Two Neurocomputational Building Blocks of Social Norm Compliance

Abstract Current explanatory frameworks for social norms pay little attention to why and how brains might carry out computational functions that generate norm compliance behavior. This paper expands on existing literature by laying out the beginnings of a neurocomputational framework for social norms and social cognition, which can be the basis for advancing our understanding of the nature and mechanisms of social norms. Two neurocomputational building blocks are identified that might constitute the building blocks of the mechanism of norm compliance. They consist of Bayesian and Reinforcement Learning systems. It is sketched why and how the concerted activity of these systems can generate norm compliance by minimization of three specific kinds of prediction-errors.

Keywords: Social norms; Bayesian brain; reinforcement learning; uncertainty minimization

Social Norm Compliance from a Neurocomputational Perspective

Philosophers, psychologists, anthropologists, and economists have offered different accounts of social norms (e.g. Bicchieri 2006; Binmore 1994; Boyd and Richerson 2001; Elster 1989; Gintis 2010; Lewis 1969; Pettit 1990; Sugden 1986; Ullmann-Margalit 1977). Many facts are known about social norms, both at the individual and at the social level (Sripada and Stich 2007). However, much existing research is partial and piecemeal, making it difficult to know how individual findings cohere into a comprehensive picture. Relatively little effort has been spent in laying out a framework that could unify these facts and advance interdisciplinary research on social/moral[1] behavior.

[1] Although moral norms do not seem to be sharply distinct from social norms or, say, norms of disgust, there is a spectrum of social behaviors, some of which tend to be more readily called 'moral.' Specifically, behavioral patterns that typically involve a victim who has been

Computational cognitive neuroscience has the opportunity to make valuable contributions to our understanding of social/moral behavior. Such understanding can be grounded in a computational, biologically plausible framework, which can unify existing knowledge about norms, and help to guide the study of social normativity across multiple disciplines in a way where the concepts and data used by researchers are informed, constrained and modified by ideas and results from multiple disciplines.

What follows lays out the *beginnings* of a neurocomputational framework for social norms. Within this framework, two building blocks of social norm compliance are identified that might constitute the basis for the mechanism of norm compliance. They consist of Bayesian and Reinforcement Learning (RL) systems. It is canvassed why and how the concerted activity of these systems could generate norm compliance by minimization of three kinds of prediction-errors. On this account, Bayesian systems compute social representations, while RL systems draw on social representations to learn to comply with norms during social interaction.

The suggestion is that social/moral behavior piggybacks on neural computations that enable agents to process incoming sensory input so as to form probabilistic beliefs about the states of the world causing that input, and to choose actions so as to maximize the value of their future reward outcomes in the social world. Agents might learn social norms as they do other regularities in their environment, and comply with them courtesy of basic types of neural computations, which operate in both social and non-social contexts. Thus, social norms could be grounded in features of human nature, which are more fundamental than either the beliefs and preferences of individuals or the idiosyncratic characteristics of the culture in

---

harmed, whose rights have been violated, or who has been subject to some injustice seem to be more readily qualified as ‘moral’ norms.

which they live. The concerted activity of the Bayesian – RL systems would generate social norm compliance as opposed to any other form of behavior because of the *social* nature of the representations that they transform and consume.

Three points should be clear about the nature and scope of this proposal before proceeding to unpacking it. First, the approach adopted here is unlike that of a number of philosophers, cognitive scientists, and social scientists working on social norms within the tradition of rational choice theory. Most of the existing accounts of norms are *rational reconstructions* of the concept of social norm, which "specify in which sense one may say that norms are rational, or compliance with a norm is rational" (Bicchieri 2006, pp. 10-11).[2] My project is not intended to be a rational reconstruction. What sets my proposal apart is that it consists of a descriptive hypothesis, framed in terms of neural computations, about some core aspects of the mechanism of norm compliance. Hence, I am not concerned with

[2] A nice and important example of rational reconstruction is Bicchieri's (2006) account of norms. For Bicchieri, social norms should be understood in game-theoretical terms as Nash equilibria that result from transforming a mixed-motive game such as the prisoner's dilemma into a coordination game. The idea is that social norms solve social problems, in which each of the agents has a selfish interest to defect from the strategy that would provide the socially superior outcome if everybody followed it. When a social norm exists in problems of this sort, agents' preferences and beliefs will reflect the existence of this norm. Accordingly, the payoffs of the problem will change in such a way that agents playing the socially superior strategy will now play an optimal equilibrium.

normative questions such as "Under what conditions social norm compliance is rational?" My approach is not concerned with the content of the norms with which people ought to comply.[3]

Second, the Bayesian - RL approach I am advocating should be understood as a hypothesis about the functioning of the neural systems supporting social/moral cognition, rather than a proven solution to the task of complying with norms. Although a growing body of evidence from computational cognitive neuroscience strongly suggests that different perceptual systems (e.g. vision) might perform some form of Bayesian inference, and that multiple neural circuits (e.g. the basal ganglia) might implement some types of RL-algorithms, the Bayesian and RL views on neural functioning are not universally accepted (cf. Berridge 2007; Bowers and Davis 2012).

