From decisions to aesthetic values in cognitive neuroscience

Pradyumna Sepúlveda

1. Italian Academy for Advanced Studies in America, Columbia University
2. Department of Psychiatry, Columbia University

From ancient cave paintings to millions of Instagram pictures, aesthetic experience has always been a fundamental constituent of human lives. Yet, how the human brain computes aesthetic value from visual stimuli is largely unknown. In cognitive neuroscience, the study of decisions has shown that neurons in an extensive network of brain regions can track the value of stimuli in various types of decisions. While many studies have focused on how the brain updates this value based on its associated reward or punishment, much less progress has been made on an even more fundamental problem: how do we come to assign value to stimuli in the first place? While multiple mechanisms likely play a role in forming value, humans can express preferences for completely novel stimuli (e.g., a new painting seen for the first time at an art gallery). This suggests that value judgments are a much more active and dynamic process that does not depend solely on prior associative learning. Indeed, additional factors such as contextual influences and dynamic attention are key in value-based decisions.

The goal of my project is to gain an understanding of the computational principles by which value judgments are formed. For this purpose, I focus on the aesthetic value judgment of visual artworks. In the present paper, I briefly introduce some background on the study of decision-making research in cognitive science and neuroscience, highlighting some relevant aspects that can be used to understand aesthetic value computation. In the final part, I present a computational model that describes how visual aesthetic judgments could be constructed, an idea I will further elaborate on during my project at the Italian Academy.

Perceptual decisions

Even the most trivial decisions, like choosing coffee or tea, are extremely intricate, involving obscure and complex interactions of multiple factors. For this reason, initial neuroscientific efforts to formalize a methodology to study decisions rely mainly on the systematic analysis of perceptual choices, i.e., psychophysics (Shadlen and Kiani, 2013). In simple perceptual decisions, subjects use sensory stimuli, such as the contrast of lines, shades of color, or dot motion to make their choices. For example, subjects may be asked to look at two circles on a screen and instructed to select the option with higher luminosity or to look at a cloud of randomly moving dots and report the direction of apparent movement. This setup allows the experimenter a high degree of control over the evidence used in the decisions: it is more accessible and reproducible to control light intensity than the personal experiences that make you choose coffee over tea. These perceptual studies imply that a vast amount of sensory information is integrated into internal representations of relevant dimensions, e.g., there is an internal coding that tracks variations in luminosity and uses this to make the choices. Further work has shown evidence that this representation could be reflected in specific patterns in the brain, with the study of the visual system being pivotal in understanding how sensory evidence is transformed into neural signals. For example, monkeys performing demanding visual discrimination tasks have shown that activity in area MT/V5, a region of the visual cortex associated with motion processing, can track decision variability, even at a single neuronal level (Newsome et al., 1989). In this way, neuroscientists have described a sophisticated mapping of how various areas of the brain, especially the occipital cortex, process visual information in a hierarchical way, from the basic orientation of edges and contrast to more complex representations of objects and faces, and use this information to make choices.

Value-based decisions

Having these insights from the study of perceptual decisions, cognitive scientists have attempted to explore those choices that depend on information beyond mere sensory information. Value-based decisions involve selecting among different options according to the subjective valence of these alternatives. From a bee foraging to humans trading in the stock market, all these agents could be considered as making decisions in this category (Rangel et al., 2008). The concept of value is pivotal in many disciplines, overlapping with other ideas such as utility or reward. Therefore, the study of value decisions in neuroscience has been a joint effort combining models and methods developed in economics, biological principles, behavioral observations and theories from animal learning and psychology, and computational frameworks from computer science. At the core of this exploration is the idea that an internal representation, a *value* signal, guides choices to maximize future rewards. In other words, to make appropriate decisions, these values should be reliable predictors of the benefits that could result from each action.

This value concept allows us to compare objects that could not be commensurable in straight sensory or quantitative terms and bring them to the same scale. For example, in an economic decision, it is possible to compare the value of vastly distinct objects such as the Mona Lisa, the International Space Station, and a banana. This abstract nature of value has sometimes been denominated as having a “common currency” in the brain (Levy and Glimcher, 2012). To understand how these values can be learned, animal and machine learning knowledge has been integrated into the “reward hypothesis”. This proposal states that the environment provides reinforcement signals that indicate the probable costs or benefits of states and actions, which are sensed by the agents, allowing them to act in ways that maximize their utility over the short or long term (Juechems and Summerfield, 2019). This hypothesis has been developed and formalized in the reinforcement learning (RL) model, one of the standards for studying value-based choices in neuroscience (see computational framework section).

The computations in value-based decision-making have been divided into five general processes (Rangel et al., 2008). First, the representation stage considers that agents must identify internal and external states and potential actions. Second, each represented state/action needs to be assessed in the valuation stage. Third, using the assigned values, a comparison is performed to inform the action selection stage. Fourth, the outcome evaluation stage after choice assesses the desirability of the collected outcomes using the resulting feedback. Finally, the collected outcomes are used in a learning stage to update the three initial stages and generate better choices in the future. Note that this sequence assumes that learning is a continuous process, i.e., each new decision informs future choices.

