Enhancing models of social and strategic decision making with process tracing and neural data

Konovalov, Arkady; Ruff, Christian C.

DOI:
10.1002/wcs.1559

License:
Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):
Konovalov, A & Ruff, CC 2022, 'Enhancing models of social and strategic decision making with process tracing and neural data', *Wiley Interdisciplinary Reviews: Cognitive Science*, vol. 13, no. 1, e1559. https://doi.org/10.1002/wcs.1559

Enhancing Models of Social and Strategic Decision Making with Process Tracing and Neural Data

Arkady Konovalov[1,2] and Christian C. Ruff[1]

[1] Zurich Center for Neuroeconomics (ZNE), Department of Economics, University of Zurich, Zurich, Switzerland
[2] Correspondence: arkady.konovalov@uzh.ch

Abstract

Every decision we take is accompanied by a characteristic pattern of response delay, gaze position, pupil dilation, and neural activity. Nevertheless, many models of social decision making neglect the corresponding process tracing data and focus exclusively on the final choice outcome. Here we argue that this is a mistake, as the use of process data can help to build better models of human behavior, create better experiments, and improve policy interventions. Specifically, such data allow us to unlock the "black box" of the decision process and evaluate the mechanisms underlying our social choices. Using these data, we can directly validate latent model variables, arbitrate between competing personal motives, and capture information processing strategies. These benefits are especially valuable in social science, where models must predict multi-faceted decisions that are taken in varying contexts and are based on many different types of information.

1. INTRODUCTION

Social behavior is multi-faceted and understanding it is a truly daunting task. Why do we trust other people? Why do we share food and shelter with one another? How do we cooperate – and why are we competitive? How do we understand each other? For centuries, we have been building theories of social behavior, each based on a diverse set of assumptions and views about human nature (Abrams & Hogg, 2006; Homans, 1974; Katz & Kahn, 1978; Weick, 1977). These theories have led to proposals of many potential motives underlying social behavior – envy, aggression, power, altruism, affiliation, approval, and many others (MacCrimmon & Messick, 1976) – but we still do not know what drives these motives, how they operate, and how we can predict the exact motive depending on the type of social situation.

While a verbal, more qualitative approach to social cognition does have its appeal, in recent years a model-based, quantitative framework is gaining traction in social science (Hackel & Amodio, 2018; Konovalov et al., 2018; Lockwood et al., 2020; Zhang et al., 2020). Stemming from economics and mathematical psychology, this framework proposes that we can build consistent and predictive theories of behavior based on simple choice axioms and mathematical – or computational – foundations. One approach is to assume that individuals maximize the utility of their actions and this utility depends, in a numerically precise way, on the context, individual traits, and other parameters of the environment (Camerer & Fehr, 2006; Fehr & Camerer, 2007). Based on these assumptions, we can build mathematical/computational models that combine these parameters and thus allow us to make specific predictions about individuals' choices in any decision situation for which the model inputs can be quantified. This approach therefore transcends the purely descriptive or explanatory goals of verbal choice theories, by enabling us to predict behaviour in novel situations and formally compare these predictions across different contexts.

While predicting choices themselves is an important goal that is widely practiced in various social sciences, it is becoming clear that the models we need to develop for this purpose may be too complex to be fit exclusively based on choice outcomes (Schulte-Mecklenbeck et al., 2017). Choice outcomes are usually binary or purely multinomial; they therefore often do not reflect the complexity of the underlying decision process. While this limitation may in principle be mitigated by careful experimental design, this strategy would often have to result in complicated task setups that are very different from typical daily-life decisions. Fortunately, each decision – including social ones – not only produces the final outcome but is also accompanied by behavioral expressions in many other channels that contain rich information on the processes leading up to the choice (Cooper et al., 2019). These channels include, but are not limited to response times, mouse-tracking trajectories, gaze position, pupil dilation, facial expressions, skin conductance, neural activity, and many other variables that can be recorded both in the real world and in the research laboratory (Schulte-Mecklenbeck et al., 2019).

In this article, we argue that measures of these processes, often referred to as process data, are very useful for enhancing our understanding of human behaviour. Specifically, such measures are indispensable for attempts to pin down the neurocognitive mechanisms that contribute to our social decisions, such as preferences, learning, and choice processes. In the following sections, we will describe and discuss these attempts. We will first introduce the concept of process data (section 2), before reviewing their use in the literature (section 3) and identifying opportunities for future research in the field (section 4).

2. WHAT ARE PROCESS DATA AND WHY SHOULD WE USE THEM?

Any decision we take is characterized not just by what we choose, but also by how this choice unfolds over time. Think of a simple example of social choice: An individual uses an internet browser to donate an amount of money to the Red Cross. When presented with a menu of possible donations ($5, $10, $20, $50, $100, custom), the individual chooses $20. A researcher studying this decision can - provided the necessary consent and compliance with the legal framework - record not just the final chosen monetary amount but many other types of data. These provide considerable information about the choosing individual and the respective choice. We will refer to these measures as process data from now on. In the following, we will review the major types of process data used in computational social research, noting that other types of data could in principle also be collected (Schulte-Mecklenbeck et al., 2019).

2.1. Types of process data

Response time. One readily available measure is response delay or decision time, usually defined as the time between the presentation of the decision problem (e.g., opening the charity website) and the moment of the decision itself (e.g., clicking on the payment button). In the literature, decision time can also be referred to as “response time” (RT), or “reaction time”. While there is no consensus on the correct usage, and many use these terms interchangeably, it is common to use “reaction time” when the individual is reacting with a stereotypical response (such as a key press) to the presence of a stimulus (for instance, a sound) and “response time” when the individual is making a choice between 2 or more alternatives (Kosinski, 2008). Response times can be recorded easily in many online interactions and can be quantified precisely (up to several milliseconds) in laboratory experiments, often revealing the strength of internal conflict between competing motives or preferences (Clithero, 2018; Konovalov & Krajbich, 2019; Spiliopoulos & Ortmann, 2017).

Mouse tracking. Another relatively easily accessible process measure is mouse tracking (Costa-Gomes et al., 2001; Johnson et al., 2002; Stillman et al., 2018). While the precise,

continuous position of the mouse cursor might not be available outside of a research lab, websites can be designed in a way to record clicks on various elements, amount of scrolling, and time spend outside of the browser window. All these details can provide insight into information processing and choice strategy employed by the individual. In our charity example, this might include photos that the decision maker chose to zoom in, or details of the specific charity programs she accessed. In the laboratory, mouse movements can be recorded quite precisely, including specific trajectories of the cursor at any moment in time (with a typical time resolution of 60 Hz, usually restricted by the type of the mouse used).

Eye tracking. While mouse tracking is easy, it is not always useful as a proxy measure of information processing. If we observe that an individual moved her mouse over a paragraph, it does not mean that she was actively reading the text. However, if we can record that the subject was actively *looking* at the paragraph, we can confirm that she did indeed process the information in the text. Using an *eye-tracker*, one can record two valuable types of data: gaze position (manifested in X and Y coordinates, which also could be converted to higher-order data such *fixations* and *saccades* (Hessels et al., 2018)) and size of the individual's pupil at any moment in time (given a specified sampling rate, typically up to 2000 Hz) (Duchowski, 2007). High-quality laboratory eye-trackers typically use infra-red light and sensitive cameras to compute the position of the corneal reflection relative to the pupil, thus estimating the gaze direction. While the subject clearly understands that her gaze is being recorded, typically this procedure is non-invasive and does not affect behavior in any meaningful way (assuming that the experiment is carefully designed in such a way that the subject *must* look at different parts of the screen to take in the corresponding information).

Neural data. Finally, the researcher might choose to record brain activity of the subject. While there are many different types of possible neural recordings (Sejnowski et al., 2014), here we will focus on the two types most commonly used in social neuroscience: electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). EEG is a relatively inexpensive method of recording the brain's electrical activity directly from the scalp, using a set of electrodes placed in a cap (Luck, 2014). EEG has an excellent temporal resolution, recording the voltage at the rate of up to 2000 Hz (each half a millisecond), but does not allow to identify the precise source of the electrical signal within the brain. fMRI, on the other hand, is much slower (typically allowing for one measurement per 2-3 seconds) and much more expensive (in terms of both the construction and maintenance cost). However, it builds a dynamic 3D image of neural activity in the brain (estimated using the blood oxygen level), showing which regions of the brain were more active at any given time during an experiment (Huettel et al., 2004).

What does neural data allow us to investigate? First, typically we use it to understand whether choices that are theoretically distinct are indeed processed differentially by the brain. This can be evident in neural activity in a specific region (if using fMRI) or during a specific period (if using EEG), both in terms of average activity levels or *patterns* of activity across space or time. For instance, returning to the charity example, one could study whether choices involving positive (condition 1) or negative (condition 2) emotional photos evoke higher responses in the subject's temporoparietal junction (or TPJ, the brain region usually implicated in empathy and theory of mind (Kanske et al., 2015; Saxe & Kanwisher, 2003)), and whether higher activity predicts the subject's choice of the donation size. Second, we can also take a continuous variable (for instance, the size of the donation) and test whether neural activity correlates with this variable over the course of the experiment, and whether the strength of this response predicts the choice outcome (Tusche et al., 2016). This would indicate that the brain may use a representation of this variable for choice. There are numerous opportunities here. For instance, one could adjust the emotional content of the message above the donation button, map it onto a numerical scale, and test whether this index

correlates with brain responses. Third, we can use brain data to test precisely formulated quantitative models of the processes underlying choices. This approach is especially powerful and will be described in the next section. That is, one can construct a latent variable as a part of a model of the subject's decisions and try to seek whether this variable, which is assumed to be a (mechanistic) part of the choice process, is represented in the brain. We will now describe this approach in greater detail.

2.2. Using multiple types of process data to test theories about social behavior

Let us illustrate in a concrete case how process data can be useful to reveal decision mechanisms underlying social behavior. Suppose the researcher's goal is to understand which factors increase charitable donations. It is known that affective images tend to impact the probability of donations and donations size (Small & Verrochi, 2009). These analyses are typically straightforward: the researcher manipulates the stimuli (such as pictures, facial expressions, or valence of the depicted emotions) and measures the response of the subject.

While these results are certainly useful and can be immediately used to make policy recommendations, they do not answer why exactly these factors impact on the outcome and on how subjects make their decisions. Understanding this would be important for predicting subjects' behavior across different situations and for designing future experiments.

Let us now assume that the researcher has access to several types of process data described above such eye-tracking or fMRI data. It is still uncommon to see both types of data to be used in the same study, so we will illustrate the use of these two types separately. We know from behavioral experiments that affective images drive donation decisions, but how exactly do they impact on the decision, and how can we measure the emotional component of the choice process?

One way to do this is via eye tracking. One paper (Bebko et al., 2014) demonstrated that emotional valence of a photo directly captures the attention of the decision maker, impacting on both gaze durations and saccade times. These measures, in turn, are shown to be correlated with (in this case, hypothetical) donation decisions. These findings suggest that there is a specific choice mechanism that links attention to emotional stimuli and donation decisions. When coupled with a computational model of attention, this putative mechanism could provide the researcher with specific predictions for donation sizes across different stimulus sets.