Finally, the framework I put forward is intended for social norm compliance, but it can be much more encompassing (cf. Clark 2013, Friston 2010). As mentioned above, what

[3] One of the consequences of my proposal is that agents can comply with "irrational" or even "immoral" norms indeed. If evolutionary pressure does not operate primarily over what is learned (the object of learning and decision-making), but over the learning and decision-making systems themselves (how such systems learn and make decisions), it is plausible that agents sometimes can learn and comply with norms that, in some sense, are "irrational" or even "immoral" (cf. e.g. Seymur, Yoshida and Dolan 2009). Consistent with this consequence is the view that there may well be no genetically-based special purpose neural network for social/moral learning and decision-making. The acquisition and implementation of specific norms would rather depend on "downstream ecological and epistemic engineering" (Sterelny 2003). The idea is that parental, upstream generations structure the downstream informational environment where the next generation develops so that the specific social norms embedded in that environment are more easily learnt and followed.

restricts my proposed account to social norms is the social nature of the representations that it posits. The main reason why the proposal is intended to be tailored specifically to social norms and social/moral cognition is that the time is ripe for making systematic, genuinely interdisciplinary, progress in the *science of norms*, by identifying the kinds of functions that brains need to compute to generate norm compliance. My hope is that a Bayesian – RL approach will at least highlight fruitful research directions in social/moral cognitive science.

Two Computational Problems for Social Cognition

Human agents live in a world populated by other people. We are bound to act in the presence of others. We are also bound to interact with others. The behavior of two or more agents can be said to be co-adaptive if it contributes to the agents' satisfying their desires, preferences, and needs in the environment in which they are embedded. Agents are best able to make plans and satisfy their desires when they are able to predict what their environment will be like over time. Since human agents are embedded in a social environment, they are best able to make plans and satisfy their desires when they are able to predict each other's behavior and changes in their social landscape.

One way in which agents can successfully make predictions of these kinds is by relying on prediction-errors. A *prediction-error* is the difference between an actual and an expected outcome (Niv and Schoenbaum 2008). It can be used to update expectations about what the future holds in order to make more accurate predictions, and, ultimately, to facilitate adaptive learning and decision-making. The amount of prediction-error in an agent's cognitive system can be understood as the agent's *uncertainty*. The less prediction-error an outcome brings about in the agent's cognitive system, the less uncertain is the agent about that outcome, and *vice versa* (cf. Friston 2010).

It is easier to make plans and satisfy one's desires when we are surrounded by agents who routinely engage in "normal," expected behavior. As various authors including Schotter (1981), Clark (1997, Ch. 9), Ross (2005, Ch. 6-7), and Smith (2008) have emphasized, social institutions can be understood as external "scaffolds" that constrain and channel people's behavior cueing specific types of cognitive routines and actions. While social institutions may facilitate the attainment of certain needs, desires, and goals, whether at the individual or at the social level, they contribute to "normalize" human behavior making it reliably predictable.

In the words of the anthropologist Mary Douglas:

"Institutional structures [can be seen as] forms of informational complexity. Past experience is encapsulated in an institution's rules, so that it acts as a guide to what to expect from the future. The more fully the institutions encode expectations, the more they put uncertainty under control, with the further effect that behavior tends to conform to the institutional matrix […]. They start with rules of thumb, and norms; eventually, they can end by storing all the useful information" (Douglas 1986, p. 48).

Social norms are instances of social institutions that act as guides "to what to expect from the future." Social norm compliance is one prominent class of "normal" behavior. By complying with social norms, agents reduce their uncertainty about the possible outcomes that social interaction can bring about. The more fully social norms are constituted by expectations, the more "they put uncertainty under control;" under the pressure of social norms, behavior tends to acquire distinct boundaries and "disorder and confusion disappear." Social norms would then be uncertainty-minimizing devices, and social norm compliance one of the tricks that we employ to interact co-adaptively and smoothly in our social environment. By complying with norms, agents minimize uncertainty over their social interactions; and by

minimizing uncertainty over their social interactions, agents' cognitive systems tend to become "models" of the social environment in which the agents are embedded. Thus, norm compliance contributes to make social environments transparent, with agent's meeting one another's expectations.