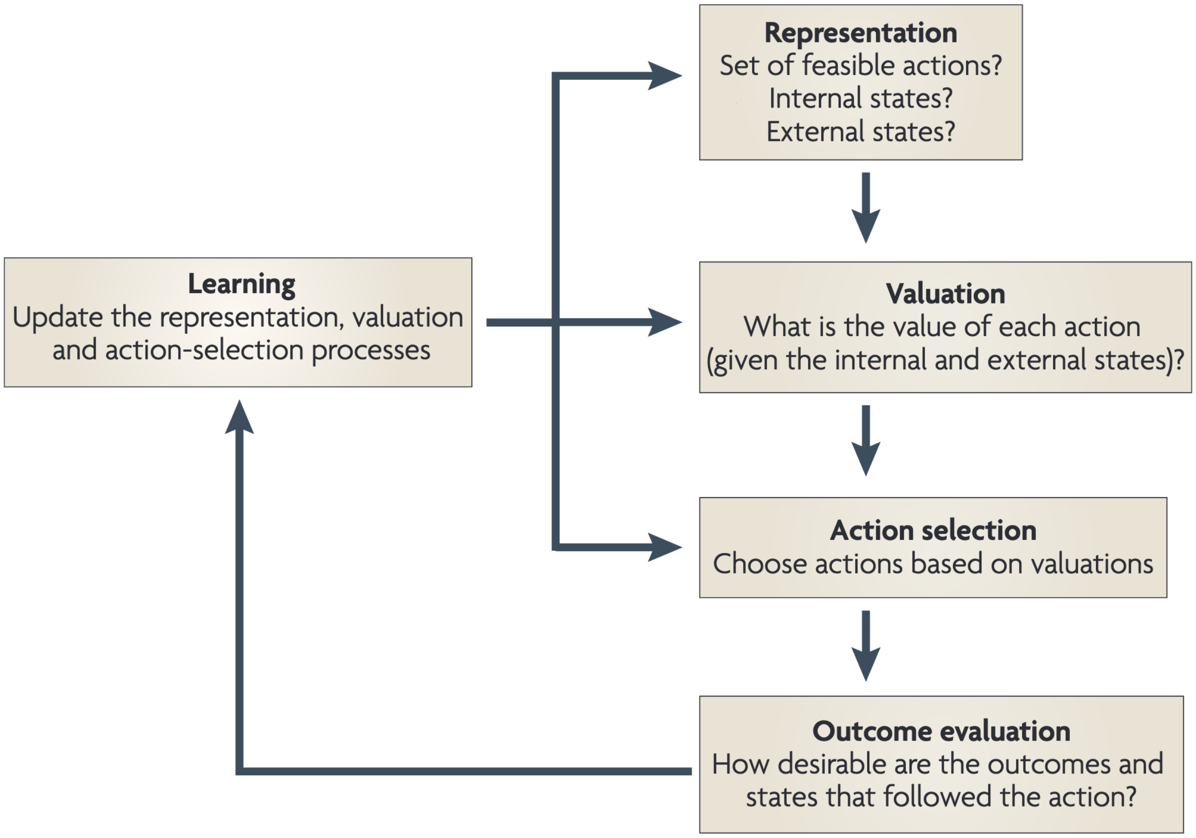
Representation
Set of feasible actions?
Internal states?
External states?
Learning
Update the representation, valuation and action-selection processes
Valuation
What is the value of each action (given the internal and external states)?
Action selection
Choose actions based on valuations
Outcome evaluation
How desirable are the outcomes and states that followed the action?

Figure 1. Basic computations required for value-based choice. From Rangel et al (2008).

The level of insight into these stages is quite dissimilar. How the brain represents internal and external states (e.g., how the possible options are identified for each situation) is a work in progress since this is a complex process that requires the integration of various sources of information. The presence of valuation systems has been generally accepted, although their exact neural implementation and characteristics are an ongoing area of research. Some standard proposals for valuation systems include the *Pavlovian system*, which assigns prominence to a restricted set of evolutionarily relevant actions, such as preparatory behaviors for food or avoiding aversive stimuli, such as heat or electric shocks. For this reason, the behavioral repertoire of Pavlovian responses is rather limited and inflexible. In contrast, the *habitual valuation system* can cover a more comprehensive set of actions, assuming they are repeatedly experienced. In the habitual system, stimulus-response associations are learned, relying on past experiences through a trial-and-error training process. This makes habitual responses a slow learning process but fast (or automatic) to deploy once learning is achieved. For example, this category includes a rat learning to press a lever for liquids in response to a sound or a smoker's desire for a cigarette after a meal. Finally, in the *goal-directed valuation system*, values depend on the association between states/actions and outcomes. Since each action can lead to multiple results, it is assumed that some mapping between actions and outcomes exists, which can be more flexible, not relying entirely on repeated training of automatic responses. However, given a more detailed assessment of the possibilities, they may require a higher investment of cognitive resources. For example, if we need to get to Columbia University and the 1 subway line is broken, we can harness our knowledge of the New York City transit system to look for alternative routes. The separation between these three systems is not necessarily clean-cut since they can interact and be modulated by additional factors, such as risk, uncertainty, and time.

Although most of the research in value-based decision-making has been related to obtaining hedonic rewards, recent work has shown that the value could represent a more general and flexible construct, which aligns with goal-directed valuation. For example, a couple of loose matches could be an inconsequential object in the context of your kitchen, but they could become your most prized possession if you are stranded on an empty island in the middle of the Pacific Ocean. Recent work has shown that the brain network tracking reward value can also contain information about the “usefulness” of an object in experiments with multiple contexts (Castegnetti et al., 2021). Indeed, the prefrontal cortex activity (usually associated with value computation) is reshaped by goals, indicating that the mapping of the value of state/actions can be quickly repurposed depending on task demands (e.g., Frömer 2019, Sepulveda et al., 2020; De Martino and Cortese, 2023).

Computational frameworks in decision making

How the brain operates is, in many ways, a black box. While we can probe the brain's anatomy, this static approach does not fully capture its dynamic flow of information. A method to understand how the brain could operate as an information-processing machine is to use mechanistic modeling of neural and behavioral processes, which has been primarily developed in computational neuroscience. David Marr, one of the founders of computational neuroscience, proposed that the brain can be understood as an information-processing system that can be analyzed from three distinct levels: 1) What computational problem does the brain solve? 2) What computational algorithms (the steps or code to solve the problem) does the brain use? 3) How does the brain implement the algorithm? (Marr, 2010).

Let's say that a person goes to the cafeteria and tries to decide what to eat. In this case, the brain is trying to solve the problem of selecting the best food item for lunch (the first level). There are multiple ways (multiple algorithms) in which this problem could be solved. For example, one way could be to compare the taste across different lunch options and pick the one with a higher taste; another could be just to pick the same snack eaten the last time I was in the cafeteria. To find the most likely algorithm that the brain uses corresponds to Marr's second level. Finally, in the third level, we care about the specific neurons or brain networks that deploy the algorithm, e.g., the firing pattern of neurons in the prefrontal cortex may implement such an algorithm, representing the taste of a food item.

