals have been observed in different species (human and non-human primates mostly) using different types of signals (from single-unit spiking to population-level hemodynamic activity). Yet two important differences have been noticed: first, electrophysiological studies in monkeys have more frequently investigated and reported value signals in the lateral OFC (Padoa-Schioppa & Assad, 2006; Rich & Wallis, 2014; Tremblay & Schultz, 1999), whereas meta-analyses of human fMRI studies have located value signals in the medial OFC and adjacent regions along the medial wall (Bartra et al., 2013; Clithero &

Rangel, 2014; Peters & Büchel, 2010), forming the so-called ventromedial prefrontal cortex area (vmPFC). This discrepancy might relate to inter-species differences in functional anatomy, to differences in behavioral tasks and training duration, or to differences in recording techniques and analytical approaches. Indeed, recent studies investigating intracerebral activity through implanted electrodes in humans have recorded similar value signals in the medial and lateral OFC (Lopez-Persem et al., 2020; Saez et al., 2018). For our purposes, we will consider OFC as a whole, with no distinction between medial and lateral parts. The idea is to develop a neural model of how a brain region like the OFC may implement the general function of providing value signals, not to account for the functional specificities of OFC subregions. A second difference between human and monkey studies is that correlation with value is systematically positive in population-level hemodynamic signals, while it is widely distributed over negative to positive coefficients across individual neurons (Abitbol et al., 2015; Bartra et al., 2013; Lopez-Persem et al., 2020; Padoa-Schioppa & Conen, 2017; Strait et al., 2014; Tom et al., 2007). We will consider this difference between micro-scale and meso-scale activities as an important pattern that must be accounted for in a neural model of OFC function.

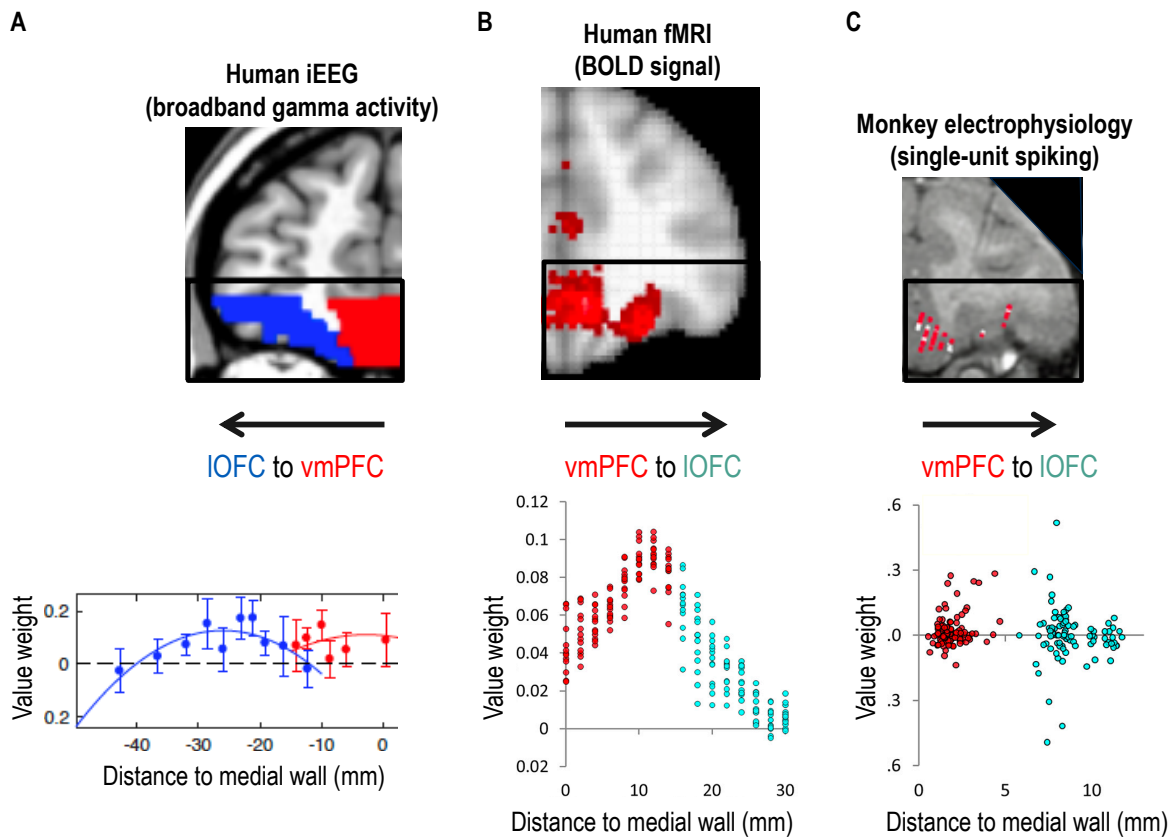


Figure 1 – Value signals across techniques and species

The three panels depict estimates of correlation between value and OFC activity recorded using iEEG, fMRI, or extra-cellular electrodes. Brain scans on the top row show IOFC and vmPFC regions (anatomically defined from the AAL atlas) sampled with local field potential recording electrodes in patients with epilepsy (A), significant clusters activated with the keyword ‘value’ in meta-analysis of human fMRI studies using the Neurosynth platform (B) and OFC regions sampled with spike recording electrodes in macaque monkeys (C). Plots on the bottom row show for different positions along the medio-lateral axis the posterior estimates of the correlation with likeability rating (A and B) or intensity of licking behavior (C). Dots are group-level estimates in bins of contact electrodes (A), in voxels normalized to the MNI template (B), and in individual neurons (C). Data were taken from previously published studies (Lopez-Persem et al. 2020 for A and Abitbol et al. 2015 for B and C).

Critically, OFC value signals have been observed in a variety of contexts (Fig. 2), not only in choice tasks where subjects select an option among two or more, but also in Pavlovian or instrumental learning tasks where subjects are shown cues and outcomes (Palminteri et al., 2012; Pauli et al., 2019), as well as in tasks where (human) participants assign numbers, such as willingness-to-pay or likeability ratings, to options presented one by one (Lebreton et al.,

2009; Plassmann et al., 2008; Polanía et al., 2019; Suzuki et al., 2017). Choice tasks present complex situations because several values, attached to the different options, must be signaled at the same time. Many different combinations of option values have been reported in these situations, with a predominance for the value of the chosen option, which generally happens to also be the option that is most attended and/or that corresponds to prior preference (Hunt et al., 2012; Ian Krajbich et al., 2010; Lopez-Persem et al., 2016; Wunderlich et al., 2009). Learning tasks also present a complication related to the progressive building of expectations at the time of cue onset, which may attenuate value signals at the time of outcome onset, according to principles of predictive coding. Many studies have consistently observed that OFC activity was correlated with both expected value at cue onset and prediction error at outcome onset (Bartra et al., 2013; Fouragnan et al., 2018; Garrison et al., 2013; Palminteri et al., 2012). Thus, the most straightforward situation to investigate is perhaps likeability rating, where a single value signal is induced at every trial by a novel item displayed on screen.

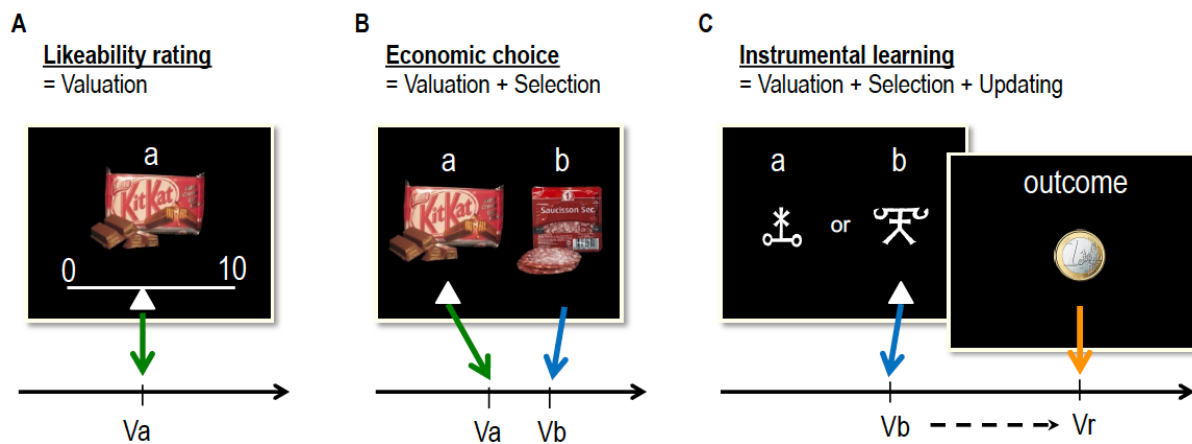


Figure 2 – Different types of value-based behavioral tasks

To investigate the valuation process, the most straightforward task is likeability rating (A), as there is only one novel option to value on each trial. Choice tasks (B) propose at least two options and thereby involve the additional stages of comparing option values and selecting the best option. On top of these two stages, instrumental learning tasks (C) additionally involve updating the value of chosen options, which gradually converge to the values of their respective outcomes.