This last claim involves some important idealizations, however. There are at least four facts related to social norm compliance that contribute to make social environments opaque: normative pluralism, normative context-sensitivity, normative clash, and normative gradability. Any adequate explanatory framework for norms should have the conceptual resources to accommodate these facts, which I now briefly discuss.

First, many agents live in social environments that are not normatively uniform. Normative pluralism is ubiquitous: there are many different social norms governing a society, which may not be reducible to each other or to some "super" social norm. This plurality makes it very hard for agents to acquire a comprehensive model of a social environment, a model of all or even most of the norms that govern social interactions.

Furthermore, social norms are context-sensitive: norm compliance is conditional on having the right kind of representation of a context (cf. Bicchieri 2006, Ch. 2). Whether we have the right representation of a context—one that calls for norm compliance—depends on which situational cues are present in the context. However, there is no straightforward mapping between the situational cues in a given context and how agents represent that context; and there is no straightforward mapping between representing a context in a certain way and compliance with a norm. Different social norms may apply to the same type of context, and different types of contexts may be governed by the same type of social norm. In North America, for example, the social norm of tipping generally applies in restaurants and after taxi rides. But it does not apply at shoe shops or at most fast foods. The fact that service is especially good in a restaurant may cue diners in North America to give a generous tip to

the waitress or the waiter. But the same fact does not generally cue the same behavior in Japan. If a feature makes a given situation as one that calls for norm compliance, it does not follow that the feature always makes the same type of situation as one that calls for norm compliance. Whether a feature in a situation counts as a cue for norm compliance for an agent, and if so, what exact role it is playing there is sensitive to other features in that situation and to the learning trajectory of the agent. This makes it hard for agents to meeting one another's expectations in all contexts courtesy of social norm compliance.

The third qualification to the claim that norm compliance makes social environments transparent is that social norm compliance involves gradability. One gradable feature is the level of confidence that agents have that a specific social norm applies in a given context. For example, during a football match in Italy, people are more confident that a social norm applies that allows abusive chants than that a norm applies against littering. A second gradable feature is that the social norms with which agents comply are more or less stable in the face of new information. For example, in the face of incoming information, people's confidence that if somebody buys you a round of drinks at the pub, then you ought to buy the next round may be more stable than their confidence that one ought to orderly queue to get a drink at a bar in Australia. A third gradable feature is the degree of importance, or value that agents assign to a social norm in some situation. People, for example, can assign high value (or high importance) to addressing in a formal way a queen, but they can assign higher value to leaving a tip to waitresses at restaurants.

Finally, norms often conflict. For instance, traditional family norms often clash with wider social expectations: agents may regard themselves as having motives to comply with each of two norms, but complying with both norms is not possible. Thus, in the face of normative conflict, agents will breach somebody's expectations no matter what they do, which will contribute to make social environments opaque.

Now, with these qualifications in place, let us specify two of the major problems a computational system needs to solve in a social environment. Specifying these problems will help us identify the sort of neurocomputational mechanism that could enable biological, adaptive agents to acquire and act upon social norms so as to reduce their uncertainty. The two problems are:

(i) To use sensory information to compute representations of social situations.

(ii) To consume these representations to determine future movements, or internal changes, in the presence of, and interaction with, other people.

Problems (i) and (ii) are not specific to social cognition. In the domain of social interaction, however, they seem much harder to tackle, since living with other agents makes our surroundings more uncertain, complex, noisy, and ambiguous. But if problems (i) and (ii) are not specific to social cognition, then reliable computational solutions for perception and action can be extended to the domain of social interaction (cf. Behrens, Hunt and Rushworth 2009; Montague and Lohrenz 2007; Wolpert, Doya and Kawato 2003). My proposal follows exactly this strategy. Given the relationships between norm compliance, uncertainty, and prediction-error, my proposal embraces the prediction-error minimization approach, which has proved fruitful to tackle computational problems with respect to perception, learning, and action (e.g. Glimcher 2011; Rao et al. 2002).

The types of prediction-errors being minimized to solve those challenges are three:

- *sensory input* prediction-error,

- *reward* prediction-error,

- *state* prediction-error.