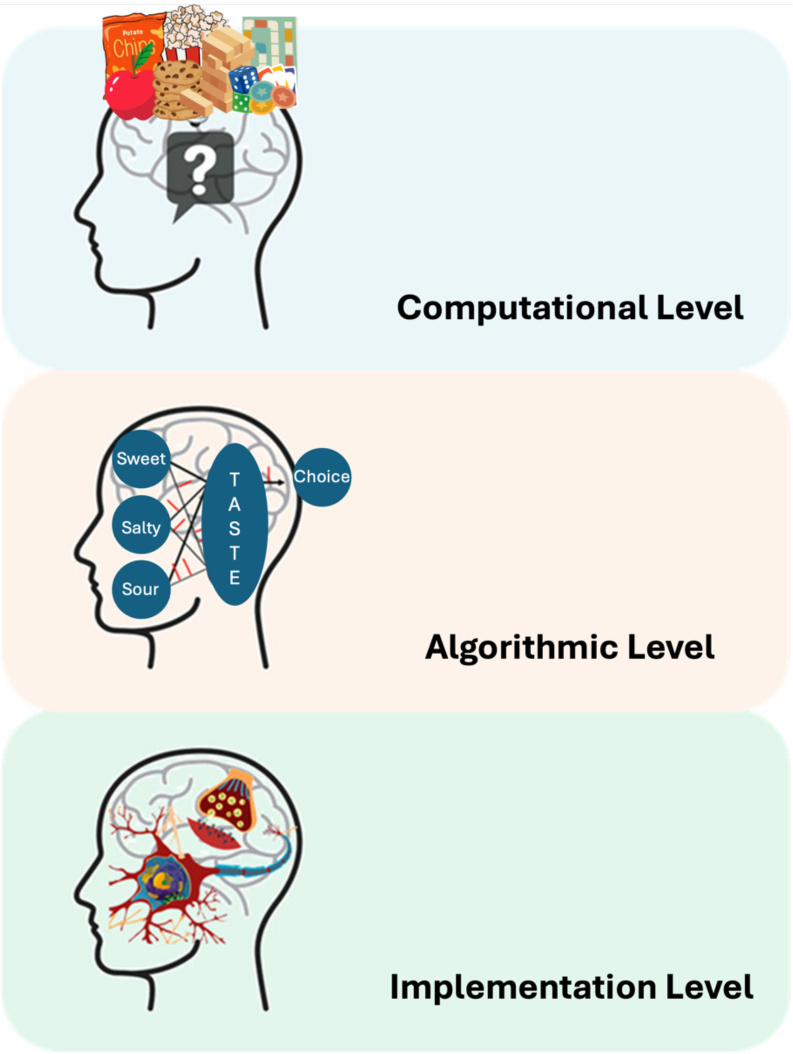

Figure 2. Illustration of Marr's three levels to understand decision-making. A person tries to choose the best snack to eat (goal, computational level). One way of doing this is by assessing different taste dimensions and picking the most satisfactory combination (algorithmic realization). In the brain, this process could be implemented through various neural interactions (physical implementation). There are many ways in which the goal can be achieved, which opens the possibility of multiple algorithms with distinct implementations in the brain. This approach also allows for the discarding of algorithms that do not have biological support.

While these three levels are complementary, most human research using this framework has focused on the second, algorithmic level. Multiple models have been proposed to describe the algorithms used to make decisions. For perceptual decision, Signal Detection Theory (Green and Swets, 1966) is a classic model employed to capture how human and non-human agents can distinguish the presence of a visual target (the signal) among multiple sources of noise, e.g., to report whether a phone is ringing in a loud, crowded room. Another example of a computational model, this time to capture value-based decisions, is reinforcement learning (RL) (Sutton and Barto, 2018), arguably one of the most popular second-level models. This model describes an algorithm so agents can learn the value of diverse objects or states in their environment, following a trial-and-error approach. For example, let's say you arrive in a new neighborhood, and you want to know the best places to eat. Initially, you have no idea if the cafeteria down the street is better than the sandwich shop around the corner, so you try both. You embrace both options with equal expectations. After tasting some awful coffee, you are a bit disappointed and downgrade the value of the cafeteria. When you have a magnificent wrap, you will be happily surprised, and your value for the sandwich shop increases. These positive or negative surprises are called *reward prediction errors* in the RL framework and work as a fundamental teaching signal. After some hits and misses (maybe after an exploration lasting some weeks), you will arrive to have an informed assessment: the sandwich shop is overall the best and most reliable source for lunchtime. In other words, you have assigned different expected values to both options and will use this to make future choices. This is a dynamic and continuous process. At some point, the cafeteria may start to put out amazing pastries and get a new espresso machine, which will be a positive surprise relative to your low expectations, increasing its expected value and making it more likely you will pick this option in the future. In that way, RL models formalize this algorithm and are fundamental for understanding learning in value-based decision-making. The prominence of this model also relies on the finding that the theoretical reward prediction error signal is consistent with the quick responses of the midbrain dopamine neurons (Schultz, 1998) when surprising events arise, which has also been reflected in human MRI experiments showing relevant activations in the ventral striatum (Pagnoni et al., 2002; Seymour et al., 2004; Pessiglione et al., 2006), strongly supporting the model's biological relevance.

The "value" described by the simple version of RL is a singular learned estimate representing an object, state, or action. However, recent studies have demonstrated that value can be a more flexible construct compounded by many features (Hare et al., 2011; Lim et al., 2013). For instance, the overall value we assign to a restaurant could rely on integrating the quality of their appetizers, entrees, desserts, beverages, location, space, etc. This feature separation has been the base of more complex models, where people learn the value of individual features, which can be traced back to specific activity in the brain. These findings have reported that areas in the prefrontal cortex represent individual components of the evaluated items. For example, the nutritious attributes of food items, such as proteins or fat composition, are represented in the orbitofrontal cortex, the lower frontal part of the brain (Suzuki et al., 2016). One caveat of this approach is that the features to describe an object could be unlimited, easily overwhelming the ability of biological and artificial agents to process information. Further studies integrating human behavior and RL in multidimensional environments have shown that attention plays a critical role in selecting the relevant features and discarding irrelevant information, fostering the generation of efficient value representations of the tasks (Niv et al., 2015; Leong et al., 2017; Cortese et al., 2021).

The study of brain activity using model-based neuroimaging

Accessing neural data in humans, which is required to describe more complex cognitive representations, poses an outstanding challenge. One of the main tools to tackle this issue is functional Magnetic Resonance Imaging (fMRI). Since its early days, this technique has fostered an explosion of non-invasive research in cognitive neuroscience (Benedettini, 2012). The principles of MRI operation allow the generation of images based on the magnetic properties of hydrogen nuclei (protons), which are abundant in water molecules. Given the varied composition and structure of different tissue types in the brain, such as grey or white matter, specific magnetic signatures can generate anatomical images of the brain, reaching even the millimetric scale without relying on invasive interventions. In the case of the brain, we also have to consider that neurons are no different from other cells in the body, and they consume energy through oxygen and glucose. The activity of neuronal groups can affect the local supply of oxygen-rich blood, i.e., whenever neurons are firing, they tend to boost the energy supply to that brain area. Given the magnetic properties of hemoglobin, a fundamental molecule in the blood, brain regions with higher neural activation (and higher blood flow) are more magnetically uniform, which increases the MRI signal, and we see them as brighter in the MR images. This is denominated the Blood Oxygen Level Dependent (BOLD) effect (**Figure 3A)**. We can obtain quick sequences of images (in the range of seconds), allowing us to track the dynamic brain changes and the "functional" areas while participants perform tasks inside the scanner, such as observing images or making decisions. In summary, fMRI allows an indirect measure of brain activity (through the changes in metabolic activity) with high spatial resolution.