Another way to measure emotional impact of a charity stimulus is by using neural data. One such study (Genevsky & Knutson, 2015) showed that photos that elicit positive arousal typically increase response in the nucleus accumbens. Using an out-of-sample prediction, the study showed that activity in this area (measured using fMRI when subjects passively observed a number of charity photos) predicted real-life donation amounts over and above other factors measured from behavior (such as self-reported willingness to donate). These results again confirm an important link between emotion and altruistic choice, and demonstrate that neural measures can help researchers predict choices better than with just verbal-willingness-reports alone.

One important tool that may bring these interesting approaches together would be the development of a mechanistic model that describes the underlying decision mechanisms affected by attention and affect. Use of such a model, coupled with different types of process data, would allow the researcher to make more specific predictions; for instance, by simulating how different experimental conditions should impact on the process data – and testing these predictions in new experiments (for instance, with different control photos). In the next section, we will discuss in more detail how process data can be used to build and validate such models of social behavior.

3. HOW CAN PROCESS DATA HELP US VALIDATE COMPUTATIONAL MODELS OF SOCIAL BEHAVIOR?

One of the goals of social science is to explain and predict social behavior. While purely verbal theories have been dominating the field of psychology for a long time, it is becoming increasingly more common to use mathematical and computational models to explain social choices. There are many benefits in doing so (Hackel & Amodio, 2018; Konovalov et al., 2018; Lockwood et al., 2020). First, computational models make precise quantitative predictions that can be tested out-of-sample; importantly, these predictions can be made for a set of decision problems or subjects that the model was not built on. Second, models can arbitrate between competing theories of underlying motives, cognitions, and decision mechanisms by explicitly stating (mathematically) the relationships between such mechanisms. Third, models can demonstrate commonalities between seemingly different types of behavior. Finally, they can link specific mechanisms to distinct brain regions and computations, shedding light on the neural underpinnings of social decisions.

There are three general classes of decision models used in social science, which cover three important aspects of social behavior: preference, choice, and learning (see Figure 1). In this section, we will briefly outline the three classes and demonstrate how process data can help build and test these models.

3.1. Models of social behavior: preference, choice, learning

Social *preference models* typically quantify how individuals value social allocations or outcomes, e.g. by producing a numerical value of an outcome based on its characteristics (how much money or goods each person is receiving, how fair or moral the outcome is, etc.). These values can be then used to predict which option an individual would prefer given a choice between a set of alternatives. While many economic theories are agnostic as to how preferences are formed, process data might help us understand whether preferences are an innate biological trait, or whether they depend dynamically on processes related to value construction (for instance, via gaze direction) or other decision mechanisms.

A rational decision maker is often assumed to have stable preferences (typically quantified by her preference model parameters) and given a choice between two options multiple times, must always choose the same option. However, real people do not always display consistent preference and often have a degree of randomness in their choices. This phenomenon led to the development of *choice models*, which explain how exactly the choice is made given the values of the options. The most common example is a simple logistic random utility model that assumes that values simply have an additive random component (McFadden, 2005). More complex models, which originated in cognitive psychology, such as the drift-diffusion model, do not just predict choice outcomes, but also explain how process data such as decision times or neural activity can be generated by the value comparison process (Fudenberg et al., 2018; Ratcliff & McKoon, 2008). Note that most of these models do not necessarily have a uniquely social component, but they have been increasingly used in studies of social behavior to explain motives underlying social decisions.

Finally, preferences and values are not necessarily stable over time and might be affected by rewards or changes in beliefs (Behrens et al., 2008, 2009). *Learning models* aim to capture this process, and value updating (e.g. through prediction errors) have been successfully used in behavioral psychology and computer science as a simple, but powerful mechanism underlying learning.

One important concept in social learning is beliefs, or representations of the individual's estimates of probabilities of certain events (such as another person actions). These include

first order beliefs ("which side of my court the other tennis player will serve to next?") and second-order beliefs ("which side of her court the other tennis player expects me to serve to next?"). By induction, higher order beliefs can also be constructed. The second- and higher-order beliefs represent a quantitative representation of "theory of mind", or mental representation of others' thoughts and beliefs. Computational models can allow us to precisely estimate these variables, and process data can help to validate their representation in the brain.

In Bayesian learning frameworks, the decision maker is assumed to perform optimal Bayesian updating of the variables in the world around them (Devaine et al., 2014). In the social domain, the most common example are actions of another person, social beliefs, or social norms (Xiang et al., 2013). These models can provide surprising insights. For example, they have been used to suggest that herding behavior, often cited as an example of suboptimal social influence, can in fact originate from perfectly rational decision making if the subjects simply use optimal Bayesian updating and learn from choices of others (Bikhchandani et al., 1992).

The reinforcement learning framework is based on the idea that the brain stores and updates values of actions (and states of the world) through so-called *prediction errors*, which are computed as a difference between the actual received reward (or the obtained state) and the expected reward (or probabilistic belief of reaching the state) (Sutton & Barto, 1998). If the prediction error is positive, the value is updated upwards, if it is negative, downwards. More sophisticated models also allow the decision maker to compute forward-looking policy using a Bellman value function (Doll et al., 2012).

Originally applied in economics to strategic behavior (Erev & Roth, 1998; Roth & Erev, 1995), learning models have been used in many domains to explain a wide variety of social choices, from conformity (Huber et al., 2015) to observational learning (Collette et al., 2017).

For all three classes of models, process data can be used as an important validation tool. While latent variables and parameters do affect decision, they are only indirectly manifested in observed choices. Process data, on the other hand, might allow us to directly measure or estimate underlying processes, thus providing supporting evidence for the model in question.

3.2. Differentiating fixed versus changing model mechanisms

Process data in combination with computational models can allow us to differentiate aspects of the decision process that are fixed and specific to a given decision-maker or context, versus those that change over time within these individual and situational contexts. That is, all three classes of models described above share some characteristics: they include *parameters,* which are fixed (such as the inequity aversion in preferences, drift rate in the drift-diffusion model, or learning rate), and latent *variables* that change over the course of the experiment or even with the single decision (with values of actions being a common example) (Figure 1).

For instance, we can estimate a weight that individuals assign to others' outcomes, and this weight might depend on whether the decision is made in the advantageous (when the individual has more money that the other person) or disadvantageous context (when the individual has less money) (Fehr & Schmidt, 1999). Process data can help in many ways here: to identify the parameter when the choice outcomes alone are inconclusive, to provide an alternative estimate of the parameter (e.g. from brain activity (Morishima et al., 2012)), or to demonstrate that individual variability in the parameter might be related to some characteristics of the choice process (e.g. information processing in gaze data (Jiang et al., 2016)).

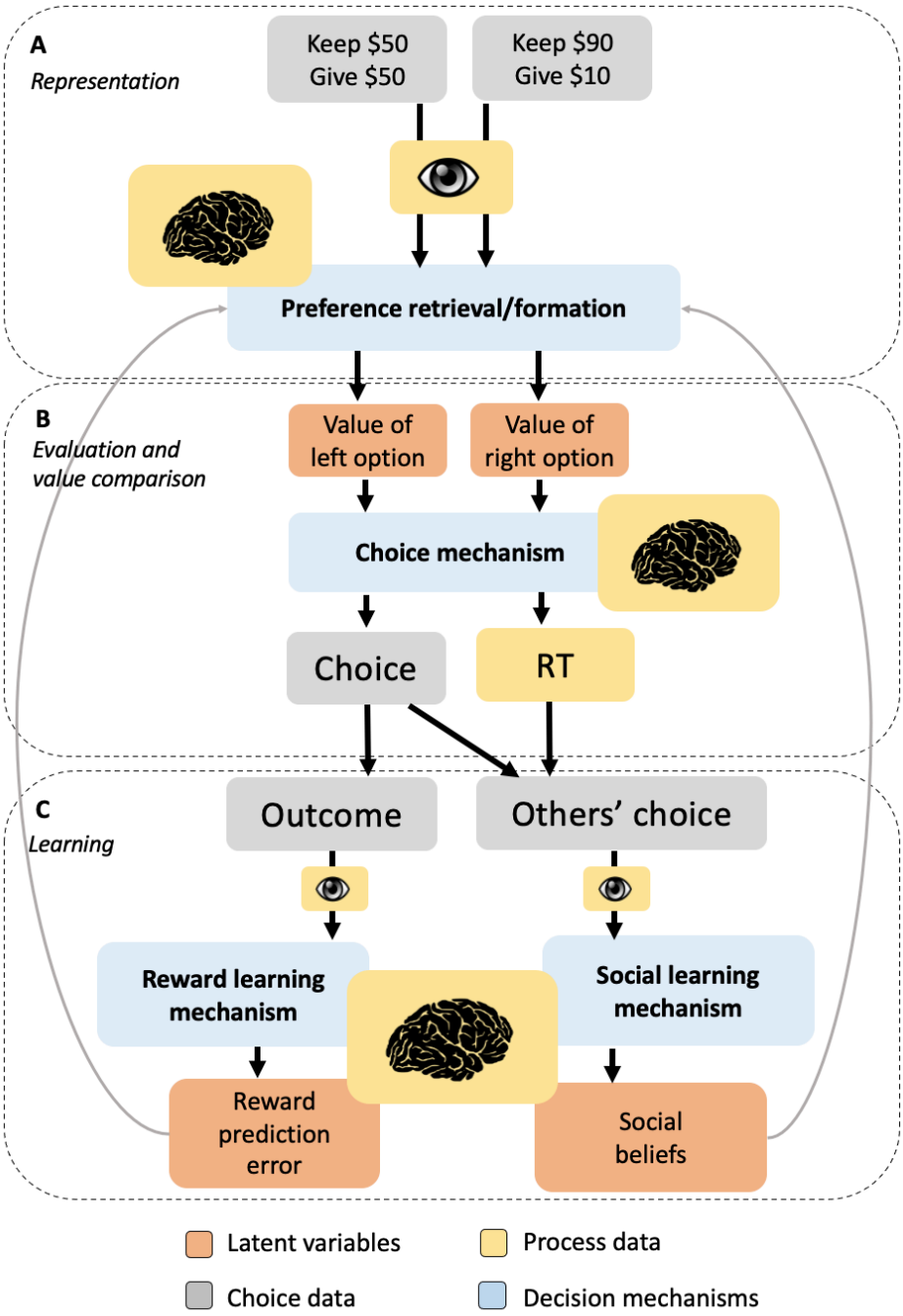

Figure 1. Process data in social decisions. An illustration of how process data can index different stages of a choice between two options: Giving another person either $50 or $10 in a trust game (left and right boxes at the top of the figure). Note that the hierarchical and sequential nature of the figure merely serves illustrative purposes; the decision process is not necessarily top-down and these stages might occur in parallel. (A) *Representation stage*: facing this decision, the individual must process the payoff information (producing gaze data). The payoff information is then mapped onto preferences (as formalized in a *preference retrieval/formation mechanism)* to form values of the options. (B) *Evaluation and value comparison stage*: the options are evaluated and compared using a *choice mechanism*, producing choice outcomes and response times (RT). (C) *Learning stage*: after another person's choice is revealed (and processed visually, producing gaze data), the outcome is then compared to the expected reward, producing a prediction error if the obtained reward is different from the expected reward (via a *reward learning mechanism*) and/or an update of the individual's beliefs if the other person reacts to the outcome (via a *social learning model*). Learning mechanisms, including errors and belief updates, can be instantiated in neural activity and might affect values of future

choices through the preference formation mechanism (gray arrows). These model-assumed processing steps can be tested with at least three types of process data, as visualized by the yellow boxes.