The first type of prediction-error enables agents to solve challenge (i); the last two types to solve challenge (ii).

Social Bayesian - RL Brains

I now introduce the main ingredients that enter a Bayesian – RL cooking recipe for social norm compliance. After these ingredients are defined at an abstract level, the familiar example of learning to comply with a norm of tipping at a restaurant will concretely illustrate some core aspects of the proposal. A more detailed discussion of how the Bayesian and RL components might interact concludes the section.

Social States and Agents' Hidden States

A *social state* is a set of social variables in a process that generates sensory input. Variables are *social* when they concern features of agents' interactions. Social states are highly structured, in that the variables constituting a social state can be correlated in complicated ways. The most important social feature is the hidden (mental) state of the other agents with whom we interact. The value of agents' hidden state both affects and is affected by the social contexts where the agents interact. Social *contexts* are sets of slowly and discretely changing parameters. These parameters comprise both slower changing variables in the internal state of agents and external variables such as features of the physical configuration of the external environment. Examples of these features are the physical arrangements of buildings and of their internal spaces. Churches, universities, cinemas, houses, parks are all examples of social contexts.

The hidden state of an agent is the most important social feature because it determines how that agent will interact with us, and how that agent will react to new sensory input. If we knew other agents' state, then we would have a model of their behavior. A model of their behavior would allow us to predict their reactions to inputs that we or the environment provide to their sensory systems. When other agents also have a model of our behavior, we

have a means to adjust our behavior to each other by predicting each other's reactions to new inputs (Wolpert et al. 2003).

However, we don't have direct access to other agents' state. Our cognitive systems need to infer it by relying on information about the social context and about other social variables like facial expression, hand gestures, posture, physical appearance, dress, speech, tone of voice, and so on. Relying on this type of information is necessary for our computationally-bounded cognitive system even if we had some direct access to other agents' internal state. Other agents' internal state, in fact, partly depends on their prior expectations about *our* state. During social interaction their behavior is both affecting and affected by our state. This would lead to an infinite hierarchy of priors in a computationally-unbounded agent. We are trying to infer another agent's state who is trying to infer our state: What I expect another agent's state is; what the other agent expects I expect about her state; what I expect another agent expects me to expect about her state, and so on. If we tried to infer other agents' states by using only information about mutual expectations about each other's state, then the infinity of priors about priors would make the computation of the state of the other agent unfeasible.

The approaches to this complexity can be twofold. On the one hand, our cognitive system can be thought as implementing finite rather than infinite prior hierarchies. There is evidence on strategic thinking in economic games suggesting that in fact people's hierarchy of priors about other agents' state comprises on average 1.5 levels (Camerer et al. 2004). On the other hand, our cognitive system can constrain inference about other agents' state by relying on learned correlations between certain external cues and the types of mental states normally entertained by agents in circumstances of a certain sort (e.g. "If the environment is dirty, then people are likely to feel disgust over there"). In this latter case, external social cues will function as proxies for the other agents' states.

Representations of external social cues need to be extracted from many modalities, integrated, and, at least initially, combined with our prior expectations about the other agent's state. After we acquire familiarity with the structure of the environment and the way in which such external cues correlate to different mental states of other agents, we need not rely on any prior expectation about other agents' priors anymore. The external cues will function as reliable proxies for knowledge about other agents' beliefs and motives. By forming accurate social representations from extensive interaction with certain types of external cues, we can arrive to reliably represent the hidden state of other agents *as though* we were trying to directly infer it. But in this case, our cognitive system does not need, in fact, to make inferences about the internal states of other agents. Other agents' hidden states would be already predicted by the social representation extracted from other relevant external cues.

Two ideas should be distinguished here. One idea is that constructing and using an inner model of other agents is sometimes less difficult than is generally supposed, courtesy of hierarchical and/or approximate algorithms. *Hierarchical* Bayesian, and RL algorithms offer one way—although certainly not the only way—to deal with domains such as the social one, that involves a large space of possible states and a large set of possible actions (see e.g. Botvnick, Niv and Barto 2009; Lee 2011). Furthermore, insofar as Bayesian or RL computations are intractable, many different *approximations*—including Monte Carlo and variational approximations—can replace exact inference in practice to account for the cognitive phenomena and behavior displayed by boundedly-rational social agents (Gershman and Daw 2012; Kwisthout and van Rooij 2013; Sanborn, Griffiths and Navarro 2010).[4]

A different idea is that very often we do not need an inner model to predict human actions. This idea resonates with a core insight in primatology as well as with some influential

[4] In what follows the shorthand 'Bayesian' refers to these types of tractable schemes.

philosophical work concerning understanding a language. With respect to the latter, Millikan (2005, Ch. 10), for example, argues that language understanding does not require mentalizing because it does not require grasping of speakers' intentions: understanding a language would be a form of direct perception of the world, instead of the speakers' intentions and thoughts. Millikan explains: "interpreting the meaning of what you hear through the medium of speech sounds that impinge on your ears is much like interpreting the meaning of what you see through the medium of light patterns that impinge on your eyes" (Millikan 2005, p. 205).