Combining the computational modeling approach with human neuroimaging allows us to test neural, third-level (implementational) models. The so-called model-based fMRI (O'Doherty et al., 2007) enables us to uncover how the computational model's signals are encoded in the brain. For this, we first identify relevant parameters in our computational models adjusted to human behavior. Using the models, we can then generate predictions about the exact time course of fMRI signals encoding the model's variables. The correlation between the model's predictions and brain signals can be estimated using a standard statistical analysis (Penny et al., 2007). For example, it is possible to generate estimations of participants' expected values using RL models and see brain areas, such as the ventromedial prefrontal cortex (vmPFC), that are "active" following this pattern (Cortese et al., 2012) (Figure 3B). Model-based fMRI can be an important avenue for understanding the decision-making mechanisms and brain representations, including aesthetic value.

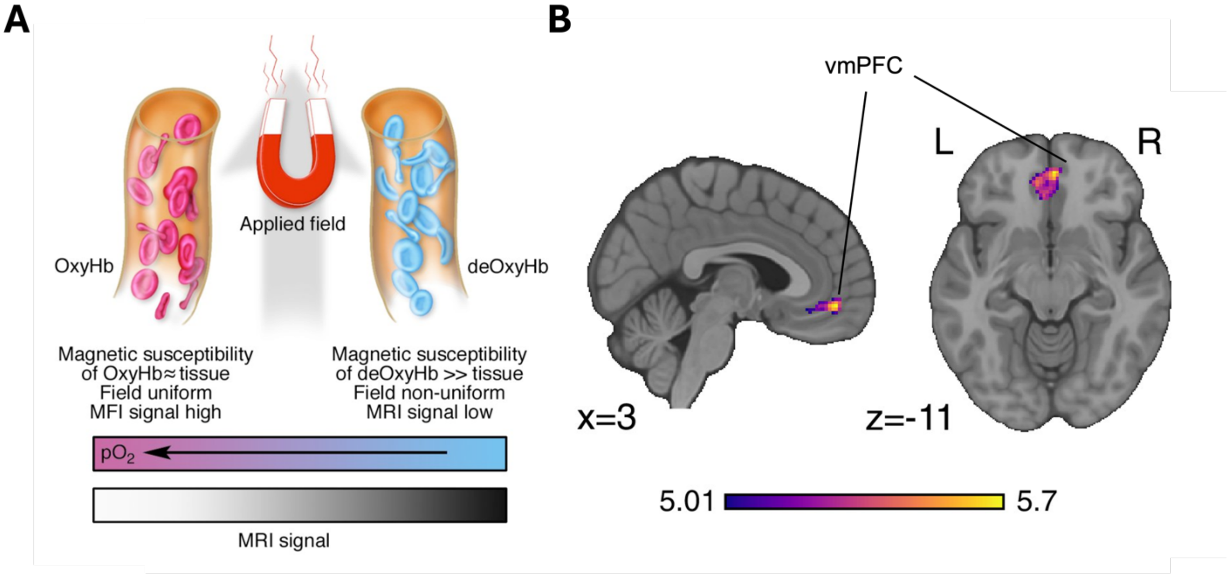
A
Applied field
OxyHb
deOxyHb
Magnetic susceptibility of OxyHb≈tissue Field uniform MFI signal high
Magnetic susceptibility of deOxyHb >> tissue Field non-uniform MRI signal low
pO_2
MRI signal
B
vmPFC
L
R
x=3
z=-11
5.01
5.7

Figure 3. (A) Hemodynamic variations by blood flow. Origin of the BOLD effect in the brain. Arterial blood is similar to tissue in terms of magnetic properties. Deoxygenated blood is paramagnetic which causes inhomogeneities in the magnetic field, and a decay in the MRI signal. Regions of the brain with higher activation (blood flow) are more magnetically uniform which increases the MRI signal. (B) Model-based fMRI. Activity associated with anticipated value extracted from an RL model in a decision-making task. vmPFC is an area recurrently associated with value tracking in the brain. Panel A from Gore et al. (2003). Panel B from Cortese et al. (2021).

What about aesthetic value?

After this brief overview, we can hint at many theoretical and methodological insights from the study of decisions that can be relevant to understanding how humans make aesthetic judgments. Indeed, aesthetic valuation is difficult to categorize, with many aspects of the decision-making process converging on it. For example, in the case of perceptual decisions, the neural basis of vision processing could inform some fundamental features used to construct aesthetic experiences. For value-based decisions, prior experience with a stimulus or object can influence value judgments, making previous memories and the history of associations relevant to estimate aesthetic preferences, e.g., I may love the song my mother used to play every day when I was a kid. On the other hand, value decisions are also flexible and highly adaptable, which in aesthetic judgments is reflected in how humans can express preferences for completely novel stimuli, somehow generating a “value” almost on the fly, e.g., I can know almost instantly if I like a painting, even if I have never seen it before. This flexibility could rely on the fact that humans may have a basal value mapping, an internal model of visual aesthetic preference, which can be used over novel objects.

Previous work in our research group has shed some light on how the brain can transform realistically complex stimuli into a simple subjective value (Iigaya et al., 2021). The brain can take high-dimensional input (for example, a complex painting) and reduce it to a one-dimensional scalar output (for example, a single rating: how much do I like this?). The dimensionality reduction process has been studied in decision-making literature using other kinds of visual inputs (e.g., Mach et al., 2020; Cortese et al., 2021), and it is crucial to identify how humans can generalize their knowledge to novel situations. Yet, little is known about how this process could operate in the context of aesthetic judgments. From the perspective of cognitive models, the diversity of visual art presents an interesting and challenging study case. Even in paintings alone, there is an overwhelmingly broad range of objects, themes, and styles. Therefore, studying artwork stimuli offers a good test for understanding the computation mechanisms the brain uses to generalize and generate values across diverse sets of stimuli.

It is important to note the field of neuroaesthetics has already used tools from cognitive neuroscience to study the aesthetic experience. For example, neuroimaging studies have shown how some brain areas increase their activity in the presence of stimuli with higher aesthetic values (Cela-Conde et al., 2004; Kawabata and Zeki, 2004). While this approach has been useful in assessing the relevance of some brain regions in this type of judgment, it is limited to probing the question of how the brain *computes* aesthetic value from visual stimuli in the first place. Using computational models (and model-based fMRI in particular) can bring some additional insights into potential algorithms and implementation.