Latent variables represent intermediate or final quantities that the individual computes during the decision process. The most common example of a latent variable often employed in decision science is value (or utility) of actions (see Figure 1 for an example). Economic theory usually assigns utilities to actions in "as-if" fashion: an individual is assumed to make decisions in a consistent manner, as if he or she is assigning a numerical value to all possible alternatives (displaying complete and transitive preferences). While it is not necessary that the brain should compute, store, and update values, substantial evidence from neuroeconomics suggests that such cardinal values might be indeed computed in the brain (Bartra et al., 2013), even during social decisions (Ruff & Fehr, 2014). If this is indeed the case, neural data can allow us to validate models of social decision-making by searching for correlates of these latent variables in the neural signal.

While some of these variables (such as value) can be directly related to choice, others are only auxiliary and represent certain steps in the decision or mechanisms underlying that decision. For instance, value might not be instantaneously computed; the value construction process might include evaluation of the attributes (such as monetary amounts divided between the individual and another person, weighted according to some preference mechanism). These latent quantities might be significantly different from explicitly presented numbers that the individual observes. Another example is learning: updating of values of actions might reflect incorporating an update of the individual's beliefs, which can again be quantified according to some learning rule (for instance, in terms of prediction errors).

Process data can be invaluable in validating the existence of these intermediate steps. For example, neural data can provide evidence for model-predicted representation of values and beliefs in the brain (in the form of prediction errors, as discussed above)(Behrens et al., 2009), response times can validate the processes formalized in the evidence accumulation framework by demonstrating that individual preferences indeed affect the subject's RT (Konovalov & Krajbich, 2019), and eye-tracking can demonstrate how formation of preferences depends on gaze patterns (Jiang et al., 2016). Thus, process data can validate the assumptions of theoretical models by demonstrating the existence of certain intermediate processing steps formulated by the model.

3.3. Quantifying social preferences

Process data have been used to investigate and validate theories of social preferences. Social preference is typically understood as individual preference over various social allocations (Fehr & Camerer, 2007; Ruff & Fehr, 2014). For instance, a person who has extra money might donate them to those in need, revealing that they prefer a more equal distribution of monetary outcomes in society. In the laboratory, social preferences are often tested using the so-called "dictator games", where subjects allocate amounts of money between themselves and other people (Engel, 2011). From these experiments we know that people often exhibit preferences for more equal allocations (sharing, on average, about 30% of their endowment (Engel, 2011)). Why do individuals value others' outcomes?

Many possible explanations have been offered by economics and psychology over the years. One popular theory is "inequity aversion": people dislike unequal outcomes, and the strength of this distaste varies across individuals. Mathematically, it can be captured with a utility function that reflects value that individuals might be assigning to different allocations or outcomes (Fehr & Schmidt, 1999; Charness & Rabin, 2002). These models typically assume that individuals assign (stable) weights to their own payoffs or payoffs of others and make decisions according to these computations (or, more precisely speaking in economic

theory terms, make consistent decisions *as if* their preferences were represented by a utility function that takes the characteristics of choice options as the input and produces a numerical value as the output (Mas-Colell et al., 1995)).

Using individual choices, we can easily estimate these weights. At the same time, without process data we might not have enough evidence to understand how these weights are constructed in different contexts and what determines them on the individual level. One possible way to improve on this is to use gaze data, which can reveal the process of this weight construction via information acquisition. We directly evaluate which payoffs subjects attend to, how long they look at those payoffs, and whether visual attention might influence the weights subjects assign to own and others' payoffs and how this then impacts on their choices (Barrafrem & Hausfeld, 2020; Fosgaard et al., 2020; Jiang et al., 2016; Smith & Krajbich, 2018).

Note that the interpretation of the model weights would differ drastically depending on whether they are influenced by what participants attend to versus if they were fixed: In the first case, it would mean that social preferences are strongly malleable and can be influenced by what we look at and attend to, versus in the second, it would mean that they represent individual traits that are invariant across different contexts. However, in these studies causality remains an issue: while they show that certain gaze patterns correlate with behavior (for instance, selfish subjects tend not to look at the others' payoffs), it does not imply that preferences are defined by gaze patterns (intuitively, selfish people willingly might not attend to certain payoffs). By designing experiments where gaze is exogenously manipulated by the researcher, one could potentially demonstrate that social preferences might be constructed at the moment of choice and thus can be manipulated by certain nudges, displays, or attention distractors.

On the other hand, some evidence suggests more intrinsic, biological underpinnings. Social preferences can also be determined by brain structure and brain activity patterns (Emonds et al., 2012; Morishima et al., 2012; Tricomi et al., 2010). Combining several types of process data (such as eye-tracking and fMRI) could potentially help us to disentangle innate versus constructed aspects of social preference. Going back to the charity experiments discussed in Section 2.2, using the two measures simultaneously could provide more substantial evidence for the impact of emotional stimuli on charitable decisions. Specifically, linking gaze data and fMRI signals could enable the researcher to directly estimate the affective impact of various parts of the stimuli (such as faces or textual cues) on choices.

3.4. Arbitrating between social choice mechanisms using response time

Preferences are typically assumed to be stable (at least in the short term, within a single experiment), but choices are not. For many decades of research in individual choices we know that people often switch their choices even within the same block of trials (Mosteller & Nogee, 1951). In most cases, choices are more stable when the decision is easy (one alternative is clearly preferred over the other) and close to randomness if the individual is indifferent (the values of two alternatives are equal). This effect is often captured by a "psychometric curve", which can be easily explained if one assumes that the choices (or preferences) have a stochastic component (for instance, an error term with a logistic distribution (McFadden, 2005)).

However, in some cases choices per se might not be enough. Imagine a situation where a subject in a laboratory experiment, presented with a series of binary allocation between themselves and another anonymous subject, always chooses selfish allocations. If all choices are the same, we cannot reliably estimate what the weights the subject might be assigning to others' payoffs (assuming that an allocation where they would choose an altruistic outcome exists, but was not presented during the experiment).

Process data that can help the experimenter here are response times (RT). Experimental evidence clearly demonstrates that people are slower making difficult choices (with utilities of the options close to each other) and faster for easy choices (the difference in utilities is large). This phenomenon is predicted by a class of choice-process models called sequential sampling models, which include the drift-diffusion models (Ratcliff & McKoon, 2008), linear ballistic accumulators (Brown & Heathcote, 2008), and others. These models, originating in psychology, have been shown to explain choices and response times in many economic domains, including social allocations preferences (Hutcherson et al., 2015; Johnson et al., 2017; Krajbich et al., 2015; Teoh et al., 2020). The simplest version of a sequential sampling model assumes that each choice is a continuous process, during which a decision variable (often referred to as "*evidence*") accumulates over time until it hits one of two *bounds* (upper and lower), which represent the alternatives being presented (Ratcliff & McKoon, 2008). The decision variable changes over time with a *drift rate*, which is often assumed to be a function of choice difficulty, represented as a difference in utilities between the options. The drift rate also has a noise component, which leads to some degree randomness in decisions; however, in general the model predicts that both choices and RT depend on the drift rate (and thus on the difficulty of choice, or decision conflict).

Many early studies in psychology and economics have observed that RTs reflect choice conflict (Dashiell, 1937; Diederich, 2003; Jamieson & Petrusic, 1977; Mosteller & Nogee, 1951; Tversky & Shafir, 1992). Computational choice models such as the drift-diffusion model can offer an explanation: it is rational for the decision maker who values her time to delay the decision until a certain threshold of confidence is reached, and this delay is longer if the values of the options are closer to each other (Alós-Ferrer et al., 2016; Busemeyer, 1985; Busemeyer & Rapoport, 1988; Busemeyer & Townsend, 1993; Echenique & Saito, 2017; Fudenberg et al., 2018; Harris et al., 2018; Hutcherson et al., 2015; Krajbich et al., 2010; Krajbich & Rangel, 2011; Moffatt, 2005; Rodriguez et al., 2014). This implies that if we present individuals with a series of decision problems, we could potentially identify their social preferences from their RTs.

Returning to our example, if the experimenter also recorded RT data, she could also estimate the amount of decision conflict on any given trial. Then a subject who always chose selfishly, but took longer time to decide between a \$90/\$10 and a \$50/\$50 allocations, could be shown to be more prosocial then another selfish subject who was very fast to pick the selfish option (Konovalov & Krajbich, 2019). This observation can thus allow us to differentiate between competing theories of social behavior.

One such open question is whether cooperation is an innate, intuitive human trait, or whether we are inherently competitive. Over the last two decades, a few studies in economics and social science presented RT as a method to differentiate fast intuitive versus slow deliberative decisions. For instance, people who choose a better strategy (e.g. closer to the equilibrium play in the game-theoretical sense) in simple decision problems and strategic settings, tend to decide longer (Arad & Rubinstein, 2012; Rubinstein, 2007, 2016), and shorter RTs often reflect errors (Rubinstein, 2013). Agranov, Caplin, and Tergiman (2015) demonstrated that, within a single decision, over time, sophisticated players tend to display higher cognitive levels. These results led to a conjecture that we could classify decision as intuitive or deliberative be merely looking at RT, and conclude whether people used their fast, habitual decision system 1 or slow, contemplative decision system 2 when they were making their decision (Kahneman, 2013). Specifically, a number of studies demonstrated that people are typically fast to make prosocial decisions and slow to make selfish decisions in simple games such as the public goods game or the prisoner's dilemma (Bear & Rand, 2016; Piovesan & Wengström, 2009; Rand, 2016; Rand et al., 2012, 2014), and that intuition

promotes cooperation: having less time to decide, people tend to go with the prosocial option (Rand et al., 2012).

This model-free approach, however, suffered from a problem of "reverse inference". It is tempting to use process data such as RT, classify this data based on a simple verbal theory, and apply this classification to a broad spectrum of decisions. But while we know that contemplative decisions tend to be long, it does not imply that each long decision is contemplative.

The drift-diffusion framework predicts that for a selfish individual picking a selfish option is an easy fast choice, and so is picking a more equal split for an altruistic individual. This implies that if we offer the same decision problem (say, a choice between a 90$/10$ and a $50/$50 split between the subject and a stranger) to a set of subjects and measure the average response time, we will observe that one type of decisions will take longer than the other simply because some subjects (for instance, very selfish ones) would find this choice very easy, while the others (somewhat altruistic ones) will take some time to consider. In the end, we might end up having faster selfish decisions, just due to the parameters of the decision problem and the set of subjects we sampled (Krajbich et al., 2015).

This example illustrates a very important point: process data can be misleading and is only truly useful in combination with a (hopefully) correct model of the data-generating process. The drift-diffusion model is the key point of the debate here: It allows to predict whether choices should be faster or slower in new contexts, based on independent measures.

Note that, like the verbal theory of the dual decision system, this theory is also based on a set of assumptions: We must assume, in a formally defined way, that people accumulate evidence on the available options when they are deciding on social allocations. However, unlike the verbal dual systems theory, the evidence accumulation framework allows us to build quantitative predictions about both choices and RTs in various decision problems, both within and across individuals, demonstrating how RTs can help us disentangle various motives underlying social decisions. Note that the two approaches are not mutually exclusive, as it may be possible to reconcile the DDM with the dual-process approach (Caplin & Martin, 2016). One way to do this is to assume that fast, automatic decisions are qualitatively different from contemplative decisions, which employ the DDM-style comparison process; this idea is yet to receive its validation with other types of process data such as fMRI.