In primatology, a core insight is that similar behavior displayed by different species can be produced by very different mechanisms. Behavior-reading and mind-reading are two such mechanisms. While it might seem that complex social cognitive skills always depend on some understanding of what others believe, want, or know, this is in fact unnecessary. Several complex social behaviors can depend only on the information provided by overt behavioral cues of other agents (see e.g. Rosati and Hare 2010 for a concise recent review).

The two ideas just distinguished are not unique to the Bayesian – RL model I shall be drawing, but they cohere with it. And it is important to bring this fact into clearer focus because a Bayesian – RL model might appear, at first glance, to offer an implausible account of evolved, biological, social/moral intelligence. The model is in fact more plausible than it appears. While there are Bayesian schemes that underwrite the idea that acquiring and using an inner model of others is less hard than is supposed (e.g. Baker, Saxe ad Tenenbaum 2011; Hamlin et al. 2013; Yoshida et al. 2010), different types of RL algorithms, which our nervous systems might implement, suggest that social norm compliance may well be driven by both behavior- and mind-reading processes (cf. Daw et al. 2005; Dickinson and Balleine 2002).

Bayesian Social Representing

The first neurocomputational building block, which can carry out the task of computing social representations from sensory input, is a hierarchical Bayesian algorithm. According to the proposal on offer, the cortex learns and infers about the causes of sensory input by implementing Bayesian inference in a multistage processing hierarchy, which allows to incorporate statistical dependencies between stimulus representations at different levels of abstraction (cf. Lee and Mumford 2003; Friston 2008). The lowest level in the cognitive system would represent basic physical features like displacement, acceleration, mass, orientation, and wavelength that are combined into increasingly complex representations, up to higher levels that represent social states. When the value of the prior on state *Y* depends on other parameters *Z* at higher levels, given perceptual input S_x, the resulting posterior probability is computed with some suitable approximation of:

[1] $$Prob\,(Y, Z \mid S_x) \propto Prob\,(S_x \mid Y)\; Prob(Y \mid Z)\; Prob\,(Z)$$

The function that our cortex should compute is the posterior probability function *Prob* $[Z \mid S_x]$ of a high-level hidden state Z given sensory input S_x. For example, S_x may be the sensory input to the nervous system when an agent faces a social environment of a certain type and Z may be the representation, at the highest level, of the social role of a certain individual encountered over there ("That person is a waitress").

In order to carry out this computation, the cortical system reverses a *generative* (or forward) model, which describes the causal process that gives rise to data assigning a probability distribution to each step in the process. Given the generative model used by the cognitive system to determine how sensory input is generated, the system can infer the hidden state dependent on the sensory data by reversing the generative model. Lower-level representations are combined in a Bayesian fashion to compute more and more abstract representations at higher levels. The feedback, in the form of a sensory input prediction-error carried by forward connections in these hierarchical Bayesian model, provides a means to

incorporate statistical dependencies between representations at different levels of abstractions (e.g. "If this is a diner in New Jersey and that person is a waitress, then her hourly pay is likely to be relatively low"). The dependencies between the representations and their weight in the hierarchical Bayesian model will vary in function of one's personal learning trajectory. While we interact with other agents, our nervous system is constantly reorganizing, so that the models of the social environment it encodes get updated, and can serve as maps we can use to smoothly navigate the social world. Prediction-error minimization of sensory input can enable us to acquire social representations. But social representations, by themselves, do not motivate us to take action.

RL Social Norm Compliance

The second piece of neurocomputational machinery that could explain how social representations are transformed to enable us to engage in social norm compliance is the Reinforcement Learning (RL) account of the cortico-basal ganglia circuit (Sutton and Barto 1998; Niv 2009). RL neural computing bootstraps us into social behavior and culture by transforming social representations so as to determine future movements or internal changes in the presence of, and interaction with, other people (on the relevance of reinforcement learning to social behavior, including altruistic behavior, see Lee 2008; Seymour, Yoshida and Dolan 2009). When RL mechanisms tap into social representations that concern the hidden state of other people, then they enable us to learn to comply with social norms by minimizing social reward prediction-error.