Surprisingly, the neural mechanisms underlying likeability ratings remain poorly understood. Most functional models have focused on how values are transformed into choices, following on economic decision theories. In these models, by construction, values are inputs meant to explain choice outputs (Fig. 3). The simplest model is certainly the softmax function (Pleskac, 2015; Sutton & Barto, 1998), which in the case of binary choice reduces to sigmoid mapping from decision value (difference between the two options) to selection probability (assigned to each option). More sophisticated models have built on the accumulation-to-bound principle, originally applied to perceptual decisions, in order to generate distributions of response time, on top of choice probabilities (Krajbich & Rangel, 2011; Ratcliff et al., 2016; Tajima et al., 2019). In these models, decision value is incorporated as the mean of drift over time, such that choice probability scales with the signed difference between option values (the best option is chosen more frequently), and response time with the unsigned difference (choice is faster for more distant options).

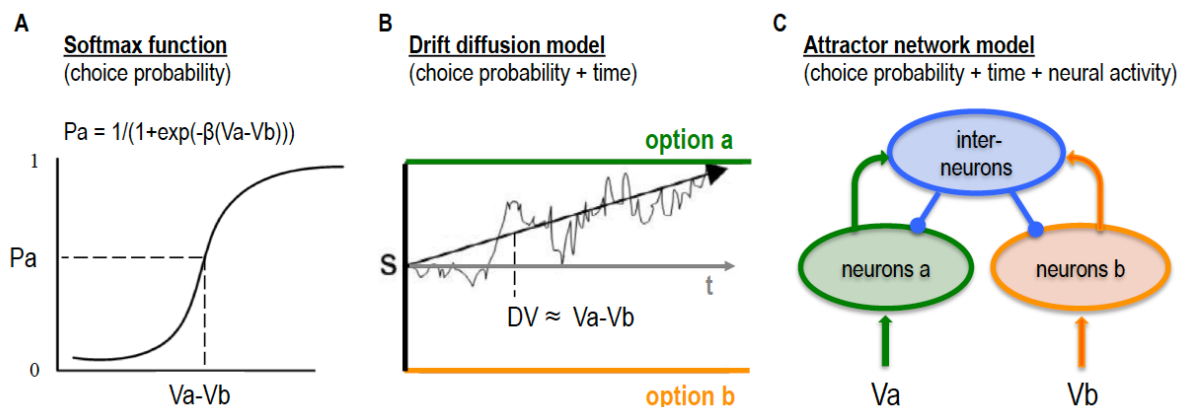


Figure 3 – Choice models are about selection (not valuation)

The three models are applied to the case of binary choice between options a and b. They are meant to explain how option values V_a and V_b influence first choice probability (A), then choice duration (B), and last underlying neural activity (C). In the softmax function (A), the decision value (V_a-V_b) is mapped onto choice probability (P_a and $P_b=1-P_a$) through sigmoidal transformation. The multiplicative slope parameter adjusts for choice stochasticity, while an additive bias parameter could account for an indifference point ($P_a=0.5$) shifted from null decision value ($V_a-V_b=0$). In the drift diffusion model (B), a decision variable is incremented at each time step by the decision value plus some noise. The decision is made at the time when it reaches one of two bounds corresponding to the two options. In addition to

decision value, choice duration depends on starting point *S* and on the distance between bounds. In the attractor network model (C), two pools of neurons compete for their favorite option, through a pool of inhibitory interneurons. When one of the two attractor states is reached, the winning pool is then signaling which option has been selected by the network.

All these models account for how values are transformed into choices, with increasingly rich predictions, but not for how values might be computed.

Finally, neural network models, again inspired by perceptual decision, have been developed to account for economic choice. They typically implement two populations of neurons, each voting for its preferred option (taking its value as an input) and competing with the other through a pool of inhibitory interneurons. The network converges to an attractor state where one of the populations reaches stable ceiling activity, while the other is silenced, following a winner-takes-all logic. On top of reproducing the functions relating values to choice probability and response time, these attractor networks make predictions about neural activity, which could account for some of OFC value signals (Hunt et al., 2012; Rustichini & Padoa-Schioppa, 2015; Wang, 2002). Yet all these models of choice behavior and underlying neural activity focus on how the best option is selected; they remain entirely agnostic about the issue of how values are computed in the first place. We would argue, along with other authors (Juechems & Summerfield, 2019; O'Doherty, 2014), that understanding the generation of value signals is the hard problem, compared to which the selection of the best option is an easy problem.

Our intention in this paper is to suggest a neural network architecture that provides values as outputs. By value we mean an absolute position on a scale, equivalent to a neural common currency, and not just a rank order. This is mandatory because we intend to explain how OFC activity can solve the task of reporting likeability feelings, for which relative preference is not enough. There is a crucial difference between our concept of value construction and earlier suggestions based on probabilistic population coding (Polanía et al.,

2019; Pouget et al., 2013). These earlier suggestions stem from an analogy with perception: some specific neurons would encode values with tuning functions (accounting for receptive fields), just as other populations of neurons would encode speed or direction of a luminous stimulus. We believe that the idea of explaining value signals with sensory coding principles is ill-conceived for two reasons. The first is empirical: if different OFC neurons have tuning curves that peak at different values, following a basic assumption of probabilistic population codes, there is no reason why population-level activity (measured with fMRI or iEEG) would correlate with the value of the item expressed in likeability rating. This is because every value would activate a similar fraction of neurons (those with ad hoc receptive fields), and hence yield similar total activity, unless we throw in some post hoc assumptions, like neurons coding for high values are more numerous or fire more. The second reason is conceptual: while it makes sense for a piece of brain to encode some perceptual feature of the outside world, such that the rest of the brain can use the information, it seems absurd for a brain region to encode subjective value. Indeed, subjective value is not a dimension of the outside world, it is an internal feature that would necessarily come from another brain region, which would therefore have already done the job of making value information available. This is why we insist on the idea that some brain region must be constructing value, while there is no need for a brain region encoding value. Although we focus here on OFC as the most likely candidate for implementing the value construction process, it is certainly not the only one and other brain regions might be constructed other values in parallel as well.

The solution: a neural network that translates features into values

If value is the output of OFC computations, the question is about what the input can be. Here, it is important to recollect that value is not the only information that is represented in OFC activity. Several recent studies, seemingly shifting away from any value-related function, have

demonstrated that the position of a particular stimulus or trial, within the space of possibilities offered by the task, can be decoded in OFC activity (Schuck et al., 2016; Wilson et al., 2014). Some structure of task space encoding in OFC activity has even been identified. Notably, in tasks where stimuli are systematically varied along two dimensions, some 60-degree periodicity found in OFC activity provided evidence for hexagonal paving of task space, similar to the grid-cell type of coding for physical space (Bao et al., 2019; Constantinescu et al., 2016). To reconcile this new literature with the old one, a neural network model of OFC function must therefore represent both stimulus features and subjective values. A natural solution is to consider that the function of the OFC is precisely to transform stimulus features into subjective values.

We are not the first of course to suggest that stimuli may first be decomposed into features (or attributes), which would then be combined to generate subjective values. Common features are for instance constituent nutrients (Suzuki et al., 2017), or healthiness versus tastiness for food items (Hare et al., 2009); they could be extended to any relevant dimension along which the item under valuation may vary. Our aim here is to derive some consequences of the postulate that the integration of features for the construction of subjective values occurs within the OFC. For this transformation to occur, the minimal architecture is a two-layer neural network, with an input layer that sparsely encodes information about the item (its coordinates on relevant features indicating its position in task space), and an integration layer that combines these pieces of information to generate a distributed value signal (Fig. 4A). A efferent neuron downstream to the integration layer, outside the OFC, could then decode the value through a weighted sum of single-unit activities.