3.5. Tracking social learning and choice processes

Gaze data are typically used as a measure of attention, which is often divided into the bottom-up (observing and exploring the choice environment for the first time, guided by external stimuli) and top-down (goal-directed, internally initiated visual search) components (Katsuki & Constantinidis, 2014). Both are important for studies of decisions: Investigating bottom-up attention can help us understand how context and environmental factors (for instance, the size and order of choice options) can affect our choices, while top-down attention reflects the steps of internal information acquisition strategy (Coricelli et al., 2020).

One cheap and easy way to study attention is mouse-tracking. Many studies of behavior in strategic games used the so-called "mouse-lab" technique, where certain options or payoffs in the game were hidden behind boxes that the subject had to click on to observe the number. While this approach has yielded many useful observations (such as that in complex games individuals do not explore all possible options (Bigoni, 2010; Bigoni & Fort, 2013; Brocas et al., 2014, 2018; Chen et al., 2018; Costa-Gomes et al., 2001; Gordon-Hecker et al., 2020; Johnson et al., 2002)), some argue that the process of clicking is too invasive and costly for the participants (Glöckner & Betsch, 2008). Eye-tracking is not strongly affected by these considerations, since it is minimally invasive and does not require any additional actions from the subject.

In the social decision domain, eye-tracking has been used extensively to study the process of decisions in strategic games (Coricelli et al., 2020). In these studies, the subject typically observes the payoff matrix of a game, where each row (or column) represents their own choice, and the column (or the row) shows the choice of the opponent. Each cell then shows the outcome of each choice combination, namely the payoffs of both players. One famous example is the Prisoner's Dilemma (PD): both players choose either to cooperate or defect, and while mutual cooperation is the best social outcome, for both players defection is the *dominant* strategy: independent of the opponent's choice, it is more beneficial (individually) to defect.

Classic game-theoretic models assume that individuals are rational and fully informed, so they use all available information (in our case, payoffs of themselves and the opponent) to make their choices. If all players are rational, the game results in the *equilibrium* behavior: upon the outcome, no player has an incentive to deviate from the chosen strategy (Camerer, 2003). In the PD game, the rational thing to do is to always defect. However, real people (in the laboratory) often deviate from the equilibrium play (for instance, cooperate in the PD), even with experience, displaying *bounded rationality*. Process data such as eye-tracking can help us understand whether this may reflect their information acquisition strategy, which may be based on their individual preferences or beliefs.

If the matrix presented on the screen is large enough, we can assume that the subject will look at each payoff to process that information (since their peripheral vision is not accurate enough) and that thus, if a payoff was not gazed at, the subject cannot possibly use it in her decision. For instance, if the subject did not look at the payoffs of other player in the PD, it might indicate several possibilities: Perhaps their choice model does not include social preferences, or they fail to understand that the other player's payoff impact the outcome of the game.

Using eye tracking, several studies showed that individuals indeed greatly vary in their information processing, and that impacts on their strategic choices (Costa-Gomes & Crawford, 2006; Devetag et al., 2016; Hausfeld, Fischbacher, et al., 2020; Hausfeld, von Hesler, et al., 2020; Knoepfle et al., 2009; Polonio et al., 2015; Polonio & Coricelli, 2019; Stewart et al., 2016; Wang et al., 2010; Zonca et al., 2019, 2020). Some individuals just compare their own payoffs to the opponent's payoffs within each cell, others do not consider the opponent's payoffs at all, and some participants carefully consider the whole payoff matrix. The studies revealed that these attentional strategies are stable within individual and predict deviations from the equilibrium play.

It is important to recognize here that we still do not completely understand whether these information sampling strategies simply reflect subjects' preferences (see section 3.1) or are a result of habits or heuristics that are completely independent from social behavior. Eye- and mouse-tracking data, combined with control non-social tasks, could give insight into the role of information acquisition strategies in social choice and determine under which conditions it might be influenced by bottom-up factors (such as display design) or top-down considerations (such as goals and preferences of individuals). Eventually, these insights can help us build new models of strategic interactions that explicitly include the attentional component.

3.6. Neural data and model validation

Many early fMRI studies used simple verbal behavioral models or identified correlates of explicit decision variables (such as rewards, monetary amounts, and types of decisions). These studies identified specific regions that represent social values (Bartra et al., 2013; Clithero & Rangel, 2014) or others' intentions (Cooper et al., 2010), maintain trust (Chang et al., 2011a; Sanfey et al., 2003), make us win in auctions (Delgado et al., 2008), guide our

strategic interactions (Coricelli & Nagel, 2009) and cooperation and competition decisions (Decety et al., 2004), provide us with theory of mind (Saxe, 2006; Saxe & Kanwisher, 2003), to name a few. These studies created a concept of the "social brain", a network of regions that includes the dorsomedial prefrontal cortex (dmPFC), precuneus, left and right temporoparietal junctions (TPJ), and left and right temporal poles. All these regions are typically implicated in a wide variety of social tasks.

Many studies have now suggested that we should use neural data to build a more mechanistic model of the decisions in the brain, where the same regions perform the same algorithmic computations in multiple domains including both social and non-social tasks (Lockwood et al., 2020). This approach can potentially allow us to understand whether the brain employs some kind of "common value currency" for social and non-social decisions (both social and non-social evaluations, for instance, evoke responses in the ventromedial prefrontal cortex and the ventral striatum (Lin et al., 2012)), or whether there is some "social-specific cognition" implemented in the brain (Ruff & Fehr, 2014).

Specifically, using neural data can help us to validate the specific components of computational models (i.e., latent variables that reflect various stages of the decision process). One prominent example of validation of behavioral models using neural data is the study of reward learning and belief updating. A seminal study in monkeys (Schultz, 1997) demonstrated that the dopamine neurons in the midbrain encode prediction errors: fire when the reward is not expected, but delivered (positive prediction error), and suppress activity when the reward is expected, but not delivered (negative prediction error).

Later experiments using neural data found evidence for similar prediction error-based updating in social beliefs in humans. One example is strategic beliefs (Hampton et al., 2008): it has been shown that first-order beliefs correlate with activity in the dorsomedial prefrontal cortex (dmPFC) (Behrens et al., 2008; Bhatt & Camerer, 2005; Hampton et al., 2008; Zhu et al., 2012), while second-order beliefs correlate with activity in the right temporoparietal junction (rTPJ) (Bhatt et al., 2010; Hampton et al., 2008; Hill et al., 2017). These studies demonstrated how these two latent components (first- and second-order beliefs) can be independently identified using neural data, by detecting the regions potentially performing these computations. The goal of building better models of social interactions clearly overlaps here with the goal of better understanding the human brain.

The same computational approach can be applied to many other domains, from learning of social value (FeldmanHall et al., 2017, 2018) and updating social impressions (Ma et al., 2012; Mende-Siedlecki et al., 2013) to observational learning (Burke et al., 2010; Charpentier et al., 2020; Dunne et al., 2016; Lindström et al., 2018; Park et al., 2019; Suzuki et al., 2012), conformity (Huber et al., 2015; Klucharev et al., 2009) and morality (Crockett, 2013; Crockett et al., 2013; Hutcherson et al., 2015).

Since fMRI has been applied to basically any aspect of social decisions (preferences, choice, and learning), we cannot potentially cover all studies that use neural data here (for extensive reviews, see Charpentier & O'Doherty, 2018; Chen & Hong, 2018; Fehr & Krajbich, 2014; Hackel & Amodio, 2018; Kliemann & Adolphs, 2018; Konovalov et al., 2018; Ruff & Fehr, 2014; Stanley & Adolphs, 2013; Zhang et al., 2020). However, we can emphasize that we are still in the very early stages of our understanding of the human brain. We believe that we need to shift our focus from functional specializations of single areas to understanding of the causal relationships between larger brain networks, as well as from testing idiosyncratic social situations to more mechanistic understanding of social decisions and finding common model components across different social domains.

One notable example is the role of the social brain network and specifically the temporoparietal junction (TPJ) in the social decision making. While these "social" regions have been implicated in many social choice domains, their exact functional roles,

communication patterns, and involvement in non-social tasks remain a puzzle. It is now clear that the TPJ does not just process theory of mind or compute higher-order beliefs, but is also involved in attention re-orientation and other non-social cognitions (Carter & Huettel, 2013). It appears that activations in "social" brain regions might be just a part of more domain-general cognitive mechanisms (such as updating the cognitive map of the world, forming beliefs, evaluation and updating of rewards, and so on) that are simply pertinent to complex decisions such as social behavior. Using neural data thus can help us to identify the neurocomputational mechanisms that are shared by non-social and social decisions (for instance, belief updating or mapping states of the world to actions). Another important aspect that has been understudied is temporal dynamics of social decisions: While EEG has been extensively applied in the studies of perceptual decision making, it can be invaluable in pinning down the within-trial dynamics of social choice but has not been as widely used in social neuroscience as fMRI (Zhang, 2018).

4. STUDYING CHOICE MODELS USING PROCESS DATA: GAPS AND NEW DIRECTIONS

In this section, we will identify gaps in the existing literature and offer potential future directions for the use of process data in social science. Certain types of process data have been predominantly used to study quite specific aspects of decision models, but the same data could – and should – be used to study other domains as well. Table 1 presents an overview of studies discussed in Section 3, split by the type of model (preference, choice, and learning) and the type of process data (RT, eye-tracking, mouse-tracking, neural data, and brain stimulation); while some cells of this table contain many studies, other ones still have strong potential for new applications.

While there has been a lot work investigating the relationship between social preferences, decision mechanisms, and RT (see Spiliopoulos & Ortmann (2017) for a deeper review), work on how learning mechanisms can be investigated with RTs, especially in social settings, is scarce. One promising approach is the combination of the DDM-type choice mechanism and learning models (Spiliopoulos, 2018; Tarantola et al., 2017), using computational models to make predictions about choices and RT at the same time. These models can pin down the role of priors in social decision making, as well as the dynamics of social belief updating, using not just binary choice data, but also continuous estimates of RT. One important question here is whether social cognition employs a similar value updating and comparison process as simple perceptual and economic decisions.

Another promising line of work is the relationship between strategic sophistication and RT in various types of game settings (Alós-Ferrer & Buckenmaier, 2020; Alós-Ferrer & Ritschel, 2018; Rubinstein, 2016). As we discussed above, individuals tend to make better decisions in social settings if they take longer time to decide. However, first, there are many factors that influence these decisions, such motor skills, demographic characteristics, preferences, decision strategies, cognitive abilities, and many others. One important future challenge is to categorize and classify these types of influences in social decisions. Second, if decision times can be observed, other individuals can make their own inferences about the others' preferences and decisions and might adjust their behavior by simply perceiving others' response times. This effect might have important implications for many economics interactions such as bargaining (a seller might infer a buyer's strength of preference even if she declines the initial offer).