Imagine that you arrive in some foreign country. You have certain beliefs (or priors) about how situations of type *Z* look like and about how people typically behave in *Z*: you have priors concerning a social state. In particular, you have a prior over the hidden state of other people in *Z*. Yet you are uncertain about what "grammar" governs situations of type *Z* in

that country, as you have a low degree of confidence about the mapping between sensory input and social representation of *Z* in that country, you are uncertain about the state transition $T(z, a, z')$, which determines how the environment evolves as you take actions, and you are uncertain about the reward structure of the environment $R: Z \times A \rightarrow \Re$, which determines the patterns of rewards/punishments you will incur by taking certain actions in certain states. If you want to interact adaptively with other people in that country in the environment *Z*, then you must learn and use the "grammar" people live by in *Z* in that country.

One task that your cognitive system should carry out in order to learn that "grammar" is to update your prior over the social environment *Z* in light of the information provided by the data generated by states in that environment. This task can be carried out courtesy of the Bayesian strategy sketched above. Suppose, for example, that you arrive to represent *Z* as a "diner" with high confidence. Your representation of *Z* correlates in specific ways to the hidden states of people who happen to be in *Z*. So, by relying on this representation, you expect that people in that environment have certain social roles in *Z*, which prompt them to behave in specific ways over there. By relying on your social representation of *Z*, you also expect that the environment has a certain causal structure. Because you are not confident about the state transition function, and about the reward contingencies in *Z* in that new country, you have to learn them if you wish to behave co-adaptively.

To learn these pieces of "social grammar" your cognitive system can rely on model-based and model-free RL algorithms, which have relatively clear neural implementations (Daw et al. 2005; Dayan 2008). These two types of algorithms differ in how they draw on experience to estimate quantities relevant to make choices and how they transform these quantities to reach a decision. *Model-based* algorithms draw on experience to build a model of the state transition and reward structure of the environment. These describe how different states of the environment, with their associated rewards, are connected to each other. Model-

based algorithms make choices by searching this model to find the most valuable action. Such a strategy is time-consuming and computationally costly, though it leads to more accurate choices.

In contrast, *model-free* algorithms draw on experience to learn action values directly, without building and searching any explicit model of the environment. What drives learning and action selection in model-free algorithms is a reward prediction-error, which reinforces successful actions without relying on explicit knowledge about state transitions or reward structure of the environment. This makes model-free computation more tractable, but less accurate than model-based strategies.

Depending on one's knowledge of the environment, cognitive resources, and time constraints, adaptive behavior can be best served by model-based or model-free algorithms. Given limited experience with that new environment, behavior may reflect model-based processes (Gläscher et al. 2010). Initially, you rely completely on an estimated model of the environment of the form:

[2] *Prob* (*new state* | *state*, *action*).

This estimated model of the environment can be constrained with information about the structure associated with your social representation of *Z*. Using this model you can perform a simulation of the consequences of your actions given current state *z*: If you take action a_t from current state *z*, then it's likely that you will end up in state *z'*. You utilize experience with state transitions to update an estimated state transition function $T(z, a, z')$. Upon each of your choices, a *state prediction-error* is computed:

[3] $\delta_{spe} = 1 - T(z, a, z')$.

This state prediction-error is used to update the probability of the observed transition thus:

[4] $T(z, a, z') = T(z, a, z') + \eta\delta_{spe}$

where η is a learning rate parameter.

Behavior shaped by model-based computation reflects a goal-directed process in which a particular desired outcome, like having a good meal and avoiding frictions with other people in *Z*, is used to flexibly determine any complex sequence of actions needed to achieve it. Action selection is carried out by searching your model of the environment: you work out the consequences of each action available to you in *z*, and select the action that is more likely to lead you towards your desired outcome. This allows action selection to be sensitive to changes in the structure of the environment and in your motivational state. If, for example, you notice that people suddenly react differently than usual given *z*, or your motivational state is abnormal, you can immediately adjust your behavior accordingly.

Searching and updating your "map" of the environment is computationally demanding both for working memory and for your "mentalizing competence." You need to remember situations you encountered in the past similar to the one at hand, you need to work out what other people's expectations may be, you have to consider many different actions and outcomes, and work out which is the best to achieve your goals. This can reduce the capacity for alternative computations and the smoothness of interaction, as the model-based computation would engage valuable cognitive resources to identify which action you should implement given your state and your goals. By relying on a model-based controller, learning and complying with social norms can be effortful and time consuming.