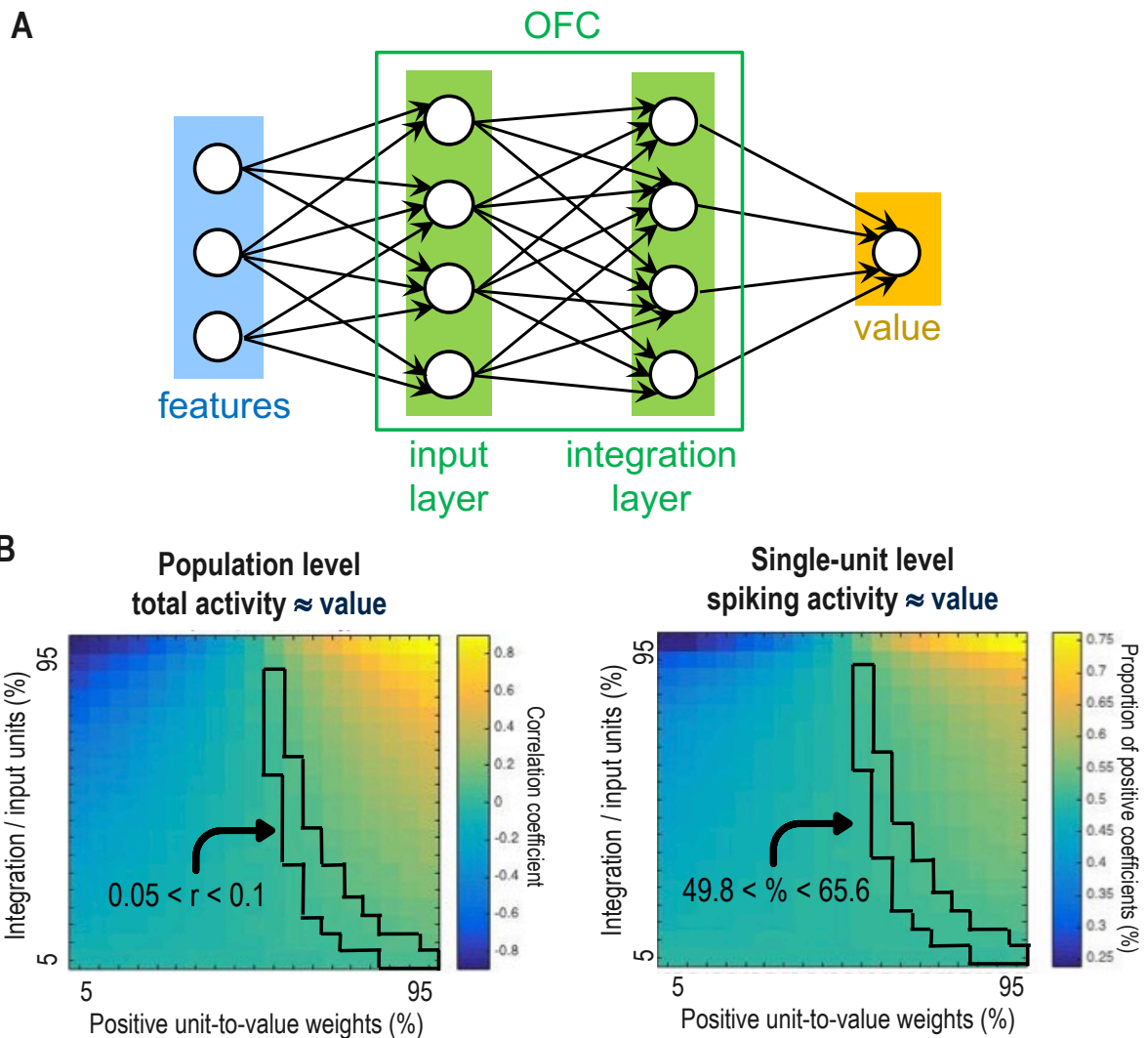


Figure 4 – A neural network for value construction

In our model (A), the OFC is composed of two layers: an input layer that sparsely encodes some features of a particular event and an integration layer that generates a decodable value signal. The tuning (input/output) functions implemented in neurons and the pattern of connections between neurons would not change the correlations between value and neural activity. However, simulations (B) show that two factors do change the (color-coded) correlations at both the single-unit and population levels: the proportion of integration vs. input neurons in the OFC network (y-axis), as well as the proportion of positive vs. negative weights in the output decoder (x-axis). The region of the two-dimensional space framed in black shows the global correlation (left graph) that is typically observed in human fMRI or iEEG studies, between hemodynamic or broadband gamma OFC activity and a proxy for value such as likeability rating. To produce this level of global correlation (between 0.05 and 0.1) between value and total activity (sum over units), the network was found (right graph) to have a rather balanced proportion of positively vs. negatively correlated neurons (between 49.8 and 65.6 %).

Simulation of this simple architecture (Fig. 4B) shows that it can, under a certain range of specifications, generate the expected pattern: when correlating activity with value, coefficients are widely distributed over single units, but positive at the population level (taking total network activity as the sum over all neurons). In the simulations (see methods), we did not incorporate any temporal dynamics: on every trial presenting a single item to value, the activity in each neuron is represented by one number. Correlations between these numbers and values (across trials) depend on two factors: first, the proportion of neurons in the input and integration layers, and second, the proportion of positive weights in the decoder that reads value out of the integration layer. The effects of these factors are quite intuitive: if all weights were positive, and if there was no input layer but just an integration layer, the correlation between activity and value would be trivially positive, at both the single-unit and population levels. In contrast, with negative weights in the decoder (integration units whose activities decrease with value) and/or units in the input layer (whose activity is tuned to features, not value), correlation coefficients are more largely distributed over neurons, and the global correlation can even be negative.

For every point of the space formed by these two factors (see Fig. 4B), we correlated value to each single-unit activity and also to total activity. This simulated total activity should approximate the BOLD signal recorded with fMRI and the broadband gamma activity recorded with iEEG, which are both linearly linked to the aggregation of firing rates over neurons (Logothetis et al., 2001; Ray et al., 2008; Scheeringa et al., 2011; Bastin et al., 2012). We therefore restricted the simulation space to those regions showing global correlation coefficient within the range typically observed in human fMRI or iEEG studies, which is quite low (between 0.05 and 0.1) in our datasets (see Fig. 2). In those regions, the proportion of neurons showing positive regression coefficients varied between 49.8 and 65.6%, with a mean at 52.2%. In a previously published dataset of single-unit recordings from both the medial and lateral

OFC in monkeys (Abitbol et al., 2015), where value was inferred from lipping behavior following the presentation of visual cues associated to liquid rewards of different magnitudes, the proportion of positive coefficients was 53.8%.

These results suggest that the positive correlation between subjective value and population-level hemodynamic or electrophysiological activity (typically observed with human fMRI or iEEG studies) is perfectly compatible with a near 50/50 proportion of neurons with positive and negative correlation coefficients. Although the numbers are not always reported, we believe that a balanced mix of positive and negative coefficients is a frequent observation in monkey single-unit recordings (for another example, see Morrison & Salzman, 2009). However, we acknowledge that we do not know the ground truth about the proportion of positive and negative coefficients among OFC neurons. It might be less balanced in reality, and the low correlation observed at the population level could instead arise from noise in fMRI or iEEG recording. Note that the simulations are symmetrical, such that in principle, a mix of positive and negative coefficients at the single-cell level can also coexist with a negative correlation at the population level, which has been documented in a recent study using fMRI in monkeys (Papageorgiou, 2017).

The above results were obtained irrespective of other factors that were drawn at random for simulations, such as the pattern of connections between neurons and the peak or shape (sigmoidal vs. gaussian) of their tuning functions. Thus, the model reliably reproduces an observation that has been taken (wrongly, in our opinion) as evidence against a role for the OFC in value-related functions: the position of a particular trial within the task space can be decoded in OFC activity even when controlling for its value. This is exactly what should be obtained from a neural network transforming features into values. Indeed, position in task space may be nothing but a set of coordinates along the different dimensions, i.e. precisely the information

about the level of relevant features that is encoded in the input layer of the neural network, and recombined to generate subjective value signals in the integration layer.

How the value-making network may contribute to decision-making

Now that we have addressed the basic situation of likeability rating, where there is only one value judgment to provide, the next question is how the network can deal with choice situations, where there are at least two options to value. We know that the OFC contributes to economic choice, because damage to this region (more precisely to the vmPFC, in patient studies) impairs the rationality of decisions (Schneider & Koenigs, 2017). One rationality criterion in particular, the transitivity of preferences, is degraded following damage to the OFC area in stroke patients (Camille et al., 2011). Intransitive preferences mean that options cannot be ordered on a value scale, because for example patients declare preferring C to A, even if they previously stated that they prefer A to B and B to C. Virtual lesions removing part of a neural network used to simulate OFC dysfunction should therefore reproduce this pathological behavior. It is obviously expected that choices may be affected by an alteration of the neural network that provides subjective values to the decision-making machinery. However, a specific increase in transitivity violations does not trivially follow from the OFC functional architecture that we have proposed so far. To account for this observation, we must first explain how the neural network can represent more than one value at a time.