Table 1. Studies using process data in model-based social decision-making

	Preferences	Choice	Learning
Response times	Hutcherson et al., 2015; Johnson et al., 2017; Chen & Krajbich, 2018; Bottemanne & Dreher, 2019; Konovalov & Krajbich, 2019; Hausfeld et al., 2020	Rubinstein, 2007, 2016; Piovesan & Wengström, 2009; Arad & Rubinstein, 2012; Rand et al., 2012, 2014; Krajbich, Bartling, et al., 2015; Hutcherson et al., 2015; Agranov et al., 2015; Rubinstein, 2016; Bear & Rand, 2016; Rand, 2016; Alós-Ferrer & Ritschel, 2018; Alós-Ferrer & Buckenmaier, 2020; Golman et al., 2020; Alós-Ferrer & Ritschel, 2021	Spiliopoulos, 2018; Tarantola et al., 2017
Eye-tracking	Fiedler et al., 2013; Gharib et al., 2015; Jiang et al., 2016; Hausfeld et al., 2020; Barrafrem & Hausfeld, 2020; Fosgaard et al., 2020; Teoh et al., 2020	Costa-Gomes & Crawford, 2006; Knoepfle et al., 2009; Wang et al., 2010; Devetag et al., 2016; Polonio et al., 2015; Stewart et al., 2016; Polonio & Coricelli, 2019; Zonca et al., 2019, 2020; Hausfeld, von Hesler, et al., 2020; Teoh et al., 2020	Knoepfle et al., 2009
Mouse-tracking	Brocas et al., 2014; Hausfeld et al., 2020; Gordon-Hecker et al., 2020	Costa-Gomes et al., 2001; Johnson et al., 2002; Bigoni, 2010; Bigoni & Fort, 2013; Brocas et al., 2014; Brocas et al., 2018; Chen et al., 2018	Bigoni, 2010; Bigoni & Fort, 2013
Neural data	Tricomi et al., 2010; Chang et al., 2011; Morishima et al., 2012; Emonds et al., 2012; Crockett, 2013; van den Bos et al., 2013; Crockett et al., 2013; Hutcherson et al., 2015; Lockwood et al., 2016; Soutschek et al., 2017; Holper et al., 2018; Harris et al., 2018	Bhatt & Camerer, 2005; Crockett, 2013; Crockett et al., 2013; Hutcherson et al., 2015; Tusche & Hutcherson, 2018; Harris et al., 2018	Bhatt & Camerer, 2005; Behrens et al., 2008; Hampton et al., 2008; Klucharev et al., 2009; Burke et al., 2010; Bhatt et al., 2010; Cooper et al., 2010; Zhu et al., 2012; Ma et al., 2012; Suzuki et al., 2012; van den Bos et al., 2013; Mende-Siedlecki et al., 2013; Huber et al., 2015; Lockwood et al., 2015; Dunne et al., 2016; Lockwood et al., 2016; Hill et al., 2017; FeldmanHall et al., 2017, 2018; Lindström et al., 2018; Lockwood et al., 2018; Park et al., 2019; Charpentier et al., 2020
Brain stimulation	Knoch et al., 2006; Silani et al., 2013; Young et al., 2010	Knoch et al., 2009; Wout et al., 2005; Ruff et al., 2013	Hill et al., 2017

Finally, we also know from non-social tasks that RTs can reflect uncertainty, surprise, and prediction errors, also represented in neural data (Chumbley et al., 2014; Konovalov & Krajbich, 2018; O'Reilly et al., 2013; Schiffer et al., 2012). In a standard serial reaction time task (SRT), a subject is instructed to respond to a stimulus with a corresponding button press. Typically, if the stimulus is expected (surprise is low, and probabilistic belief of the individual that this specific stimulus will appear on the screen is high), then individuals tend to react faster than in the case when the uncertainty of beliefs about which stimulus to expect is high. This approach can be easily applied to the study of belief formation if the researcher separates the process of choice from the process of observing the outcome (by additionally requiring an outcome reaction). This way reaction times could provide a rich set of data directly measuring first-order beliefs in repeated social interactions such as competitive games, trust games, and cooperative decisions.

A similar approach can be employed using mouse-tracking. For instance, the researcher can use mouse cursor position to infer decision conflict, subjective beliefs, and other latent variables that do not just reflect information acquisition strategies (as discussed in section 3). Studies in individual decision making demonstrated that mouse-tracking can be used in multiple ways: to track cursor trajectories and map them to drift-diffusion models (Sullivan et

al., 2015), estimate subjective beliefs (Konovalov & Krajbich, 2020), and track categorization decisions (Stillman et al., 2018).

While there has been significant progress in our understanding of strategic decisions, there are still gaps in the literature in terms of applying the eye-tracking technique to study preference formation, learning, and belief updating. Jiang et al. (2016) and Fiedler et al. (2013) show that the degree of attention to other's payoffs can explain social value orientation, but there is still much work to be done to demonstrate the causal role of attention in social decisions. One potential line of work here is causal manipulation of attention in social decisions (for instance, nudging people to donate more for charity by adjusting the display, or achieving better outcomes in public goods problems). Another application of eye-tracking that could be borrowed from individual decision making (Bakst & McGuire, 2020) is the use of predictive gaze paradigms (in which a stimulus appears on the screen and its location has to be predicted by shifting one's gaze towards it) to quantify social beliefs.

Another promising direction for eye-tracking is the use of pupillometry. It has been noted that pupil size might carry social information (Ebitz et al., 2014; Kret & De Dreu, 2019), signal dishonest behavior (van Breen et al., 2018), or condition and promote trust (Kret & De Dreu, 2017; Prochazkova et al., 2018). However, pupil data have been rarely used to validate computational models – particularly so in the social domain. Nevertheless, recent evidence from studies on perceptual and individual value-based decision making indicates that pupil dilation can be linked to specific decision-mechanisms such as the drift-diffusion process (Cavanagh et al., 2014; de Gee et al., 2014; Ebitz & Platt, 2015; Urai et al., 2017). An interesting avenue for future studies is thus to characterize potentially corresponding mechanisms at work during social decision making.

Finally, we want to emphasize that in most studies of social behavior, process data are identified as a mere correlate of potential decision mechanisms (or latent variables). It is therefore important to use the insights from process data analyses to build *mechanistical* models of social decisions. These models can provide predictions for (a) new experimental interventions that could vary conditions and contexts that influence specific decision mechanisms, or (b) brain stimulation protocols that affect behavior by disrupting or enhancing activity in the brain regions that are shown to perform the computations underlying these mechanisms. While the process data analysis could potentially allow us to identify information sampling strategies or specific brain regions responsible for specific computations, only interventional techniques (brain stimulation, pharmacological interventions, or new experimental designs) can provide a causal test of the role of these specific computations for social choice.

On the most basic level, process data can motivate new experimental designs. If RT, gaze data, or neural data confirm that a specific stimulus or parameter of the experiment can change the underlying mechanism (measured by process data), which eventually affects behavior, the experimenter could change this stimulus or parameter and causally demonstrate, using process data, that this specific mechanism indeed determines choice (for instance, by manipulating the size of payoff in a strategic game, showing that attention to payoffs impacts strategic choices).

A more advanced approach is non-invasive brain stimulation and pharmacological interventions that affect neural processes. Transcranial magnetic stimulation (TMS) is a popular non-invasive technique that uses brief, high-intensity magnetic field to excite or inhibit a small brain area close to the scalp. Many studies, inspired by model-based fMRI, showed that disrupting the brain activity in the dorsolateral prefrontal cortex (dlPFC) or TPJ can influence individual behavior. For instance, it can change fairness preferences (Knoch et al., 2006), change behavior in the ultimatum game (Knoch et al., 2009; Wout et al., 2005), decrease the use of second-order beliefs (Hill et al., 2017), change social norm compliance

(Ruff et al., 2013), social judgement (Silani et al., 2013), and moral judgement (Young et al., 2010). While the use of this method is restricted to the surface regions of the brain, it can provide invaluable causal validation of the functional role of specific neural processes in the decision mechanisms. As a more concrete example, we knew from fMRI studies that activity in the temporoparietal junction (TPJ) is correlated with second-order belief updates (Hampton et al., 2008). However, mere correlation does not imply that this region is involved in processing these beliefs and making strategic decisions; for instance, this activity might be just a co-manifestation of activity in another region that is functionally connected to the TPJ. Based on these considerations, a combined TMS-fMRI study demonstrated this causality, by showing that disruption of activity in the TPJ indeed affects both TPJ activity/connectivity as well as behavior - making individuals rely less on second-order belief updating – as predicted by a computational model of TPJ function in this context (Hill et al., 2017).

Another potential way of influencing neural processes is pharmacological interventions. For instance, some studies demonstrate that the hormone oxytocin might affect social behavior (Aydogan et al., 2017; Baumgartner et al., 2008), however some of these effects tend to be weak when tested in larger subject samples (Declerck et al., 2020; Nave et al., 2015). Nevertheless, these manipulations, combined with neural data, can potentially help us establish more direct links between process data about biological processes and motives and cognitive processes specified by computational models of social behavior.

5. CONCLUSION

Process data are increasingly used in the studies of social decisions. Here we described the main types of such data (response times, mouse- and eye-tracking, and neural data) and how they are used to validate and improve computational models of social behavior. We believe that this is a promising approach that will allow us to better understand human behavior and create better policy interventions, for at least four reasons. First, process data can allow us to validate mechanistic components of computational models such as parameters and latent variables. Second, we can use process data to differentiate between competing theories of social behavior, by building models that make predictions for new decision situations and new sets of subjects. Third, process data can track distinct components of the decision process such as information sampling and pin down the specific role of these components in preference formation and belief updating. Finally, process data can motivate new interventions such as new behavioral experiments and causal manipulations of neural activity or attention.

While some methods (such as fMRI) have been extensively applied to the study of social decisions, some remain uncommon, but could be used creatively in future studies. As we pointed out, some of the process data methods are more popular within specific model frameworks, but remain underused in other settings. As an example, we would welcome new applications of EEG, eye-tracking, and mouse-tracking to study social preference formation, as well the use of RTs to investigate the process of belief updating in strategic settings. Process data can also stimulate development of models that do not just predict choice, but also the process data themselves. While prediction of RTs with evidence accumulation models is very common, it might be beneficial to use the same model to predict the gaze data (Krajbich et al., 2010) or the neural data (Turner et al., 2013) at the same time as the choices. Explaining and predicting choices is an important goal, but richer computational models that also predict corresponding processes (and the corresponding data) can help us arbitrate between competing theories of social motives. While brain stimulation has been applied in certain paradigms, it is still not as common as fMRI, and simultaneous use of fMRI and TMS could provide valuable insights into functional roles of the cortical brain regions (Polanía et al., 2018).

Funding Information
This publication has been supported with funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 725355 BRAINCODES).

References

Abrams, D., & Hogg, M. A. (2006). *Social identifications: A social psychology of intergroup relations and group processes*. Routledge.

Agranov, M., Caplin, A., & Tergiman, C. (2015). Naive play and the process of choice in guessing games. *Journal of the Economic Science Association*, *1*(2), 146–157. https://doi.org/10.1007/s40881-015-0003-5

Alós-Ferrer, C., & Buckenmaier, J. (2020). Cognitive sophistication and deliberation times. *Experimental Economics*. https://doi.org/10.1007/s10683-020-09672-w

Alós-Ferrer, C., Granić, Đ.-G., Kern, J., & Wagner, A. K. (2016). Preference reversals: Time and again. *Journal of Risk and Uncertainty*, *52*(1), 65–97.

Alós-Ferrer, C., & Ritschel, A. (2018). The reinforcement heuristic in normal form games. *Journal of Economic Behavior & Organization*, *152*, 224–234.