One crucial aspect of social-decision problems is that they typically recur. So, with more experience with situation *Z* in that country, you need not to rely on the model-based strategy. After you have regularly encountered situations of type *Z*, the sensory data generated by *Z* have led your representation of *Z* to be more and more accurate. Your prior about the structure of that environment can impose further constraints on the state and action space, on which your learning and decision-making systems tap. You may rely on model-free

computation which drives learning and decision making by means of social reward prediction-errors.

A *reward* prediction-error is a difference between two values associated with executing actions in some state. The *value* of a state is the expected sum of future rewards and punishments that can be achieved starting to act from that state. In general, rewards can be understood as stimuli, objects or states that make us come back for more. Punishments, conversely, can be understood as stimuli, objects or states that make us *not* come back for more (Schultz 2007). Although the distinction between social and non-social rewards is not a sharp one, a social reward can be defined as a stimulus provided by another individual of the same or some other closely related species by means of some movement, sound, utterance, gesture, posture, or facial expression. Consistently with this definition, examples of non-social rewards can be money, food, water, and a variety of other inanimate objects and signs.

By picking up on social rewards, you acquire ways of evaluating or predicting the long-term consequences associated with executing a particular action in a social context. You need not mentalize with others or search any map of the environment in order to comply with a norm. You can come to comply with social norms automatically, quickly and at little computational costs. It is important to emphasize, however, that the shift from model-based to model-free control is not sequential nor instantaneous, but highly parallel and dynamic. The early phase of model-free learning processes take place while behavior still appear to be controlled by a model-based strategy (Tricomi et al. 2009). Moreover, model-based and model-free computations in the human brain may not be neatly separated (Daw et al. 2011), which suggests significant cross-talk and interaction between different types of RL algorithms that might be implemented by neural activity.

Nonetheless, model-free algorithms seem to operate most effectively, with little computational demands, in familiar situations. They operate on "cached" values, which store

experience about the overall future worth of a particular action. Such values can be used to implement certain behavioral responses in the face of stimuli that were consistently associated to a rewarding outcome in the past. Given reliable co-variation between situational cues and certain behavioral patterns of people in *Z*, the reward values of the behavioral responses become conditioned onto the cues. Features of the environment become to encode information about the reward structure of the environment, and you can outsource behavioral control on them. The cues present in the environment signal opportunities to perform particular "rewarding" actions. In this way, as your training with social situation *Z* proceeds, goal-directed behavior becomes habitual and cue-driven. The representation of *Z* itself can drive behavior with no need to work out what other people expect you to do in *Z* or to keep track of state transitions underlying *Z*. Features of *Z*, that is, acquire the capacity to motivate you to directly act upon your social representation of *Z*.

Norm compliance in this case becomes a perceptually-based, habit. And social interaction becomes a fluid, context-specific, inferential response to incoming sensory inputs and their values. It enables co-adaptive, smooth interaction without access to hidden states of other agents in the world. Insofar as other accounts of social behavior entail that the preference to comply with social norms is always dependent on mutual expectations, then this theoretical proposal makes a novel, testable prediction: norm compliance can become a habit that involves no mentalizing.

A Social Game of Tipping

How model-based and model-free RL algorithms can solve the problem of learning a social norm is best illustrated by examining a specific case. Imagine once again that you are visiting for a certain number of days a foreign country, about which you know little. Let us assume that, courtesy of Bayesian neural computing, you represent a given social environment in that

country as a restaurant, and that you intend to repeatedly dine at that restaurant. There may be some social norm of tipping in restaurants in that country, but you are not sure. You want to learn this social norm so as to display adaptive behavior in the social environment you are navigating.

Now, the structure of a learning problem such as this one can be simplified thus.[5] Each time you enter the restaurant (call this, state *s*1), you must choose either of two tables L or R. While Miss L waits table L generally providing either excellent (*s*2) or good service (*s*3), table R is waited by Miss R, who typically provides average (*s*4) or bad service (*s*5). After your dinner is over, you must pay your bill and decide which amount to tip. After your decision, the person who waited your table collects your tip and says goodbye to you. The goodbye can be uttered with either an angry or a cheerful tone of voice, accompanied respectively by either an angry or a happy facial expression. Such emotional reactions can be understood as positive or negative social reward outcomes, which you can use to learn how much is the social norm of tipping over there.