Standard models of decision-making posit that in the case of a binary choice, two distinct populations of neurons are voting for the two options A and B (Fig. 3C). However, labels A and B are artificially imposed on the neural network simulations. It is unclear how the network by itself could flexibly adapt to a new choice situation, unless the framing A vs. B has a very general meaning. Let us consider the main frames that have been suggested in the decision-making literature. A spatial frame (like ‘left vs. right’ options) has been originally

inspired by perceptual decision-making models (Hunt et al., 2012; Wong, 2006), opposing neurons with different spatial receptive fields. Although spatial framing can be easily applied to economic choice tasks showing two options side by side on a visual screen (Palminteri et al., 2009), it is difficult to figure how it could work in the case of options that succeed each other in time and/or that are only envisaged in imagination. Another kind of framing is ‘default vs. alternative’ options, originally inspired from foraging theory about decisions whether to stay and exploit a given patch or switch and explore new patches. The notion of default, defined as the option that the agent would take if no deliberation is engaged, has been used in a variety of choice situations (Fleming et al., 2010; Hayden et al., 2011; Kolling et al., 2012; Lopez-Persem et al., 2016; Tsetsos et al., 2014). However, it remains limited to cases where a prior preference can be identified, excluding choices between two novel alternatives. A last kind of framing is ‘attended vs. unattended’ options, which was shown to explain the link between OFC value signals and gaze fixation patterns (Krajbich et al., 2010; Lim et al., 2011). We would argue this attentional frame can be naturally generalized to any choice situation, including when the options are only imagined and not perceived as visual cues, meaning when attentional shifts are covert and not apparent in eye movements. The attentional frame is in fact confounded with a temporal frame, since options are necessarily attended one at a time, such that one option is always considered before the other. Note that it is also related to the ‘chosen vs. unchosen’ frame, because the option that is chosen in the end is also more looked at on average (Krajbich & Rangel, 2011).

Thus, we will assume in the following that the OFC network can simultaneously represent the values of at least two options (decomposed into features): the currently attended option and the previously considered option. In doing so, we keep the logic of a value-construction network, as opposed to a value-comparison network that would implement a competition between neural pools in order to select the best option. Thus, in the augmented

OFC network (Fig. 5A), two values can be read out from neuronal activity in the integration layer, with different codes (i.e., different vectors of weights). We prefer not assuming that the two values are represented in distinct neurons, as in standard decision-making models, for two reasons. First, this would mean losing generality, as two distinct populations is just a subcase of two distinct codes (with 0 weight on some neurons and 1 on the others). Second, multivariate pattern analysis of vmPFC single-unit recordings, in monkeys choosing between two successive options, showed that current and past option values could be decoded from the same neurons, by multiplying their activity with different weights (Tim Behrens, personal communication).

This double value coding imposes an important constraint on the network that precludes random connections between neurons. Indeed, for the network to be reliable, an option value must be the same when applying the current value code (at a given time) and when applying the previous value code (at the next time step). Nevertheless, mathematical derivations identified analytical solutions (i.e., a class of connection patterns) that ensure the stability of option values across the two neural codes (see methods). These solutions involve recurrent feedback connections in the integration layer, creating a sort of working memory. For the sake of simplicity, these derivations were worked out with linear integration of input activities, and for two value decoders only. In principle, the mechanism of multiple value decoding could be generalized to more than two options, although there is no strong evidence, to our knowledge, that OFC activity can represent more than two option values at the same time. This would imply that when people choose among more than two options, they need additional working memory systems implemented in brain regions outside the OFC. In other words, binary choice would be an easy case for the brain, because it would fit with the OFC specific architecture for value construction.

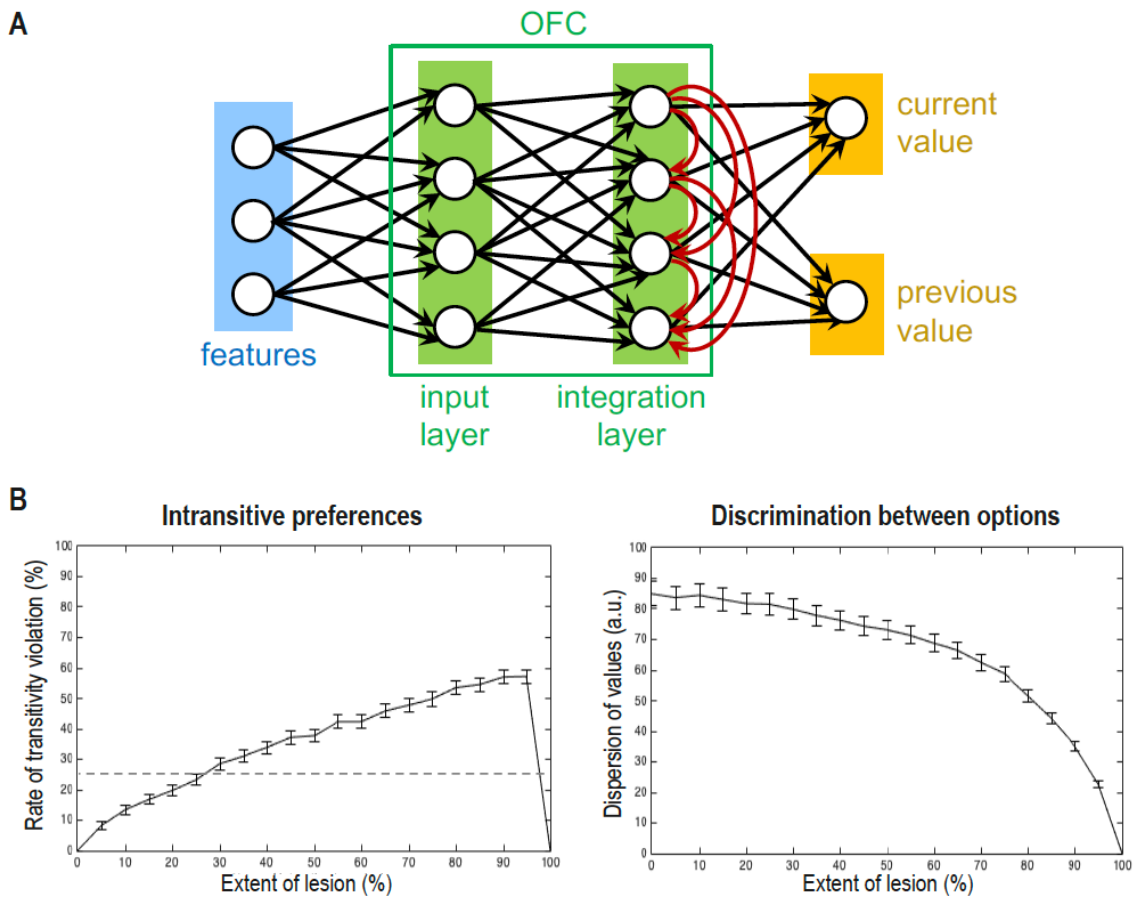


Figure 5 – A double value-making neural network for binary choice

The OFC neural network model (A) has been augmented with (red) recurrent feedback connections in the integration layer. For the sake of visibility, only half the possible connections (those going from top to bottom nodes) are shown, while the other half (those going from bottom to top nodes) are omitted. When these feedback connections are appropriately set, the neural network implements a form of working memory that enables decoding the values of both currently and previously attended items, at any time step. Graphs (B) show the effect of cutting neurons out of the integration layer, after connections are adjusted so as to reliably signal both current and previous values. Simulations of the lesioned network were used to generate the values assigned to first and second options in a series of three binary choices (a vs. b, b vs. c, c vs. a). Preferences are classified as intransitive (left) here when the three options cannot be ordered on a value scale (e.g., when $V_a > V_b$, $V_b > V_c$ and $V_c > V_a$). Chance level is 25% (dotted line) because under the null (with random values), transitivity violations can only be assessed in 50% of trials, and have a 50% probability to occur. Value dispersion (right) was calculated as the standard deviation of decision value (difference between option values), averaged over choices. When the integration layer is entirely removed, the network loses its capacity to discriminate between values, and hence the capacity to generate detectable transitivity violations. In both plots, error bars show standard errors across simulations (networks with different connections / tuning functions).

Now that our network is proven to function properly, translating the features of previous and current options into decodable and stable values, let us examine the consequences of virtual lesions. The intuition is that lesions will not affect in the same manner the values associated through the two neural codes to current and previous options. Indeed, if a random set of integration neurons is cut out of the network, it will likely remove distinct contributions to current and previous option values. Therefore, a same option would be valued differently when considered first or second in a binary choice, degrading the transitivity of preferences. This was verified in simulations of trials including three binary choices: options a vs. b, options b vs. c, options c vs. a. In this series of choices, each option alternates between first and second positions. Transitive preferences mean that the three options can be ordered on a scale, with for example $V_a > V_b > V_c$. Thus, we did not simulate the choice process, but just read out the values generated by the network for every three-option set (see methods).