A model of aesthetic value construction

The basis of the model is that the brain actively constructs the value of a stimulus by integrating its basic attributes or features. These features are basic enough to be found in any visual stimulus, i.e., they can used to generalize to new images. For instance, any work of visual art is composed of different colors, intensities, textures, and shapes and can be characterized as abstract or concrete, dynamic or still, and so on; all these are features that can be used to characterize a painting. To compute an overall value signal, the model considers that the brain implements an algorithm with three fundamental parts (Figure 4A):

1) *Feature weights of the agent*: The model considers that each agent can have a unique set of weights that characterize each one of the features. These weights reflect an individual's subjective judgment about how much a particular feature should count toward the overall value of a stimulus. These weights are assumed to be stable for each individual, allowing the model to define a *feature space*. Although the current version does not consider it, dynamic weight updates could be implemented in the model. This could capture scenarios of learning, e.g., some features could become relevant to our aesthetic valuation after art training.
2) *Feature metrics for novel images*: Each "new" image to be judged by the agent is decomposed into its basic features metrics. For example, the agent quantifies the novel painting's global luminosity, hue, symmetry, etc. These feature metrics will change for every painting to be judged by the agent; this is the *input* to the model.
3) *Integration of features*: For the novel image, the value of the features and their respective weights are used to compute the overall value of a stimulus, which can be used for further decisions, such as indicating an aesthetic value rating. This integration is assumed to follow a linear function (i.e., adding up the weighted metrics for all the features), although it could be implemented using non-linear functions. This stage generates the *output* of the model, the subjective value.

As you probably noted, the definition of the features is crucial for the effectiveness of the model. In the current implementation, two main sets of features are extracted: low-level features and high-level features. Low-level features can be generated by multiple methods of computer vision, which rely entirely on global characteristics of the image (e.g., color distributions, brightness effects, hue, saturation, etc.) or local features extracted from segmentations (e.g., the color of the largest segment in the painting). High-level features are generated based on more abstract dimensions, not easily computed by a simple algorithm such as concreteness, dynamics, or emotional valence value for each image (Chatterjee et al., 2010; Vaidya et al., 2018). For these high-level features, annotation by individuals with experience in art was requested. Analysis showed that low-level features can predict part of the variance of high-level features, hinting at a hierarchical organization in the feature generation.

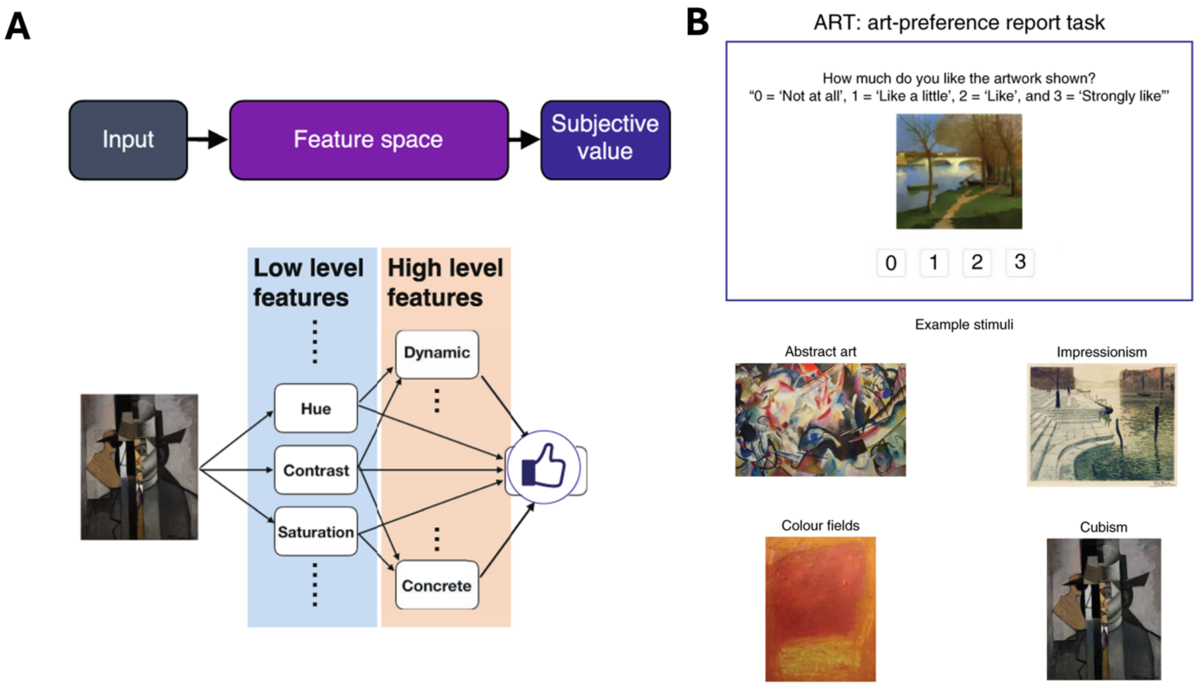

Figure 4. The model for prediction of aesthetic value. A) The idea of value construction captured by the model implies that input is projected into a feature space, in which the subjective value judgment is performed. Importantly, the feature space is shared across images, enabling this mechanism to generalize across various paintings and photos, including novel ones. In the model, a visual stimulus (e.g., artwork) is decomposed into various low-level features (e.g., mean hue, mean contrast) and high-level features (e.g., concreteness, dynamics). One hypothesis of the model is that high-level features can be constructed from low-level features. Subjective value is constructed from a linear combination of all features (low and high levels). (B) The task (ART: art-liking rating task). Participants were asked to report how much they liked a stimulus (a piece of artwork) shown on the screen using a four-point Likert rating ranging from 0 to 3. Example of some of the paintings presented to participants. Some of the styles presented to participants were Cubism, Impressionism, abstract art, and color fields. Reproduced from Iigaya et al. (2021).

To validate the model, participants with no expertise in art reported their aesthetic ratings for hundreds of images of different styles (e.g., color field, impressionism, cubism) during in-person and online experiments (Figure 4B). Naturally, each participant can have unique preferences regarding the type of artwork they prefer. The model can account for this variability since it is possible to find the best set of weights that allow the model to capture individual participant responses (Figure 5). The process of adjusting parameters to subject responses is called “Model Fitting” in computational modeling. To assess how good a model is (and eventually compare it with alternative models), we can check how well the fitted model can predict the value of images that the model has never seen. The aesthetic value construction model generates predictions that are significantly correlated with participants' responses, even when it is used to predict the aesthetic preferences for non-painting visual stimuli, such as photography. These observations suggest that a non-negligible proportion of the variance in participants' visual aesthetic ratings can be captured by how they assess simple features (although certainly there is room to improve the model). Some features were found to be consistently relevant across participants (e.g., concreteness), while in other cases, features presented varied relevance across participants (Figure 5).