Alós-Ferrer, C., & Ritschel, A. (2021). Multiple behavioral rules in Cournot oligopolies. *Journal of Economic Behavior & Organization*, *183*, 250–267. https://doi.org/10.1016/j.jebo.2020.12.034

Arad, A., & Rubinstein, A. (2012). Multi-dimensional iterative reasoning in action: The case of the Colonel Blotto game. *Journal of Economic Behavior & Organization*, *84*(2), 571–585. https://doi.org/10.1016/j.jebo.2012.09.004

Aydogan, G., Furtner, N. C., Kern, B., Jobst, A., Müller, N., & Kocher, M. G. (2017). Oxytocin promotes altruistic punishment. *Social Cognitive and Affective Neuroscience*, *12*(11), 1740–1747. https://doi.org/10.1093/scan/nsx101

Bakst, L., & McGuire, J. T. (2020). Eye movements reflect adaptive predictions and predictive precision. *Journal of Experimental Psychology: General*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/xge0000977

Barrafrem, K., & Hausfeld, J. (2020). Tracing risky decisions for oneself and others: The role of intuition and deliberation. *Journal of Economic Psychology*, *77*, 102188. https://doi.org/10.1016/j.joep.2019.102188

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. https://doi.org/10.1016/j.neuroimage.2013.02.063

Baumgartner, T., Heinrichs, M., Vonlanthen, A., Fischbacher, U., & Fehr, E. (2008). Oxytocin shapes the neural circuitry of trust and trust adaptation in humans. *Neuron*, *58*, 639–650.

Bear, A., & Rand, D. G. (2016). Intuition, deliberation, and the evolution of cooperation. *Proceedings of the National Academy of Sciences*, *113*(4), 936–941. https://doi.org/10.1073/pnas.1517780113

Bebko, C., Sciulli, L. M., & Bhagat, P. (2014). Using Eye Tracking to Assess the Impact of Advertising Appeals on Donor Behavior. *Journal of Nonprofit & Public Sector Marketing*, *26*(4), 354–371. https://doi.org/10.1080/10495142.2014.965073

Behrens, T. E. J., Hunt, L. T., & Rushworth, M. F. S. (2009). The Computation of Social Behavior. *Science*, *324*(5931), 1160–1164. https://doi.org/10.1126/science.1169694

Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature*, *456*(7219), 245–249. https://doi.org/10.1038/nature07538

Bhatt, M. A., Lohrenz, T., Camerer, C. F., & Montague, P. R. (2010). Neural signatures of strategic types in a two-person bargaining game. *Proceedings of the National Academy of Sciences*, *107*(46), 19720–19725. https://doi.org/10.1073/pnas.1009625107

Bhatt, M., & Camerer, C. F. (2005). Self-referential thinking and equilibrium as states of mind in games: FMRI evidence. *Games and Economic Behavior*, *52*(2), 424–459.

Bigoni, M. (2010). What do you want to know? Information acquisition and learning in experimental Cournot games. *Research in Economics*, *64*(1), 1–17. https://doi.org/10.1016/j.rie.2009.12.001

Bigoni, M., & Fort, M. (2013). Information and learning in oligopoly: An experiment. *Games and Economic Behavior*, *81*, 192–214. https://doi.org/10.1016/j.geb.2013.05.006

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5), 992–1026.

Brocas, I., Carrillo, J. D., & Sachdeva, A. (2018). The path to equilibrium in sequential and simultaneous games: A mousetracking study. *Journal of Economic Theory*, *178*, 246–274. https://doi.org/10.1016/j.jet.2018.09.011

Brocas, I., Carrillo, J. D., Wang, S. W., & Camerer, C. F. (2014). Imperfect Choice or Imperfect Attention? Understanding Strategic Thinking in Private Information Games. *The Review of Economic Studies*, *81*(3), 944–970. https://doi.org/10.1093/restud/rdu001

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178.

Burke, C. J., Tobler, P. N., Baddeley, M., & Schultz, W. (2010). Neural mechanisms of observational learning. *Proceedings of the National Academy of Sciences*, *107*(32), 14431–14436. https://doi.org/10.1073/pnas.1003111107

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 538–564. https://doi.org/10.1037/0278-7393.11.3.538

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision making. *Journal of Mathematical Psychology*, *32*, 91–134.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*(3), 432–459.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Camerer, C. F., & Fehr, E. (2006). When does" economic man" dominate social behavior? *Science*, *311*(5757), 47–52.

Caplin, A., & Martin, D. (2016). THE DUAL-PROCESS DRIFT DIFFUSION MODEL: EVIDENCE FROM RESPONSE TIMES. *Economic Inquiry*, *54*(2), 1274–1282. https://doi.org/10.1111/ecin.12294

Carter, R. M., & Huettel, S. A. (2013). A nexus model of the temporal–parietal junction. *Trends in Cognitive Sciences*, *17*(7), 328–336.

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General*, *143*(4), 1476–1488. https://doi.org/10.1037/a0035813

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011a). Triangulating the Neural, Psychological, and Economic Bases of Guilt Aversion. *Neuron*, *70*, 560–572.

Chang, L. J., Smith, A., Dufwenberg, M., & Sanfey, A. G. (2011b). Triangulating the neural, psychological, and economic bases of guilt aversion. *Neuron*, *70*(3), 560–572. https://doi.org/10.1016/j.neuron.2011.02.056

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *Quarterly Journal of Economics*, 817–869.

Charpentier, C. J., Iigaya, K., & O'Doherty, J. P. (2020). A Neuro-computational Account of Arbitration between Choice Imitation and Goal Emulation during Human Observational Learning. *Neuron*, *106*(4), 687-699.e7. https://doi.org/10.1016/j.neuron.2020.02.028

Charpentier, C. J., & O'Doherty, J. P. (2018). The application of computational models to social neuroscience: Promises and pitfalls. *Social Neuroscience*, *13*(6), 637–647. https://doi.org/10.1080/17470919.2018.1518834

Chen, C.-T., Huang, C.-Y., & Wang, J. T. (2018). A window of cognition: Eyetracking the reasoning process in spatial beauty contest games. *Games and Economic Behavior*, *111*, 143–158. https://doi.org/10.1016/j.geb.2018.05.007

Chen, P., & Hong, W. (2018). Neural Circuit Mechanisms of Social Behavior. *Neuron*, *98*(1), 16–30. https://doi.org/10.1016/j.neuron.2018.02.026

Chumbley, J. R., Burke, C. J., Stephan, K. E., Friston, K. J., Tobler, P. N., & Fehr, E. (2014). Surprise beyond prediction error: Surprise Beyond Prediction Error. *Human Brain Mapping*, *35*(9), 4805–4814. https://doi.org/10.1002/hbm.22513

Clithero, J. A. (2018). Response times in economics: Looking through the lens of sequential sampling models. *Journal of Economic Psychology*, *69*, 61–86. https://doi.org/10.1016/j.joep.2018.09.008

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.

Collette, S., Pauli, W. M., Bossaerts, P., & O'Doherty, J. (2017). Neural computations underlying inverse reinforcement learning in the human brain. *ELife*, *6*.

Cooper, D. J., Krajbich, I., & Noussair, C. N. (2019). Choice-Process Data in Experimental Economics. *Journal of the Economic Science Association*, *5*(1), 1–13. https://doi.org/10.1007/s40881-019-00075-z

Cooper, J. C., Kreps, T. A., Wiebe, T., Pirkl, T., & Knutson, B. (2010). When Giving Is Good: Ventromedial Prefrontal Cortex Activation for Others' Intentions. *Neuron*, *67*(3), 511–521. https://doi.org/10.1016/j.neuron.2010.06.030

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(23), 9163–9168.

Coricelli, G., Polonio, L., & Vostroknutov, A. (2020). The process of choice in games. In *Handbook of Experimental Game Theory*. Edward Elgar Publishing.

Costa-Gomes, M. A., & Crawford, V. P. (2006). Cognition and Behavior in Two-Person Guessing Games: An Experimental Study. *American Economic Review*, *96*(5), 1737–1768. https://doi.org/10.1257/aer.96.5.1737

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and Behavior in Normal-Form Games: An Experimental Study. *Econometrica*, *69*(5), 1193–1235. https://doi.org/10.1111/1468-0262.00239

Crockett, M. J. (2013). Models of morality. *Trends in Cognitive Sciences*, *17*(8), 363–366. https://doi.org/10.1016/j.tics.2013.06.005

Crockett, M. J., Braams, B. R., Clark, L., Tobler, P. N., Robbins, T. W., & Kalenscher, T. (2013). Restricting Temptations: Neural Mechanisms of Precommitment. *Neuron*, *79*(2), 391–401. https://doi.org/10.1016/j.neuron.2013.05.028

Dashiell, J. F. (1937). Affective value-distances as a determinant of esthetic judgment-times. *The American Journal of Psychology*.

de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences*, *111*(5), E618–E625.

Decety, J., Jackson, P. L., Sommerville, J. A., Chaminade, T., & Meltzoff, A. N. (2004). The neural bases of cooperation and competition: An fMRI investigation. *NeuroImage*, *23*(2), 744–751. https://doi.org/10.1016/j.neuroimage.2004.05.025

Declerck, C. H., Boone, C., Pauwels, L., Vogt, B., & Fehr, E. (2020). A registered replication study on oxytocin and trust. *Nature Human Behaviour*, *4*(6), 646–655. https://doi.org/10.1038/s41562-020-0878-x

Delgado, M. R., Schotter, A., Ozbay, E. Y., & Phelps, E. A. (2008). Understanding Overbidding: Using the Neural Circuitry of Reward to Design Economic Auctions. *Science*, *321*(5897), 1849–1852. https://doi.org/10.1126/science.1158860

Devaine, M., Hollard, G., & Daunizeau, J. (2014). The social Bayesian brain: Does mentalizing make a difference when we learn? *PLoS Comput Biol*, *10*(12), e1003992.

Devetag, G., Di Guida, S., & Polonio, L. (2016). An eye-tracking study of feature-based choice in one-shot games. *Experimental Economics*, *19*(1), 177–201. https://doi.org/10.1007/s10683-015-9432-5

Diederich, A. (2003). MDFT account of decision making under time pressure. *Psychonomic Bulletin & Review*, *10*(1), 157–166. https://doi.org/10.3758/BF03196480

Doll, B. B., Simon, D. A., & Daw, N. D. (2012). The ubiquity of model-based reinforcement learning. *Current Opinion in Neurobiology*, *22*(6), 1075–1081. https://doi.org/10.1016/j.conb.2012.08.003

Duchowski, A. (2007). Eye tracking techniques. In *Eye Tracking Methodology* (pp. 51–59). Springer.

Dunne, S., D'Souza, A., & O'Doherty, J. P. (2016). The involvement of model-based but not model-free learning signals during observational reward learning in the absence of choice. *Journal of Neurophysiology*, *115*(6), 3195–3203. https://doi.org/10.1152/jn.00046.2016

Ebitz, R. B., Pearson, J. M., & Platt, M. L. (2014). Pupil size and social vigilance in rhesus macaques. *Frontiers in Neuroscience*, *8*, 100.

Ebitz, R. B., & Platt, M. L. (2015). Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. *Neuron*, *85*(3), 628–640.

Echenique, F., & Saito, K. (2017). Response time and utility. *Journal of Economic Behavior & Organization*, *139*, 49–59.

Emonds, G., Declerck, C. H., Boone, C., Vandervliet, E. J., & Parizel, P. M. (2012). The cognitive demands on cooperation in social dilemmas: An fMRI study. *Social Neuroscience*, *7*(5), 494–509.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610. https://doi.org/10.1007/s10683-011-9283-7

Erev, I., & Roth, A. E. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review*, 848–881.

Fehr, E., & Camerer, C. F. (2007). Social neuroeconomics: The neural circuitry of social preferences. *Trends in Cognitive Sciences*, *11*(10), 419–427.