For each simulated network, connections were drawn at random but then adjusted such that the network was perfectly reliable, hence the preferences perfectly transitive. As expected, the proportion of trials with transitivity violation increased with the proportion of integration units taken out of the network (Fig. 5B). The raw number of transitivity violations was obviously dependent on the degree of redundancy in the integration layer. If many neurons exist with the same weights on values, then a given lesion might spare some copies and hence have a lesser impact. Yet when the lesion killed more than 70% of integration neurons, the rate of transitivity violations was augmented beyond chance level (corresponding to what would be obtained with a network generating random values). Interestingly, when the extent of virtual lesions reached the totality of integration neurons, the number of transitivity violations dropped to zero. On closer inspection, this was due to a drastic reduction in value dispersion: when the network is deprived of integration neurons, it loses its ability to discriminate between option values. Note however that if we introduce choice in the model (with some stochasticity), and

define transitivity violation based on options actually selected (and not on their values), then closer option values would result in a higher rate of transitivity violations. Thus, there are (at least) two explanations for why OFC damage may result in intransitive choices. One, rather trivial, is that without a system able to differentiate option values, choices would be more random, and would therefore violate transitivity by accident. The other, more interesting, is that with a distorted system assigning different values to a same option considered first or second, the resulting preferences would be intransitive in a systematic way. We have shown that such distortion can account for increased intransitivity because this was the deficit observed in patients with OFC lesions, but the same mechanism would also predict increased inconsistency if the choices were repeated with the same two options presented in a reversed sequential order.

While the interest of this double value-making system is evident for making binary decisions, it is rather dubious for making likeability judgments. If, as we assumed, the double code stems from feedback recurrent connections, then the OFC neural network would mechanistically generate two values, even when items are presented one by one for likeability rating. It is possible that the previous item value is simply ignored by the neural system in charge of producing a response (i.e., assigning a likeability rating to the current item). Alternatively, it is possible that the previous item value is used to compute a distance with current item value and help with positioning the cursor on the rating scale. Further investigations are needed to elucidate whether and how value comparisons are achieved across trials during likeability rating tasks. The impact of OFC damage on likeability rating is more elusive than with economic choice, mostly because rational norms of decision theory (such as transitivity axiom) have been defined with respect to preferences (ordering of options) as revealed by choices, not ratings. Nevertheless, a difficulty in integrating several dimensions normally contributing to value judgment has been recently documented (Pelletier & Fellows, 2019; Vaidya et al., 2018). This is what could be expected from the view that the OFC network

is integrating features to generate values. If the network is only partially lesioned, and its inputs topographically organized, the weights on some features may be attenuated, while some other features would become dominant. In this case, preferences could be systematically shifted, as it does happen in fronto-temporal dementia, a neurodegenerative disease affecting the OFC (Lebreton et al., 2013; Piguet et al., 2011). If the OFC network is totally impaired, by a stroke for instance, patients might instead base their value judgment on a single, easy-to-quantify dimension of the stimulus.

Extension to other properties of the OFC valuation signal

Recent neuroimaging studies in humans have identified some core properties of the brain valuation system, verified in both the vmPFC and lateral OFC activity, recorded using either fMRI or iEEG. In the following, we examine whether these functional properties are compatible with the OFC neural network model suggested in this paper.

A first property is generality, meaning that the OFC associates value signals to items belonging to distinct categories (Chib et al., 2009; Lebreton et al., 2009; Lopez-Persem et al., 2020), thereby providing a so-called ‘common neural currency’ (Levy & Glimcher, 2012). It is an important property because it enables choosing between seemingly incommensurable options, like doing paperwork versus surfing on the net. Our neural network is general by construction, as we have not specified the kind of features that may be fed into the OFC. These features could be concrete attributes, either directly perceived, such as color and shape for a fruit, or extracted by inference, such as sugar content for a food item. More abstract features could also enter the valuation network, either semantic information like delay in reward delivery, or prospective episodes elaborated from memory fragments, like vacations in Paris. There has been a debate (see Cisek & Kalaska, 2010; Padoa-Schioppa, 2011) about whether the OFC only integrates dimensions of the good space (to compute the outcome value) or whether

it also integrates attributes of actions (to compute a net value). In principle, action-related features could input the valuation network, which would accord with increasing evidence that OFC activity reflects expected rewards discounted by costs such as effort intensity (Aridan et al., 2019; Lopez-Gamundi et al., 2021).

A second property of OFC function is the ability to signal confidence (in the behavioral response) on top of stimulus value. Confidence is defined here as the subjective accuracy of the response, which can be reduced to the probability of being correct in the case of a multiple-choice question. For our OFC network to generate confidence, it would be sufficient to feed it with the response (just intended or already given), and the evidence on which the response is based, as input features. Correlates of confidence have been observed in OFC activity during both likeability rating and economic choice (De Martino et al., 2013; Lebreton et al., 2015; Lopez-Persem et al., 2020), as well as during perception and memory tasks (Gherman & Philiastides, 2018; Hebscher et al., 2016). While the OFC is the most prominent region activated by positive correlation with confidence, a large set of brain regions, in both the parietal and prefrontal cortex, show negative correlations. Thus, the OFC is far from being a unique or central spot for processing confidence, or more generally for implementing meta-cognitive processes. One may argue that the reason why the OFC activates with confidence level is because confidence is inherently valuable. We would even argue that what is maximized when picking a response is precisely confidence (in making the right choice, or providing the right judgment). In other words, the OFC would compute confidence because it is a general valuation device, and confidence is just one particular kind of value.

A third functional property is automaticity, meaning that the OFC can signal the values of items even when doing so is not instrumental for solving the task at hand (Abitbol et al., 2015; Lebreton et al., 2009). A key condition for observing automatic value signals is that the item should be attended, for other purposes or for no particular purpose. This property has been

exploited in passive or Pavlovian paradigms, where participants have no voluntary action to perform in response to stimuli (Harvey et al., 2010; Levy et al., 2011). It has also been demonstrated in rating tasks where the dimension to estimate is decorrelated from stimulus value, such as the age of buildings or paintings (Lebreton et al., 2009; Lopez-Persem et al., 2020). We do not take automatic in the strong sense of irrepressible, but simply in the weak sense that value signals can be spontaneously generated even if not required for behavioral control. Indeed, some studies failed to detect such automatic value signals in the OFC (Grueschow et al., 2015), suggesting that they may be restricted to certain classes of stimuli or to tasks leaving sufficient time for triggering useless valuation processes. Faces for instance yield robust first-impression automatic value signals that predict later preferences between individuals (Ito et al., 2020; Kim et al., 2007). It is quite plausible indeed that stimuli whose valuation is essential for survival (such as fruits), or stimuli whose valuation is culturally trained (such as paintings), would more easily trigger automatic processing than stones or batteries. Besides, a traveler peacefully waiting for a train would have the leisure to appreciate attractive faces, whereas a parent rushing through the station to retrieve a lost child would probably miss them all. For our neural network, the issue of automaticity is again about input connections: is the routing of features to the OFC a controlled process, only engaged when a value judgment is needed, or is it mechanistically achieved for any attended object, such that the OFC would automatically generate a value? The answer could be mixed, the decomposition into features being hard-wired for some kinds of objects and more flexibly elaborated on demand for others.

A fourth functional property is hysteresis, meaning the fact that the value of preceding events, or the value of some elements in the surrounding context, influences the value assigned to the item currently presented (Lopez-Persem et al., 2020; Vinckier et al., 2018). For example, the pleasantness of a music excerpt played in the background was found to influence the likeability judgment assigned to a painting, by shifting baseline OFC activity before the

painting is displayed on screen (Abitbol et al., 2015). This sort of interference can be seen as a case of misattribution bias (leading us to believe that we like the painting when we actually like the music), which may result in irrational choice behavior. It presupposes the other functional properties: for this interference to occur, OFC-mediated valuation must be generic (applied to both auditory stimuli such as music and visual stimuli such as painting) and automatic (associating value to music even if the task is to judge the value of paintings). It may also involve the property of signaling confidence, since interference between value and confidence has been reported as a case of desirability bias, notably in tasks where confidence in meeting a challenge increases with the reward at stake (Lench et al., 2014; Tappin et al., 2017). It is tempting to relate interference between successive value judgments to the double value coding posited in our neural network. As we have shown, for any random wiring, it is possible to rewire the network such that value coding is reliable: it ensures that the value does not depend on the position (current or previous item) and on the value or the other option. However, if we relax this constraint on the network, then unwanted interferences between values might occur, opening the possibility of misattribution bias and ensuing irrational choice behavior.

Conclusion

In this paper, we have suggested a functional architecture for a neural network to account for how the OFC generates value signals. This neural network has the merits of 1) reconciling value-construction and position-in-task-space accounts of OFC function, 2) unifying the contribution of the OFC to both likeability rating and economic choice, 3) bridging the gap between single-unit and population-level OFC recordings, 4) explaining distortions of value judgment following OFC damage. In addition, our OFC neural network is compatible with the core functional properties attributed to the brain valuation system (generality, automaticity and hysteresis) and to the additional construction of confidence signals. At the neural level, our

model makes specific predictions, notably that the values of currently and previously attended options should be both decodable from OFC activity but with different vectors of weights. At the behavioral level, it predicts that choice consistency should be degraded when options switch their position in a sequential presentation, particularly after damage to the OFC.