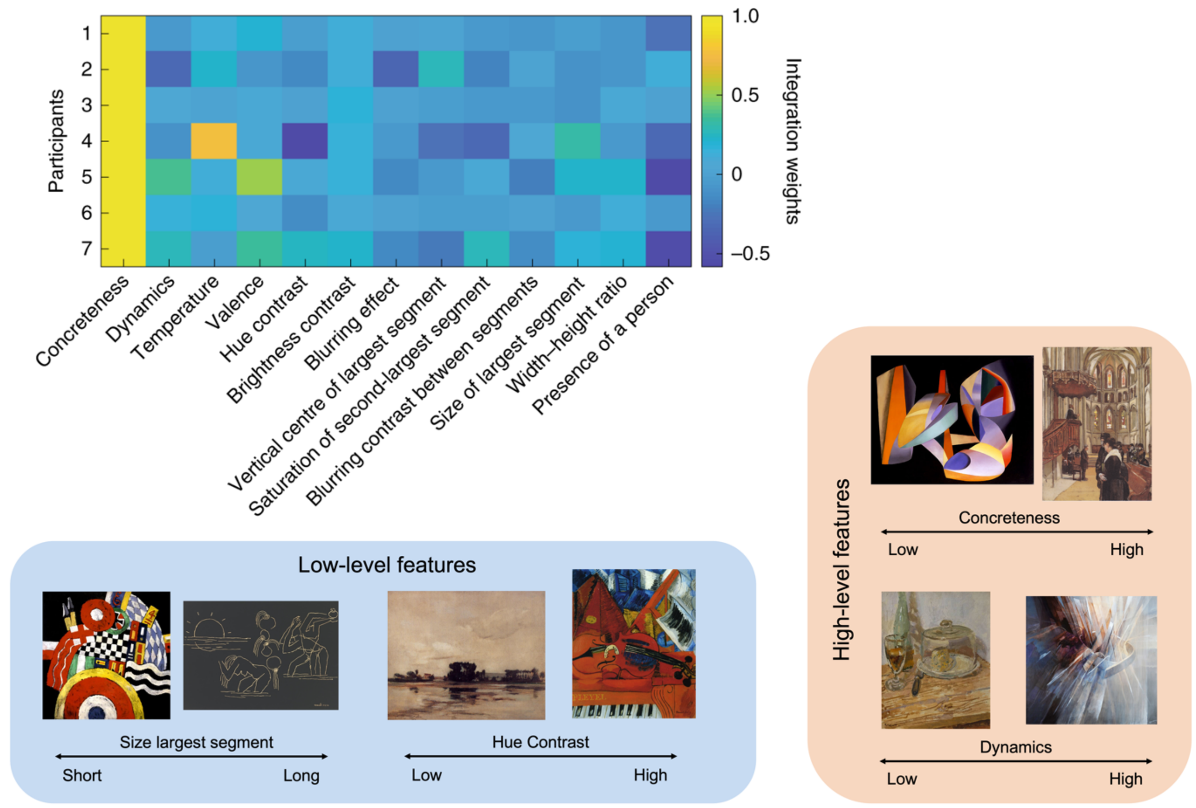

Figure 5. Value of feature weights for 7 in-person participants. Weights were estimated by fitting the model to individual participants separately. A reduced model, with some of the features most relevant features, is presented here. The colorbar in the figure indicates the value of feature weights for each participant. Positive weights indicate that a higher feature metric will increase the preference for the images (e.g., higher concreteness predicts a higher subjective value for this group of participants). Negative weights indicate that higher feature metrics will reduce the subjective value rating predicted by the model. Examples of low and high-level features are presented. Selected paintings are exemplars of images representing the feature metrics. Reproduced from Iigaya et al. (2021).

The model alone may suggest the algorithms used to make aesthetic judgments, but to elucidate the actual implementation of this process in humans, we need to include brain measures. Participants performed an equivalent artwork rating task inside the MR scanner, and a model-based fMRI analysis was used to characterize the potential neural substrates for the different parts of the model (Iigaya et al., 2023). In this analysis, it was found that areas in the visual stream (which encompasses part of the occipital and temporal cortex) tended to have a hierarchical representation of model features (Figure 6A). This means that primary visual areas (such as V1, which receives input from the retina through the thalamus) contain information mainly associated with low-level features. On the other hand, more anterior areas in the visual stream (e.g., the middle superior temporal area, MST) represented mostly the high-level features. These observations align with previous perceptual and value-based decision studies showing sensory features are strongly represented in these regions (e.g. Cortese et al., 2021; Castegnetti et al., 2021). Following the hierarchy, parietal and lateral frontal areas of the brain show mixed representations of the low- and high-level features. Anterior brain areas, especially the medial prefrontal cortex, were found to contain information about subjective aesthetic valuation (Figure 6B). This is also aligned with multiple reports of subjective value in decision experiments in cognitive neuroscience (e.g., Hare et al., 2011, Lim et al., 2013).

While these results show that feature and value information can be tracked in the brain, they do not indicate if these are independent representations or if they arise from some type of interaction. Functional connectivity analysis was used to test whether the brain areas representing features could be working together with those tracking subjective values during the task. Indeed, “feature” areas, such as the posterior parietal cortex (but not primary visual areas) and areas representing subjective values (medial prefrontal cortex), tended to be coupled during the task, suggesting that their interaction could facilitate the valuation process (Figure 6C), although this analysis does not allow to make assumptions about the directionality in the flow of information.