Fehr, E., & Krajbich, I. (2014). Social Preferences and the Brain. In *Neuroeconomics* (pp. 193–218). Elsevier. http://linkinghub.elsevier.com/retrieve/pii/B9780124160088000115

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.
FeldmanHall, O., Dunsmoor, J. E., Kroes, M. C. W., Lackovic, S., & Phelps, E. A. (2017). Associative Learning of Social Value in Dynamic Groups. *Psychological Science*, *28*(8), 1160–1170. https://doi.org/10.1177/0956797617706394
FeldmanHall, O., Dunsmoor, J., Tompary, A., Hunter, L., Todorov, A., & Phelps, E. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences*, *115*, 201715227. https://doi.org/10.1073/pnas.1715227115
Fiedler, S., Glöckner, A., Nicklisch, A., & Dickert, S. (2013). Social value orientation and information search in social dilemmas: An eye-tracking analysis. *Organizational Behavior and Human Decision Processes*, *120*(2), 272–284.
Fosgaard, T., Jacobsen, C., & Street, C. (2020). The heterogeneous processes of cheating: Attention evidence from two eye tracking experiments. *Journal of Behavioral Decision Making*, *n/a*(n/a). https://doi.org/10.1002/bdm.2200
Fudenberg, D., Strack, P., & Strzalecki, T. (2018). Speed, accuracy, and the optimal timing of choices. *American Economic Review*, *108*(12), 3651–3684.
Genevsky, A., & Knutson, B. (2015). Neural Affective Mechanisms Predict Market-Level Microlending. *Psychological Science*, *26*(9), 1411–1422. https://doi.org/10.1177/0956797615588467
Gharib, A., Mier, D., Adolphs, R., & Shimojo, S. (2015). Eyetracking of social preference choices reveals normal but faster processing in autism. *Neuropsychologia*, *72*, 70–79.
Glöckner, A., & Betsch, T. (2008). Multiple-reason decision making based on automatic processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1055–1075. https://doi.org/10.1037/0278-7393.34.5.1055
Golman, R., Bhatia, S., & Kane, P. B. (2020). The dual accumulator model of strategic deliberation and decision making. *Psychological Review*, *127*(4), 477–504. https://doi.org/10.1037/rev0000176
Gordon-Hecker, T., Schneider, I. K., Shalvi, S., & Bereby-Meyer, Y. (2020). Leaving with something: When do people experience an equity–efficiency conflict? *Journal of Behavioral Decision Making*, *n/a*(n/a). https://doi.org/10.1002/bdm.2205
Hackel, L. M., & Amodio, D. M. (2018). Computational neuroscience approaches to social cognition. *Current Opinion in Psychology*, *24*, 92–97. https://doi.org/10.1016/j.copsyc.2018.09.001
Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences*, *105*(18), 6741–6746.
Harris, A., Clithero, J. A., & Hutcherson, C. A. (2018). Accounting for Taste: A Multi-Attribute Neurocomputational Model Explains the Neural Dynamics of Choices for Self and Others. *Journal of Neuroscience*, *38*(37), 7952–7968. https://doi.org/10.1523/JNEUROSCI.3327-17.2018
Hausfeld, J., Fischbacher, U., & Knoch, D. (2020). The value of decision-making power in social decisions. *Journal of Economic Behavior & Organization*, *177*, 898–912. https://doi.org/10.1016/j.jebo.2020.06.018
Hausfeld, J., von Hesler, K., & Goldlücke, S. (2020). Strategic gaze: An interactive eye-tracking study. *Experimental Economics*. https://doi.org/10.1007/s10683-020-09655-x
Hessels, R. S., Niehorster, D. C., Nyström, M., Andersson, R., & Hooge, I. T. C. (2018). Is the eye-movement field confused about fixations and saccades? A survey among 124 researchers. *Royal Society Open Science*, *5*(8), 180502. https://doi.org/10.1098/rsos.180502

Hill, C. A., Suzuki, S., Polania, R., Moisa, M., O'Doherty, J. P., & Ruff, C. C. (2017). A causal account of the brain network computations underlying strategic social behavior. *Nature Neuroscience*, *20*(8), 1142–1149. https://doi.org/10.1038/nn.4602

Holper, L., Burke, C. J., Fausch, C., Seifritz, E., & Tobler, P. N. (2018). Inequality signals in dorsolateral prefrontal cortex inform social preference models. *Social Cognitive and Affective Neuroscience*, *13*(5), 513–524.

Homans, G. C. (1974). *Social behavior: Its elementary forms*.

Huber, R. E., Klucharev, V., & Rieskamp, J. (2015). Neural correlates of informational cascades: Brain mechanisms of social influence on belief updating. *Social Cognitive and Affective Neuroscience*, *10*(4), 589–597. https://doi.org/10.1093/scan/nsu090

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (Vol. 1). Sinauer Associates Sunderland, MA.

Hutcherson, C. A., Bushong, B., & Rangel, A. (2015). A Neurocomputational Model of Altruistic Choice and Its Implications. *Neuron*, *87*(2), 451–462.

Jamieson, D. G., & Petrusic, W. M. (1977). Preference and the time to choose. *Organizational Behavior and Human Performance*, *19*(1), 56–67.

Jiang, T., Potters, J., & Funaki, Y. (2016). Eye-tracking Social Preferences. *Journal of Behavioral Decision Making*, *29*(2–3), 157–168. https://doi.org/10.1002/bdm.1899

Johnson, D. J., Hopwood, C. J., Cesario, J., & Pleskac, T. J. (2017). Advancing Research on Cognitive Processes in Social and Personality Psychology: A Hierarchical Drift Diffusion Model Primer. *Social Psychological and Personality Science*, *8*(4), 413–423. https://doi.org/10.1177/1948550617703174

Johnson, E. J., Camerer, C., Sen, S., & Rymon, T. (2002). Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining. *Journal of Economic Theory*, *104*(1), 16–47. https://doi.org/10.1006/jeth.2001.2850

Kahneman, D. (2013). *Thinking, Fast and Slow* (Reprint edition). Farrar, Straus and Giroux.

Kanske, P., Böckler, A., Trautwein, F.-M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain-behavior relations for empathy and Theory of Mind. *NeuroImage*, *122*, 6–19. https://doi.org/10.1016/j.neuroimage.2015.07.082

Katsuki, F., & Constantinidis, C. (2014). Bottom-up and top-down attention: Different processes and overlapping neural systems. *The Neuroscientist*, *20*(5), 509–521.

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (Vol. 2). Wiley New York.

Kliemann, D., & Adolphs, R. (2018). The social neuroscience of mentalizing: Challenges and recommendations. *Current Opinion in Psychology*, *24*, 1–6. https://doi.org/10.1016/j.copsyc.2018.02.015

Klucharev, V., Hytönen, K., Rijpkema, M., Smidts, A., & Fernández, G. (2009). Reinforcement Learning Signal Predicts Social Conformity. *Neuron*, *61*(1), 140–151. https://doi.org/10.1016/j.neuron.2008.11.027

Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science (New York, N.Y.)*, *314*(5800), 829–832. https://doi.org/10.1126/science.1129156

Knoch, D., Schneider, F., Schunk, D., Hohmann, M., & Fehr, E. (2009). Disrupting the prefrontal cortex diminishes the human ability to build a good reputation. *Proceedings of the National Academy of the United States of America*, *106*(49), 20895–20899.

Knoepfle, D. T., Wang, J. T., & Camerer, C. F. (2009). Studying Learning in Games Using Eye-tracking. *Journal of the European Economic Association*, *7*(2–3), 388–398.

Konovalov, A., Hu, J., & Ruff, C. C. (2018). Neurocomputational approaches to social behavior. *Current Opinion in Psychology*, *24*, 41–47. https://doi.org/10.1016/j.copsyc.2018.04.009

Konovalov, A., & Krajbich, I. (2018). Neurocomputational Dynamics of Sequence Learning. *Neuron*, *98*(6), 1282-1293.e4. https://doi.org/10.1016/j.neuron.2018.05.013

Konovalov, A., & Krajbich, I. (2019). Revealed strength of preference: Inference from response times. *Judgment and Decision Making*, 14.

Konovalov, A., & Krajbich, I. (2020). Mouse tracking reveals structure knowledge in the absence of model-based choice. *Nature Communications*, *11*(1), 1893. https://doi.org/10.1038/s41467-020-15696-w

Kosinski, R. J. (2008). A literature review on reaction time. *Clemson University*, *10*(1).

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

Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, *6*, 7455. https://doi.org/10.1038/ncomms8455

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.

Kret, M. E., & De Dreu, C. K. W. (2017). Pupil-mimicry conditions trust in partners: Moderation by oxytocin and group membership. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1850), 20162554. https://doi.org/10.1098/rspb.2016.2554

Kret, M. E., & De Dreu, C. K. W. (2019). The power of pupil size in establishing trust and reciprocity. *Journal of Experimental Psychology: General*, *148*(8), 1299–1311. https://doi.org/10.1037/xge0000508

Lin, A., Adolphs, R., & Rangel, A. (2012). Social and monetary reward learning engage overlapping neural substrates. *Social Cognitive and Affective Neuroscience*, *7*(3), 274–281. https://doi.org/10.1093/scan/nsr006

Lindström, B., Haaker, J., & Olsson, A. (2018). A common neural network differentially mediates direct and social fear learning. *NeuroImage*, *167*, 121–129. https://doi.org/10.1016/j.neuroimage.2017.11.039

Lockwood, P. L., Apps, M. A. J., & Chang, S. W. C. (2020). Is There a 'Social' Brain? Implementations and Algorithms. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2020.06.011

Lockwood, P. L., Apps, M. A. J., Roiser, J. P., & Viding, E. (2015). Encoding of Vicarious Reward Prediction in Anterior Cingulate Cortex and Relationship with Trait Empathy. *Journal of Neuroscience*, *35*(40), 13720–13727. https://doi.org/10.1523/JNEUROSCI.1703-15.2015

Lockwood, P. L., Apps, M. A. J., Valton, V., Viding, E., & Roiser, J. P. (2016). Neurocomputational mechanisms of prosocial learning and links to empathy. *Proceedings of the National Academy of Sciences*, *113*(35), 9763–9768. https://doi.org/10.1073/pnas.1603198113

Lockwood, P. L., Wittmann, M. K., Apps, M. A. J., Klein-Flügge, M. C., Crockett, M. J., Humphreys, G. W., & Rushworth, M. F. S. (2018). Neural mechanisms for learning self and other ownership. *Nature Communications*, *9*(1), 4747. https://doi.org/10.1038/s41467-018-07231-9

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

Ma, N., Vandekerckhove, M., Baetens, K., Van Overwalle, F., Seurinck, R., & Fias, W. (2012). Inconsistencies in spontaneous and intentional trait inferences. *Social*