However, we have left in the dark many processes, both upstream and downstream to the OFC network. On the input side, the principles underlying the decomposition into features are yet to be specified. A purely analytical decomposition of the object itself, into dimensions like for instance nutrients concentration for food items, is not very plausible. It is likely that associations to previous experiences, with similar objects stored in memory, play a role. Another kind of decomposition would involve dimensions such as the degree of similarity with a particular cluster of objects whose value is already well established. We have left open the anatomical location of the input and integration layers of the OFC network. These layers could be homogeneously implemented all over the OFC, or segregated between distinct areas, for instance in more lateral parts for input features and more medial ones for output values. More generally, what specific information (about which features) is integrated in the different OFC subregions requires further exploration. A possible line of divide could stand between the medial OFC, processing internal information (coming from the body or memory/imagination), and the lateral OFC, processing external information (coming from perceptions of the environment).

On the output side, the processes through which values are mapped onto a rating scale, or compared so as to select the best option, are still to be specified. We have excluded neural competition processes from our network, but these may very well be implemented within the OFC, on top of value construction. The alternatives would be that the comparison of options takes place in other brain regions linked to cognitive control, such as the dorsolateral prefrontal cortex, or closer to motor outputs, such as the supplementary motor area. Further theoretical

and experimental work is needed to understand how made-in-OFC value signals are used to guide the behavior. Finally, we have not addressed the important issue of how the mapping from features to value is learned. Our network model accounts for a steady state, in which the weights are stable. If values change across contexts in this steady state, it is because different features are integrated in the network, not because weights are adapted. Yet the question of how the weights are tuned for the network to generate adaptive values remains a critical puzzle.

References

- Abitbol, R., Lebreton, M., Hollard, G., Richmond, B. J., Bouret, S., & Pessiglione, M. (2015). Neural Mechanisms Underlying Contextual Dependency of Subjective Values: Converging Evidence from Monkeys and Humans. *Journal of Neuroscience*, *35*(5), 2308–2320.
- Aridan, N., Malecek, N. J., Poldrack, R. A., & Schonberg, T. (2019). Neural correlates of effort-based valuation with prospective choices. *NeuroImage*, *185*, 446–454.
- Bao, X., Gjorgieva, E., Shanahan, L. K., Howard, J. D., Kahnt, T., & Gottfried, J. A. (2019). Grid-like Neural Representations Support Olfactory Navigation of a Two-Dimensional Odor Space. *Neuron*, *102*(5), 1066-1075.e5.
- Bartra, O., McGuire, J. T., & 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.
- Bastin, J., Lebranchu, P., Jerbi, K., Kahane, P., Orban, G., Lachaux J.P., Berthoz, A. (2012). Direct recordings in human cortex reveal the dynamics of gamma-band [50–150Hz] activity during pursuit eye movement control. *NeuroImage* *63*, 339-347.
- Camille, N., Griffiths, C. A., Vo, K., Fellows, L. K., & Kable, J. W. (2011). Ventromedial Frontal Lobe Damage Disrupts Value Maximization in Humans. *Journal of Neuroscience*, *31*(20), 7527–7532.
- Chib, V. S., Rangel, A., Shimojo, S., & O’Doherty, J. P. (2009). Evidence for a Common Representation of Decision Values for Dissimilar Goods in Human Ventromedial Prefrontal Cortex. *Journal of Neuroscience*, *29*(39), 12315–12320.
- Cisek, P., & Kalaska, J. F. (2010). Neural Mechanisms for Interacting with a World Full of Action Choices. *Annual Review of Neuroscience*, *33*(1), 269–298.
- Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, *9*(9), 1289–1302.
- Constantinescu, A. O., O’Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468.
- De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nature Neuroscience*, *16*(1), 105–110.
- Fleming, S. M., Thomas, C. L., & Dolan, R. J. (2010). Overcoming status quo bias in the human brain. *Proceedings of the National Academy of Sciences*, *107*(13), 6005–6009.
- Fouragnan, E., Retzler, C., & Philiastides, M. G. (2018). Separate neural representations of prediction error valence and surprise: Evidence from an fMRI meta-analysis. *Human Brain Mapping*, *39*(7), 2887–2906.
- Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *37*(7), 1297–

1310.

Gherman, S., & Philiastides, M. G. (2018). Human VMPFC encodes early signatures of confidence in perceptual decisions. *ELife*, *7*, e38293.

Grueschow, M., Polania, R., Hare, T. A., & Ruff, C. C. (2015). Automatic versus Choice-Dependent Value Representations in the Human Brain. *Neuron*, *85*(4), 874–885.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System. *Science*, *324*(5927), 646–648.

Harvey, A. H., Kirk, U., Denfield, G. H., & Montague, P. R. (2010). Monetary Favors and Their Influence on Neural Responses and Revealed Preference. *Journal of Neuroscience*, *30*(28), 9597–9602.

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, *14*(7), 933–939.

Hebscher, M., Barkan-Abramski, M., Goldsmith, M., Aharon-Peretz, J., & Gilboa, A. (2016). Memory, Decision-Making, and the Ventromedial Prefrontal Cortex (vmPFC): The Roles of Subcallosal and Posterior Orbitofrontal Cortices in Monitoring and Control Processes. *Cerebral Cortex*, *26*(12), 4590–4601.

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F. S., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience*, *15*(3), 470–476.

Ito, A., Yoshida, K., Takeda, K., Sawamura, D., Murakami, Y., Hasegawa, A., Sakai, S., & Izuma, K. (2020). The role of the ventromedial prefrontal cortex in automatic formation of impression and reflected impression. *Human Brain Mapping*, *41*(11), 3045–3058.

Juechems, K., & Summerfield, C. (2019). Where Does Value Come From? *Trends in Cognitive Sciences*, *23*(10), 836–850.

Kim, H., Adolphs, R., O'Doherty, J. P., & Shimojo, S. (2007). Temporal isolation of neural processes underlying face preference decisions. *Proceedings of the National Academy of Sciences*, *104*(46), 18253–18258.

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science*, *336*(6077), 95–98.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857.

Krajbich, Ian, Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298.

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, *18*(8), 1159–1167.

- Lebreton, M., Bertoux, M., Boutet, C., Lehericy, S., Dubois, B., Fossati, P., & Pessiglione, M. (2013). A Critical Role for the Hippocampus in the Valuation of Imagined Outcomes. *PLoS Biology*, *11*(10), e1001684.
- Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron*, *64*(3), 431–439.
- Lench, H. C., Smallman, R., Darbor, K. E., & Bench, S. W. (2014). Motivated perception of probabilistic information. *Cognition*, *133*(2), 429–442.
- Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, *22*(6), 1027–1038.
- Levy, I., Lazzaro, S. C., Rutledge, R. B., & Glimcher, P. W. (2011). Choice from Non-Choice: Predicting Consumer Preferences from Blood Oxygenation Level-Dependent Signals Obtained during Passive Viewing. *Journal of Neuroscience*, *31*(1), 118–125.
- Lim, S.-L., O’Doherty, J. P., & Rangel, A. (2011). The Decision Value Computations in the vmPFC and Striatum Use a Relative Value Code That is Guided by Visual Attention. *Journal of Neuroscience*, *31*(37), 13214–13223.
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150-157.
- Lopez-Gamundi, P., Yao, Y.-W., Chong, T. T.-J., Heekeren, H. R., Herrero, E. M., & Pallares, J. M. (2021). The neural basis of effort valuation: A meta-analysis of functional magnetic resonance imaging studies [Preprint]. *bioRxiv*. Neuroscience.
- Lopez-Persem, A., Bastin, J., Petton, M., Abitbol, R., Lehongre, K., Adam, C., Navarro, V., Rheims, S., Kahane, P., Domenech, P., & Pessiglione, M. (2020). Four core properties of the human brain valuation system demonstrated in intracranial signals. *Nature Neuroscience*, *23*(5), 664–675.
- Lopez-Persem, A., Domenech, P., & Pessiglione, M. (2016). How prior preferences determine decision-making frames and biases in the human brain. *ELife*, *5*, e20317.
- Morrison, S. E., & Salzman, C. D. (2009). The Convergence of Information about Rewarding and Aversive Stimuli in Single Neurons. *Journal of Neuroscience*, *29*(37), 11471-11483.
- O’Doherty, J. P. (2014). The problem with value. *Neuroscience & Biobehavioral Reviews*, *43*, 259–268.
- Padoa-Schioppa, C. (2011). Neurobiology of Economic Choice: A Good-Based Model. *Annual Review of Neuroscience*, *34*(1), 333–359.
- Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, *441*(7090), 223–226.