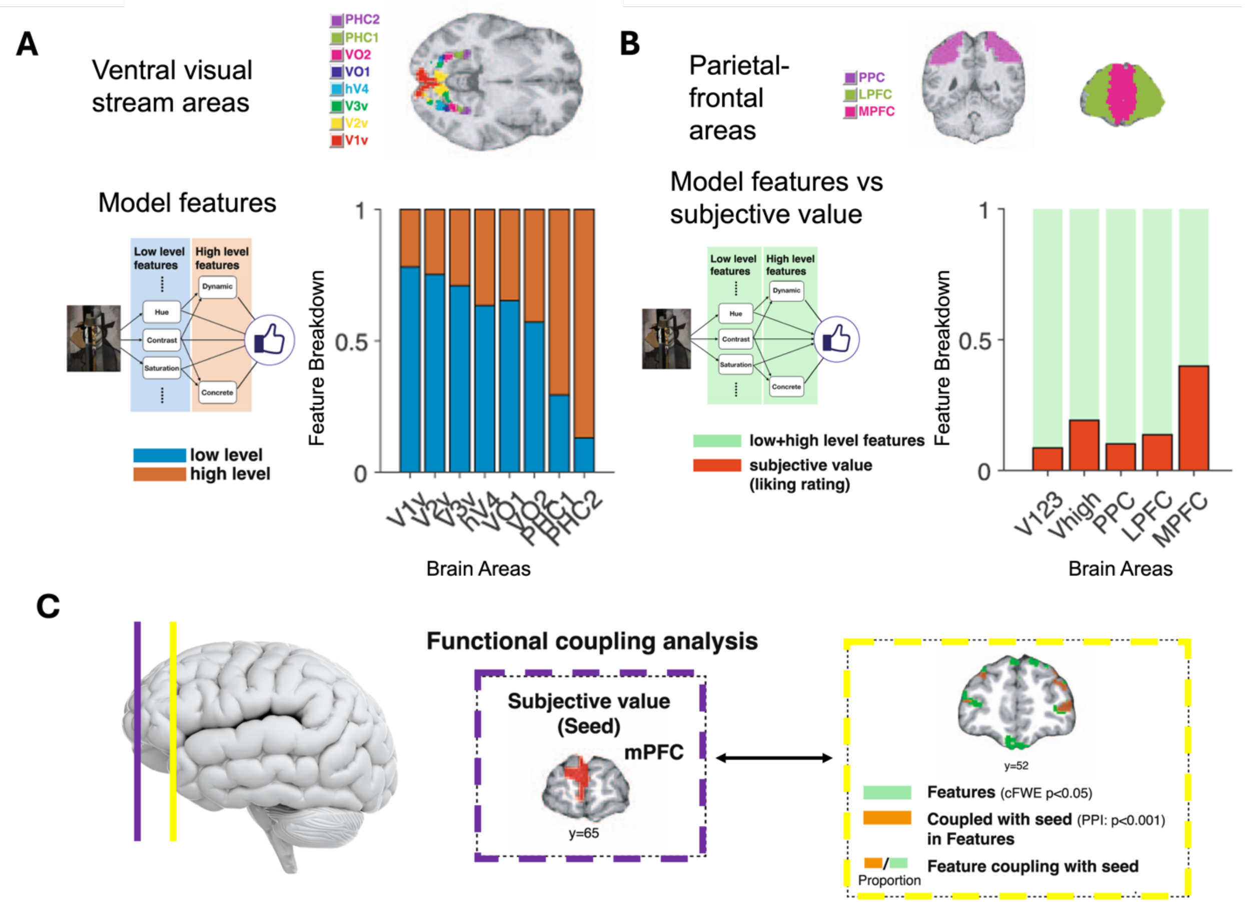

Figure 6. Encoding of low (blue) and high-level (orange) features in the visual ventral-temporal stream in a graded hierarchical manner. Deeper areas in the ventral stream show a preferential encoding for high-level features, as would be expected from the visual hierarchy processing. Brain parcellation with different colors for the areas in the visual stream is presented. B) Encoding of low- and high-level features (green) and liking ratings (red) across brain regions in frontal areas. Frontal areas Note that the ROIs for the visual areas are now grouped as V1-2-3 (V1, V2, and V3) and V-high (Visual areas higher than V3). Brain parcellation with different colors for the posterior parietal cortex (PPC), lateral prefrontal cortex (lPFC), and medial prefrontal cortex (mPFC) is presented. C) Functional coupling analysis to test how feature representations are coupled with subjective value. Regions that encode features are indicated in green. The orange color indicates regions that are coupled to the mPFC, an area that encodes subjective value (i.e., liking rating) during stimulus presentation. The brain on the left indicates the position of the coronal cuts for the brains presented in the center (subjective value) and right panels (feature representations). Encoding was obtained from one example participant. Reproduced from ligaya et al., (2023).

A complementary finding is that a Deep Convolutional Neural Network (DCNN), a computational model with a brain-like architecture optimized for visual object recognition, could also predict human subjective value after additional training. This type of network contains multiple levels of artificial neurons organized in sequential "layers." Additionally, these artificial neural networks do not use explicit and interpretable image features (e.g., image luminosity, hue, etc.) as input since their early layers are directly fed by the pixel information from the image. However, latent representations like visual features emerge spontaneously in the initial layers of the network after the optimization of the DCNN. Even more, after training the DCNN network to predict human aesthetic values for multiple paintings, it was found that deeper layers (later stages of information processing) also represented subjective preference variables, akin to the hierarchical processing observed across layers of the biological brains. This suggests that this hierarchical information processing, shaping areas from basic visual features to higher-level value assessments, could be a natural consequence of neuron-like architectures. This could hint at why some basic features seem to be recurrent for aesthetic valuations across people.

There are questions that this model does not answer. For example, while we can characterize the relative relevance of some features to generate aesthetic reports, how these weights are fixed for each subject is still a big question. Developing computational models to gain further insight into the generation of aesthetic values is a promising avenue to solve this and other important issues on aesthetic judgment. A recent proposal has taken direct insights from RL models to characterize how aesthetic value can evolve from an interaction between internal models and dynamic sensory experiences (Brielmann and Dayan, 2022). This type of model could help to characterize some observations that are central in other aesthetic frameworks, such as Processing Fluency Theories. These interesting implementations are beyond the reach of the current project but are worth keeping in mind since they could complement a feature model like the one presented here, allowing the integration of dynamic updates and learning.

Final remarks

The objective of this overview was to connect the study of decision-making from the perspective of cognitive neuroscience with the complex question of aesthetic evaluation and to introduce a model that captures some aspects of how these values could be constructed. During my time at the Italian Academy, I plan to expand this framework, using the presented computational model to assess the impact that goals and contexts could have during aesthetic evaluation. Value decisions are highly flexible, and behavior and brain are quickly adapted to novel task demands. This project will explore how aesthetic value computations, based on potential common mechanisms to other value and decision systems, could also be exposed to these manipulations. Furthermore, the process of creating each value judgment is highly dynamic; the process of aesthetic valuation is not automatic and instantaneous. I propose that including visual attention in the modeling could help understand how feature information can be sampled and prioritized in the time leading to aesthetic judgments.