Cognitive and Affective Neuroscience, *7*(8), 937–950. https://doi.org/10.1093/scan/nsr064
MacCrimmon, K. R., & Messick, D. M. (1976). A framework for social motives. *Behavioral Science*, *21*(2), 86–100.
Mas-Colell, A., Whinston, M. D., & Green, J. R. (1995). *Microeconomic theory* (Vol. 1). Oxford university press New York.
McFadden, D. L. (2005). Revealed stochastic preference: A synthesis. *Economic Theory*, *26*(2), 245–264.
Mende-Siedlecki, P., Cai, Y., & Todorov, A. (2013). The neural dynamics of updating person impressions. *Social Cognitive and Affective Neuroscience*, *8*(6), 623–631. https://doi.org/10.1093/scan/nss040
Moffatt, P. G. (2005). Stochastic choice and the allocation of cognitive effort. *Experimental Economics*, *8*(4), 369–388.
Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking Brain Structure and Activation in Temporoparietal Junction to Explain the Neurobiology of Human Altruism. *Neuron*, *75*(1), 73–79. https://doi.org/10.1016/j.neuron.2012.05.021
Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, *59*(5), 371–404.
Nave, G., Camerer, C., & McCullough, M. (2015). Does Oxytocin Increase Trust in Humans? A Critical Review of Research. *Perspectives on Psychological Science*, *10*(6), 772–789. https://doi.org/10.1177/1745691615600138
O'Reilly, J. X., Schuffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *110*(38), E3660–E3669. https://doi.org/10.1073/pnas.1305373110
Park, S. A., Sestito, M., Boorman, E. D., & Dreher, J.-C. (2019). Neural computations underlying strategic social decision-making in groups. *Nature Communications*, *10*. https://doi.org/10.1038/s41467-019-12937-5
Piovesan, M., & Wengström, E. (2009). Fast or fair? A study of response times. *Economics Letters*, *105*(2), 193–196. https://doi.org/10.1016/j.econlet.2009.07.017
Polanía, R., Nitsche, M. A., & Ruff, C. C. (2018). Studying and modifying brain function with non-invasive brain stimulation. *Nature Neuroscience*, *21*(2), 174–187. https://doi.org/10.1038/s41593-017-0054-4
Polonio, L., & Coricelli, G. (2019). Testing the level of consistency between choices and beliefs in games using eye-tracking. *Games and Economic Behavior*, *113*, 566–586. https://doi.org/10.1016/j.geb.2018.11.003
Polonio, L., Di Guida, S., & Coricelli, G. (2015). Strategic sophistication and attention in games: An eye-tracking study. *Games and Economic Behavior*, *94*, 80–96. https://doi.org/10.1016/j.geb.2015.09.003
Prochazkova, E., Prochazkova, L., Giffin, M. R., Scholte, H. S., Dreu, C. K. W. D., & Kret, M. E. (2018). Pupil mimicry promotes trust through the theory-of-mind network. *Proceedings of the National Academy of Sciences*, *115*(31), E7265–E7274. https://doi.org/10.1073/pnas.1803916115
Rand, D. G. (2016). Cooperation, Fast and Slow: Meta-Analytic Evidence for a Theory of Social Heuristics and Self-Interested Deliberation. *Psychological Science*, *27*(9), 1192–1206. https://doi.org/10.1177/0956797616654455
Rand, David G., Greene, J. D., & Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature*, *489*(7416), 427–430. https://doi.org/10.1038/nature11467

Rand, David G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., & Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, *5*. https://doi.org/10.1038/ncomms4677

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*(4), 873–922.

Rodriguez, C. A., Turner, B. M., & McClure, S. M. (2014). Intertemporal Choice as Discounted Value Accumulation. *PloS One*, *9*(2), e90138.

Roth, A. E., & Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behavior*, *8*(1), 164–212.

Rubinstein, A. (2007). Instinctive and cognitive reasoning: A study of response times*. *The Economic Journal*, *117*(523), 1243–1259.

Rubinstein, A. (2013). Response time and decision making: An experimental study. *Judgment and Decision Making*, *8*(5), 540–551.

Rubinstein, A. (2016). A typology of players: Between instinctive and contemplative. *The Quarterly Journal of Economics*, *131*(2), 859–890.

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing Social Norm Compliance with Noninvasive Brain Stimulation. *Science*, *342*(6157), 482–484. https://doi.org/10.1126/science.1241399

Ruff, Christian C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience*, *15*(8), 549–562. https://doi.org/10.1038/nrn3776

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The Neural Basis of Economic Decision-Making in the Ultimatum Game. *Science*, *300*, 1755–1758.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." *Neuroimage*, *19*(4), 1835–1842.

Schiffer, A.-M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at All the Entropy: Hippocampal, Caudate and Midbrain Contributions to Learning from Prediction Errors. *PLoS ONE*, *7*(5), e36445. https://doi.org/10.1371/journal.pone.0036445

Schulte-Mecklenbeck, M., Johnson, J. G., Böckenholt, U., Goldstein, D. G., Russo, J. E., Sullivan, N. J., & Willemsen, M. C. (2017). Process-Tracing Methods in Decision Making: On Growing Up in the 70s. *Current Directions in Psychological Science*, *26*(5), 442–450. https://doi.org/10.1177/0963721417708229

Schulte-Mecklenbeck, M., Kühberger, A., & Johnson, J. G. (2019). *A handbook of process tracing methods*. Routledge.

Schultz, W. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Sejnowski, T. J., Churchland, P. S., & Movshon, J. A. (2014). Putting big data to good use in neuroscience. *Nature Neuroscience*, *17*(11), 1440–1441. https://doi.org/10.1038/nn.3839

Silani, G., Lamm, C., Ruff, C. C., & Singer, T. (2013). Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(39), 15466–15476. https://doi.org/10.1523/JNEUROSCI.1488-13.2013

Small, D. A., & Verrochi, N. M. (2009). The face of need: Facial emotion expression on charity advertisements. *Journal of Marketing Research*, *46*(6), 777–787.

Smith, S. M., & Krajbich, I. (2018). Attention and Choice Across Domains. *Journal of Experimental Psychology: General*, *in press*.

Soutschek, A., Burke, C. J., Beharelle, A. R., Schreiber, R., Weber, S. C., Karipidis, I. I., Ten Velden, J., Weber, B., Haker, H., & Kalenscher, T. (2017). The dopaminergic reward system underpins gender differences in social preferences. *Nature Human Behaviour*, *1*(11), 819–827.

Spiliopoulos, L. (2018). The determinants of response time in a repeated constant-sum game: A robust Bayesian hierarchical dual-process model. *Cognition*, *172*, 107–123. https://doi.org/10.1016/j.cognition.2017.11.006

Spiliopoulos, L., & Ortmann, A. (2017). The BCD of response time analysis in experimental economics. *Experimental Economics*. https://doi.org/10.1007/s10683-017-9528-1

Stanley, D. A., & Adolphs, R. (2013). Toward a Neural Basis for Social Behavior. *Neuron*, *80*(3), 816–826. https://doi.org/10.1016/j.neuron.2013.10.038

Stewart, N., Gächter, S., Noguchi, T., & Mullett, T. L. (2016). Eye Movements in Strategic Choice. *Journal of Behavioral Decision Making*, *29*(2–3), 137–156. https://doi.org/10.1002/bdm.1901

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How Mouse-tracking Can Advance Social Cognitive Theory. *Trends in Cognitive Sciences*, *22*(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Sullivan, N., Hutcherson, C., Harris, A., & Rangel, A. (2015). Dietary Self-Control Is Related to the Speed With Which Attributes of Healthfulness and Tastiness Are Processed. *Psychological Science*, *26*(2), 122–134. https://doi.org/10.1177/0956797614559543

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Suzuki, S., Harasawa, N., Ueno, K., Gardner, J. L., Ichinohe, N., Haruno, M., Cheng, K., & Nakahara, H. (2012). Learning to Simulate Others' Decisions. *Neuron*, *74*(6), 1125–1137. https://doi.org/10.1016/j.neuron.2012.04.030

Tarantola, T., Kumaran, D., Dayan, P., & De Martino, B. (2017). Prior preferences beneficially influence social and non-social learning. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00826-8

Teoh, Y. Y., Yao, Z., Cunningham, W. A., & Hutcherson, C. A. (2020). Attentional priorities drive effects of time pressure on altruistic choice. *Nature Communications*, *11*(1), 3534. https://doi.org/10.1038/s41467-020-17326-x

Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, *463*(7284), 1089–1091.

Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, *72*, 193–206.

Tusche, A., Böckler, A., Kanske, P., Trautwein, F.-M., & Singer, T. (2016). Decoding the Charitable Brain: Empathy, Perspective Taking, and Attention Shifts Differentially Predict Altruistic Giving. *Journal of Neuroscience*, *36*(17), 4719–4732. https://doi.org/10.1523/JNEUROSCI.3392-15.2016

Tusche, A., & Hutcherson, C. A. (2018). Cognitive regulation alters social and dietary choice by changing attribute representations in domain-general and domain-specific brain circuits. *ELife*, *7*, e31185. https://doi.org/10.7554/eLife.31185

Tversky, A., & Shafir, E. (1992). Choice under conflict: The dynamics of deferred decision. *Psychological Science*, *3*(6), 358–361.

Urai, A. E., Braun, A., & Donner, T. H. (2017). Pupil-linked arousal is driven by decision uncertainty and alters serial choice bias. *Nature Communications*, *8*(1), 1–11.

van Breen, J. A., De Dreu, C. K. W., & Kret, M. E. (2018). Pupil to pupil: The effect of a partner's pupil size on (dis)honest behavior. *Journal of Experimental Social Psychology*, *74*, 231–245. https://doi.org/10.1016/j.jesp.2017.09.009

van den Bos, W., Talwar, A., & McClure, S. M. (2013). Neural Correlates of Reinforcement Learning and Social Preferences in Competitive Bidding. *Journal of Neuroscience*, *33*(5), 2137–2146. https://doi.org/10.1523/JNEUROSCI.3095-12.2013

Wang, J. T., Spezio, M., & Camerer, C. F. (2010). Pinocchio's Pupil: Using Eyetracking and Pupil Dilation to Understand Truth Telling and Deception in Sender-Receiver Games. *American Economic Review*, *100*(3), 984–1007. https://doi.org/10.1257/aer.100.3.984

Weick, K. (1977). Social psychology. *Psyccritiques*, *22*(5).

Wout, M. van't, Kahn, R., Sanfey, A. G., & Aleman, A. (2005). Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport*, *16*(16), 1849–1852.

Xiang, T., Lohrenz, T., & Montague, P. R. (2013). Computational Substrates of Norms and Their Violations during Social Exchange. *Journal of Neuroscience*, *33*(3), 1099–1108. https://doi.org/10.1523/JNEUROSCI.1642-12.2013

Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, *107*(15), 6753–6758.

Zhang, D. (2018). Computational EEG Analysis for Hyperscanning and Social Neuroscience. In C.-H. Im (Ed.), *Computational EEG Analysis: Methods and Applications* (pp. 215–228). Springer. https://doi.org/10.1007/978-981-13-0908-3_10

Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J., & Lamm, C. (2020). Using reinforcement learning models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Social Cognitive and Affective Neuroscience*, *15*(6), 695–707. https://doi.org/10.1093/scan/nsaa089

Zhu, L., Mathewson, K. E., & Hsu, M. (2012). Dissociable neural representations of reinforcement and belief prediction errors underlie strategic learning. *Proceedings of the National Academy of Sciences*, *109*(5), 1419–1424. https://doi.org/10.1073/pnas.1116783109

Zonca, J., Coricelli, G., & Polonio, L. (2019). Does exposure to alternative decision rules change gaze patterns and behavioral strategies in games? *Journal of the Economic Science Association*, *5*(1), 14–25. https://doi.org/10.1007/s40881-019-00066-0

Zonca, J., Coricelli, G., & Polonio, L. (2020). Gaze patterns disclose the link between cognitive reflection and sophistication in strategic interaction. *Judgment and Decision Making*, 16.