- Padoa-Schioppa, C., & Conen, K. E. (2017). Orbitofrontal Cortex: A Neural Circuit for Economic Decisions. *Neuron*, *96*(4), 736–754.
- Palminteri, S., Boraud, T., Lafargue, G., Dubois, B., & Pessiglione, M. (2009). Brain Hemispheres Selectively Track the Expected Value of Contralateral Options. *Journal of Neuroscience*, *29*(43), 13465–13472.
- Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Czernecki, V., Karachi, C., Capelle, L., Durr, A., & Pessiglione, M. (2012). Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. *Neuron*, *76*(5), 998–1009.
- Papageorgiou, G. K., Sallet, J., Wittmann, M.K., Chau, B.K.H., Schuffelgen, U., Buckley, M.J., Rushworth, M.F.S. (2017). Inverted activity patterns in ventromedial prefrontal cortex during value-guided decision-making in a less-is-more task. *Nature Communications*, *8*(1), 1886.
- Pauli, W. M., Gentile, G., Collette, S., Tyszka, J. M., & O’Doherty, J. P. (2019). Evidence for model-based encoding of Pavlovian contingencies in the human brain. *Nature Communications*, *10*(1), 1099.
- Pelletier, G., & Fellows, L. K. (2019). A Critical Role for Human Ventromedial Frontal Lobe in Value Comparison of Complex Objects Based on Attribute Configuration. *The Journal of Neuroscience*, *39*(21), 4124–4132.
- Peters, J., & Büchel, C. (2010). Neural representations of subjective reward value. *Behavioural Brain Research*, *213*(2), 135–141.
- Piguet, O., Hornberger, M., Mioshi, E., & Hodges, J. R. (2011). Behavioural-variant frontotemporal dementia: Diagnosis, clinical staging, and management. *The Lancet Neurology*, *10*(2), 162–172.
- Plassmann, H., O’Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences*, *105*(3), 1050–1054.
- Pleskac, T. J. (2015). Decision and Choice: Luce’s Choice Axiom. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 895–900). Elsevier.
- Polanía, R., Woodford, M., & Ruff, C. C. (2019). Efficient coding of subjective value. *Nature Neuroscience*, *22*(1), 134–142.
- Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: Knowns and unknowns. *Nature Neuroscience*, *16*(9), 1170–1178.
- Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion Decision Model: Current Issues and History. *Trends in Cognitive Sciences*, *20*(4), 260–281.
- Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J., & Hsiao, S. S. (2008). Neural Correlates of High-Gamma Oscillations (60–200 Hz) in Macaque Local Field Potentials and Their Potential Implications in Electrocorticography. *Journal of Neuroscience*, *28*(45), 11526–

11536.

Rich, E. L., & Wallis, J. D. (2014). Medial-lateral Organization of the Orbitofrontal Cortex. *Journal of Cognitive Neuroscience*, 26(7), 1347–1362.

Rustichini, A., & Padoa-Schioppa, C. (2015). A neuro-computational model of economic decisions. *Journal of Neurophysiology*, 114, 1382-1398.

Saez, I., Lin, J., Stolk, A., Chang, E., Parvizi, J., Schalk, G., Knight, R. T., & Hsu, M. (2018). Encoding of Multiple Reward-Related Computations in Transient and Sustained High-Frequency Activity in Human OFC. *Current Biology*, 28(18), 2889-2899.e3.

Scheeringa, R., Fries, P., Petersson, K.M., Oostenveld, R., Grothe, I., Norris, D.G., Hagoort, P., Bastiaansen, C.M. (2010). Neuronal Dynamics Underlying High- and Low-Frequency EEG Oscillations Contribute Independently to the Human BOLD Signal. *Neuron*, 69, 572–583.

Schneider, B., & Koenigs, M. (2017). Human lesion studies of ventromedial prefrontal cortex. *Neuropsychologia*, 107, 84–93.

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, 91(6), 1402–1412.

Strait, C. E., Blanchard, T. C., & Hayden, B. Y. (2014). Reward Value Comparison via Mutual Inhibition in Ventromedial Prefrontal Cortex. *Neuron*, 82(6), 1357–1366.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*.

Suzuki, S., Cross, L., & O’Doherty, J. P. (2017). Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nature Neuroscience*, 20(12), 1780–1786.

Tajima, S., Drugowitsch, J., Patel, N., & Pouget, A. (2019). Optimal policy for multi-alternative decisions. *Nature Neuroscience*, 22(9), 1503–1511.

Tappin, B. M., van der Leer, L., & McKay, R. T. (2017). The heart trumps the head: Desirability bias in political belief revision. *Journal of Experimental Psychology: General*, 146(8), 1143–1149.

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The Neural Basis of Loss Aversion in Decision-Making Under Risk. *Science*, 315(5811), 515–518.

Tremblay, L., & Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature*, 398(6729), 704–708.

Tsetsos, K., Wyart, V., Shorkey, S. P., & Summerfield, C. (2014). Neural mechanisms of economic commitment in the human medial prefrontal cortex. *ELife*, 3, e03701.

Vaidya, A. R., Sefranek, M., & Fellows, L. K. (2018). Ventromedial Frontal Lobe Damage Alters how Specific Attributes are Weighed in Subjective Valuation. *Cerebral Cortex*, 28(11), 3857–3867.

Vinckier, F., Rigoux, L., Oudiette, D., & Pessiglione, M. (2018). Neuro-computational account of how mood fluctuations arise and affect decision making. *Nature Communications*, 9(1), 1708.

Wang, X.-J. (2002). Probabilistic Decision Making by Slow Reverberation in Cortical Circuits. *Neuron*, 36(5), 955–968.

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., & Niv, Y. (2014). Orbitofrontal Cortex as a Cognitive Map of Task Space. *Neuron*, 81(2), 267–279.

Wong, K.-F. (2006). A Recurrent Network Mechanism of Time Integration in Perceptual Decisions. *Journal of Neuroscience*, 26(4), 1314–1328.

Wunderlich, K., Rangel, A., & O’Doherty, J. P. (2009). Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences*, 106(40), 17199–17204.

Supplementary methods:

Likeability rating

What we describe hereafter is an artificial neural network (ANN) that transforms a set of features into a value signal that can be decoded to provide a response in a likeability rating task presenting a novel item on every trial (see illustration in Fig. 4A).

Without loss of generality, let $[u_t^{(1)}, u_t^{(2)}, \dots, u_t^{(i)}, \dots, u_t^{(n)}]$ be the coordinates of the item under the attentional focus (in an nD-features space), at time or trial t .

We assume that the input feature $u_t^{(i)}$ is encoded into the activity of neurons $[x_t^{(i,1)}, x_t^{(i,2)}, \dots, x_t^{(i,j)}, \dots, x_t^{(i,m_i)}]$ of its dedicated "input layer", where m_i is the number of neurons in the input layer i . What we mean here is that the neuron j in the input layer i responds to $u_t^{(i)}$ as follows:

$$x_t^{(i,j)} = f(u_t^{(i)}, \theta_{i,j}) \quad (1)$$

where $f(\bullet, \theta_j)$ is the neuronal activation function (here either a sigmoid or a gaussian mapping). The parameters $\theta_{i,j}$ capture idiosyncratic properties of the neuron j in the input layer i (e.g., its firing rate threshold). This means that, collectively, the activity vector $[x_t^{(i,j)}]_{j=1, \dots, m_i}$ forms a representation of input $u_t^{(i)}$ in the form of a population code.

Then activity is passed to the so-called integration layer $[z_t^{(1)}, z_t^{(2)}, \dots, z_t^{(k)}, \dots, z_t^{(q)}]$, which mixes and rescales the representations of input features $u_t^{(i)}$, i.e. the neuron k of the integration layer responds to

$$[x_t^{(i,j)}]_{j=1, \dots, m_i}^{i=1, \dots, n} \text{ as follows:}$$
$$z_t^{(k)} = f\left(\sum_{i=1}^n \sum_{j=1}^{m_i} C_k^{(i,j)} x_t^{(i,j)}, \phi_k\right) \quad (2)$$

where $C_k^{(i,j)}$ is the connection weight from the neuron j in the input layer i to the neuron k of the integration layer, and ϕ_k capture idiosyncratic properties of the integration neuron k .

The value V_t of the attended item at time t is then read out from the integration layer as follows:

$$V_t = \sum_{k=1}^q W^{(k)} z_t^{(k)} \quad (3)$$

where the $W^{(k)}$ can be thought of as connection weights to another system that would use the value signal (e.g., to make decisions).

Crucially, value is not *encoded* by this ANN. Rather, it is *constructed* out of input features, in the sense that what the ANN does is to transform the input features and output value. This can be seen by replacing Equations 1-2 into Equation 3:

$$V_t = \sum_{k=1}^q W^{(k)} f \left(\sum_{i=1}^n \sum_{j=1}^{m_i} C_k^{(i,j)} f(u_t^{(i)}, \theta_{i,j}), \phi_k \right) \quad (4)$$

Provided there are enough neurons in input and integration layers, this ANN architecture can capture any value function defined on the nD-feature space.

The first question we may want to ask with this model is why does aggregated OFC population activity (as measured with, e.g., fMRI) positively correlate with value whereas single-unit recordings do not?