References

Bandettini, P. A. (2012). Twenty years of functional MRI: The science and the stories. *NeuroImage*, *62*(2), 575–588. https://doi.org/10.1016/j.neuroimage.2012.04.026

Brielmann, A. A., & Dayan, P. (2022). A computational model of aesthetic value. *Psychological Review*, *129*(6), 1319–1337. https://doi.org/10.1037/rev0000337

Castegnetti, G., Zurita, M., & De Martino, B. (2021). How usefulness shapes neural representations during goal-directed behavior. *Science Advances*, *7*(15), eabd5363. https://doi.org/10.1126/sciadv.abd5363

Cela-Conde, C. J., Marty, G., Maestú, F., Ortiz, T., Munar, E., Fernández, A., Roca, M., Rosselló, J., & Quesney, F. (2004). Activation of the prefrontal cortex in the human visual aesthetic perception. *Proceedings of the National Academy of Sciences*, *101*(16), 6321–6325. https://doi.org/10.1073/pnas.0401427101

Chatterjee, A., Widick, P., Sternschein, R., Smith, W. B., & Bromberger, B. (2010). The Assessment of Art Attributes. *Empirical Studies of the Arts*, *28*(2), 207–222. https://doi.org/10.2190/EM.28.2.f

Cortese, A., Yamamoto, A., Hashemzadeh, M., Sepulveda, P., Kawato, M., & De Martino, B. (2021). Value signals guide abstraction during learning. *eLife*, *10*, e68943. https://doi.org/10.7554/eLife.68943

De Martino, B., & Cortese, A. (2023). Goals, usefulness and abstraction in value-based choice. *Trends in Cognitive Sciences*, *27*(1), 65–80. https://doi.org/10.1016/j.tics.2022.11.001

Frömer, R., Dean Wolf, C. K., & Shenhav, A. (2019). Goal congruency dominates reward value in accounting for behavioral and neural correlates of value-based decision-making. *Nature Communications*, *10*(1), 4926. https://doi.org/10.1038/s41467-019-12931-x

Gore, J. C. (2003). Principles and practice of functional MRI of the human brain. *Journal of Clinical Investigation*, *112*(1), 4–9. https://doi.org/10.1172/JCI200319010

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. 1969-2012). New York: Wiley.

Hare, T. A., Malmaud, J., & Rangel, A. (2011). Focusing Attention on the Health Aspects of Foods Changes Value Signals in vmPFC and Improves Dietary Choice. *Journal of Neuroscience*, *31*(30), 11077–11087. https://doi.org/10.1523/JNEUROSCI.6383-10.2011

Iigaya, K., Yi, S., Wahle, I. A., Tanwisuth, K., & O'Doherty, J. P. (2020). *Aesthetic preference for art emerges from a weighted integration over hierarchically structured visual features in the brain*. https://doi.org/10.1101/2020.02.09.940353

Iigaya, K., Yi, S., Wahle, I. A., Tanwisuth, K., & O'Doherty, J. P. (2021). Aesthetic preference for art can be predicted from a mixture of low- and high-level visual features. *Nature Human Behaviour*, *5*(6), 743–755. https://doi.org/10.1038/s41562-021-01124-6

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of sciences*, *1104*(1), 35-53.

Juechems, K., & Summerfield, C. (2019). Where Does Value Come From? *Trends in Cognitive Sciences*, *23*(10), 836–850. https://doi.org/10.1016/j.tics.2019.07.012

Kawabata, H., & Zeki, S. (2004). Neural Correlates of Beauty. *Journal of Neurophysiology*, *91*(4), 1699–1705. https://doi.org/10.1152/jn.00696.2003

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, *93*(2), 451–463. https://doi.org/10.1016/j.neuron.2016.12.040

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. https://doi.org/10.1016/j.conb.2012.06.001

Lim, S.-L., O'Doherty, J. P., & Rangel, A. (2013). Stimulus Value Signals in Ventromedial PFC Reflect the Integration of Attribute Value Signals Computed in Fusiform Gyrus and Posterior Superior Temporal Gyrus. *The Journal of Neuroscience*, *33*(20), 8729–8741. https://doi.org/10.1523/JNEUROSCI.4809-12.2013

Mack, M. L., Preston, A. R., & Love, B. C. (2020). Ventromedial prefrontal cortex compression during concept learning. *Nature Communications*, *11*(1), 46. https://doi.org/10.1038/s41467-019-13930-8

Marr, D. (2010). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. The MIT Press. https://doi.org/10.7551/mitpress/9780262514620.001.0001

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *341*(6237), 52–54. https://doi.org/10.1038/341052a0

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., & Wilson, R. C. (2015). Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. *The Journal of Neuroscience*, *35*(21), 8145–8157. https://doi.org/10.1523/JNEUROSCI.2978-14.2015

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-Based fMRI and Its Application to Reward Learning and Decision Making. *Annals of the New York Academy of Sciences*, *1104*(1), 35–53. https://doi.org/10.1196/annals.1390.022

Pagnoni, G. Z., Caroline F. ;. Montague, P. Read; Berns, Gregory S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, *5*(2), 97–98. https://doi.org/10.1038/nn802

Pessiglione, M. S., Ben; Flandin, Guillaume; Dolan, Raymond J. ;. Frith, Chris D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. https://doi.org/10.1038/nature05051

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556. https://doi.org/10.1038/nrn2357

Schultz, W. (1998). Predictive Reward Signal of Dopamine Neurons. *Journal of Neurophysiology*, *80*(1), 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Sepulveda, P., Usher, M., Davies, N., Benson, A. A., Ortoleva, P., & De Martino, B. (2020). Visual attention modulates the integration of goal-relevant evidence and not value. *eLife*, *9*, e60705. https://doi.org/10.7554/eLife.60705

Seymour, B. O., John P. ;. Dayan, Peter; Koltzenburg, Martin; Jones, Anthony K. P. ;. Dolan, Raymond J. ;. Friston, Karl J. ;. Frackowiak, Richard S. J. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*(6992), 664–667. https://doi.org/10.1038/nature02581

Shadlen, M. N., & Kiani, R. (2013). Decision Making as a Window on Cognition. *Neuron*, *80*(3), 791–806. https://doi.org/10.1016/j.neuron.2013.10.047

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (Eds.). *Statistical Parametric Mapping*. (2007). Elsevier. https://doi.org/10.1016/B978-0-12-372560-8.X5000-1

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. https://doi.org/10.1038/s41593-017-0008-x

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. https://doi.org/10.1093/cercor/bhx246