Let $y_t = \sum_{k=1}^q z_t^{(k)} + \sum_{i=1}^n \sum_{j=1}^{m_i} x_t^{(i,j)}$ be the sum of activity across all ANN units. In our simulations, we take

y_t as a proxy for aggregated OFC population activity. For the simulation results presented in Fig.4B, we used a network structure of 128 units in total. We systematically varied the ratio of input vs. integration units and the ratio of positive weights in the decoder (from 0 to 100%, with 5% steps). For each ratio, we randomly draw the input-to-integration connectivity matrix C , and the population code W . We then simulate 100 trials, each consisting in a set of random input features. Correlations with value, for both single-unit and aggregated activities, are calculated across the 100 trials. The procedure is then repeated 1000 times, meaning that each ratio (i.e., each cell in Fig. 4B) involves 1000 simulations of 100 trials, each time with a different network structure. It turns that, under very weak assumptions, the correlation between value and single-unit activity in the ANN may be zero on average (across ANN units), still there may be a stable positive correlation between value and y_t .

Binary choice

So far, we focused on situations where the OFC only has to map a value to attended items. Here, we consider situations in which the brain has to perform a value-based choice between two (or more) items. We assume that the mechanism described above is used to map the value of any element of the choice set that is currently attended. However, we consider an additional process, by which value is transiently stored in short-term memory, using a different population code (but using the same neurons). As we will see, this is made possible through recurrent feedback connections on the integration layer (see illustration in Fig. 5B).

Recall that Equation 3 expresses how value is constructed from activity in the integration layer. In the context of choices, where more than one option is being valued, Equation 3 expresses the value V_t of the *currently attended* item. But a different population code may be used to decode the value V_{t-1} of the *previously* attended item:

$$\begin{aligned} V_t &= \sum_{k=1}^q W_0^{(k)} z_t^{(k)} \\ V_{t-1} &= \sum_{k=1}^q W_1^{(k)} z_t^{(k)} \end{aligned} \tag{6}$$

where W_0 and W_1 are the effective weights of the value population code for the currently and previously attended item, respectively. Note that both current and previous values are read from the same activity z_t in the integration layer.

We now ask whether we can find a re-wiring of the ANN that makes Equation 6 hold, irrespective of the population codes W_0 and W_1 .

Let us start with a linear simplification of the above ANN, whereby neural activation functions of the integration layer can be written as: $f(x, \phi_k) \approx a_k + b_k x$, where a_k and b_k are arbitrary constants. We also consider that previous activity of the integration layer is fed back to the integration layer through recurrent connections. Equation 2 thus takes the following form:

$$z_t = A + \text{diag}(B)(\Psi Cx_t + \Omega z_{t-1} + \omega) \quad (7)$$

where Ψ effectively re-wires the input layers to the integration layer, Ω is a recurrent feedback connection, ω is a baseline input pattern, and $A = [a_1 \ a_2 \ \dots \ a_q]^T$ and $B = [b_1 \ b_2 \ \dots \ b_q]^T$ are the vectorized affine coefficients.

The question now reduces to finding connection matrices Ψ and Ω (and baseline input pattern ω) such that Equation 6 is verified, irrespective of the network state.

To simplify notations, let us set $\tilde{W}_0 = \text{diag}(B)W_0$ and $\tilde{W}_1 = \text{diag}(B)W_1$.

Replacing Equation 7 into Equation 6 then yields:

$$\begin{aligned} V_t &= W_0^T A + \tilde{W}_0^T \Psi Cx_t + \tilde{W}_0^T \Omega z_{t-1} + \tilde{W}_0^T \omega \\ V_{t-1} &= W_1^T A + \tilde{W}_1^T \Psi Cx_t + \tilde{W}_1^T \Omega z_{t-1} + \tilde{W}_1^T \omega \\ &= W_1^T A + \tilde{W}_1^T \Psi Cx_t + \tilde{W}_1^T \Omega \left(A + \text{diag}(B)(\Psi Cx_{t-1} + \Omega z_{t-2} + \omega) \right) + \tilde{W}_1^T \omega \end{aligned} \quad (8)$$

First, note that one does not want the value of the previously attended option to depend upon the current state of the input layer, i.e. $\partial V_{t-1} / \partial x_t = 0$. From the second line of Equation 8, this yields: $\Psi^T \tilde{W}_1 = 0_q$. In other words, \tilde{W}_1 should belong to the null space of Ψ . A simple solution is to define Ψ as the residual-forming projector matrix of \tilde{W}_1 , as follows:

$$\Psi = I_q - \tilde{W}_1 \left(\tilde{W}_1^T \tilde{W}_1 \right)^{-1} \tilde{W}_1^T \quad (9)$$

Second, one wants that the value of the currently attended option to depend only upon the current state of the integration layer. In other words, V_t should not depend upon z_{t-1} , i.e.: $\partial V_t / \partial z_{t-1} = 0$. From the first line of Equation 8, this implies that: $\Omega^T \tilde{W}_0 = 0_q$ (i.e. \tilde{W}_0 belongs to the null space of Ω).

Third, one wants that the value of the previously attended option, when decoded with the population code W_1 , equates the previous attended value, i.e.: $V_{t-1} = W_1^T z_t = W_0^T z_{t-1}$. Thus, the following condition must be verified:

$$W_1^T A + \tilde{W}_1^T \Omega \left(A + \text{diag}(B) (\Psi C x_{t-1} + \Omega z_{t-2} + \omega) \right) + \tilde{W}_1^T \omega = W_0^T A + \tilde{W}_0^T \Psi C x_{t-1} + \tilde{W}_0^T \Omega z_{t-2} + \tilde{W}_0^T \omega \quad (10)$$

where we have used Equation 9 to enforce the constraint $\Psi^T \tilde{W}_1 = 0_q$.

Equation 10 must hold irrespective of z_{t-2} and x_{t-1} , which yields:

$$\frac{\partial}{\partial z_{t-2}} W_1^T z_t = \frac{\partial}{\partial z_{t-2}} W_0^T z_{t-1} \Rightarrow \Omega^T \tilde{W}_1 = W_0 \quad (11)$$

(The condition one gets with x_{t-1} is slightly more general but holds if Equation 11 holds). Equation 11 simply means that Ω maps \tilde{W}_1 onto W_0 .

Taken together, the recurrent feedback connection matrix Ω has to satisfy two constraints. A solution to this problem is given by:

$$\Omega = \begin{bmatrix} \tilde{W}_1 & \tilde{W}_0 \end{bmatrix} \left(\begin{bmatrix} \tilde{W}_1 & \tilde{W}_0 \end{bmatrix}^T \begin{bmatrix} \tilde{W}_1 & \tilde{W}_0 \end{bmatrix} \right)^{-1} \begin{bmatrix} W_0 & 0_q \end{bmatrix}^T \quad (12)$$

Replacing Equation 11 into Equation 10 finally yields:

$$\omega = -\text{diag}(B)^{-1} A \quad (13)$$

Equations 9, 12 and 13 solve the problem of how to re-wire the ANN such that applying the population codes W_0 and W_1 recover the value of the currently and previously attended option, respectively. In principle, this means that the ANN can always re-wire itself (through, e.g., hebbian plasticity) to make sure that downstream decoding of the currently and previously attended options are possible. Note that, in theory, the value memory span of OFC is not limited to one item. The same principle can be applied, eventually yielding a value memory span of the same order of magnitude as the number of units in the integration layer. However, the necessary re-wiring of the network may become too complex to sustain large memory spans.

The previous short-term memory mechanism enables a value comparison, which relies upon reading the value of both the currently and the previously attended option from the current activity in the integration layer.

Transitivity violations

Our last question is how the properties of the ensuing decisions would change as increasingly wide anatomical lesions distort the value mapping. Intuitively, if lesions randomly shut neuronal integration units down, then this will not impact the value of the currently and the previously attended item in the same way. In turn, this might cause some form of decision irrationality, e.g., violations of preference transitivity. One would thus expect that the rate of preference transitivity violations increases with the amount of lesioned units in the integration layer. We note that lesions on the input layers should not induce preference transitivity violations, because it would not yield differences between value codes of the previously and the currently attended items.

Recall that transitive preferences mean that if $V_1 > V_2$ and $V_2 > V_3$, then $V_1 > V_3$. In what follows, we measure the rate at which such relations are violated, as a function of the proportion of lesioned integration units. To simplify things, we consider linear (i.e. affine) neuronal activation functions. Our network structure involves 2 inputs, 32 units per input, and 64 integration units. In each simulation, we randomly draw offset and slope parameters (A and B), the input-to-integration connectivity matrix C , and the population codes W_0 and W_1 . We then simulate 100 “trials”, each of which consists of 3 pairwise comparisons of 3 items with random input features. Each of these trial series is repeated with increasing ratios of lesioned units in the integration layer (from 0% to 100%, with steps of 5%). For each lesion ratio, we measure the rate of preference transitivity violations. To average over the network structure, we repeat this procedure 1000 times, and average the results over all simulations. The results (Fig. 5B) are summarized in terms of the rate of preference transitivity violations as a function of lesion ratio, associated with the average dispersion of compared values.