So, starting from state *s*1 you can take either of two actions (i.e. L or R). If you take action L, then with a certain probability you will be in state *s*2, or *s*3. If you take action R, then with a certain probability you will be in state *s*4, or *s*5. At this point, you have several different actions available corresponding to different amounts you may leave as a tip—money is limited and important to you, so the number of actions you are willing to take is bounded. Finally, a reward outcome is revealed to you (i.e. a positive or negative social reaction),

[5] An experimental task of this type has been used by Colombo, Stankevicius and Seriès (ms) to address how social rewards (e.g. facial expressions), in comparison to non-social rewards (e.g. conventional feedback marks such as ticks and crosses), impact learning performance and decision-making.

which may depend stochastically on the underlying social norm you are learning and on the pair (service quality you received- amount you tipped).

Model-based strategies enable agents to learn the social norm of tipping by building and relying on an estimated map of the environment (i.e. a state–action–outcome tree). State prediction errors will help acquire and update such a map. Decisions are then made on the basis of this map, by searching through it and finding the path with the highest overall value. For example, after some experience, you may learn that the path [enter the restaurant-choose table L-receive good service-tip 15% of the bill] has highest overall value, and is conducive to adaptive social behavior.

Which path has the highest overall value for you is determined by the amount of money you have in your pocket, and how much you care about your money and about the emotional reaction you receive. All these factors have some impact on how quickly (i.e. after how many meals in that restaurant) you will behave adaptively in that environment, thereby learning the social norm of tipping. The way state prediction errors are computed "on the fly" allows you to be sensitive to changed circumstances; it allows, for example, that your own current motivational state (e.g. you are especially cheerful today) may lead you to re-valuate different decision paths.

Model-based strategies enable agents to learn the social norm of tipping directly, by trial-and-error, without acquiring an explicit map of the environment. These strategies rely "cached," learned, values for every action available to you at every state. After several meals at that restaurant, each (state-action) pair is associated with a value summarising the expected amount of social rewards that you are likely to obtain by taking a certain action in a given state. Action selection simply involves choosing the action with the highest cached value at the current state; for example, given normal service, the action with the highest value is tipping 10% of the bill. Relying on cached values is computationally simple, as it does not

require an explicit estimate of the probabilities that govern state transitions or a search of all possible paths in the environment. This, however, comes at the cost of less flexibility. Values do not immediately change with changes in, for example, your motivational state: your being especially grumpy today will not make any difference.

Finally, besides these two learning strategies, verbal instruction is obviously an incredibly efficient means to learn how to navigate the social environment. Recent computational and neuroimaging work indicates that verbal information can have significant impact on reward-based learning (e.g. Doll et al. 2009; Li et al. 2011). Although it remains unclear how exactly, and under which circumstances, verbal instructions influence learning and social behavior, we may assign less weight to observed feedback when reliable verbal instructions are available, which can spare us multiple errors, and learn more quickly.[6]

Interaction between Bayesian and RL components

The full neurocomputational account of social norm compliance is not so simple as the proposal thus far may have suggested. Bayesian and RL components have been treated separately, as if there is no rich dynamical interaction between Bayesian and RL-systems. More plausibly, Bayesian and RL processing are intimately related. Evidence indicates that

[6] Doll *et al.* (2009) have developed two neurocomputational models that could explain the precise effect of verbal information on reward learning: an 'override' and a 'bias model'. In the first, the striatum—a subcortical brain region and major target of dopaminergic neurons—learns cue-reward probabilities as experienced, but is overridden by the prefrontal cortex—where instructed information would be encoded—at the level of the decision output. In the bias model action selection and learning supported by the striatum are biased by rules and instructions encoded in the prefrontal cortex.

these two kinds of systems display dynamical interaction, but also that a separation between probabilistic inference and value might not be empirically adequate (Gershman and Daw 2012, Sec. 3.3; see also Sec. 5 for two types of proposals on how the segregation between perception and action might be weakened or even abandoned).

Now I outline one way in which Bayesian and RL systems might interact in producing social norm compliance. The basic hypotheses are twofold: On the one hand, RL systems piggyback on top of the Bayesian system. On the other hand, reward modulates Bayesian inference at all levels of sensory processing. Let me explain.

What could it mean that RL systems piggyback on Bayesian inference? The idea is not only that the Bayesian system computes social representations that feed into RL processing, but Bayesian computing is also intimately involved in RL model construction and action selection. Specifically, learning a social norm and norm compliance might be performed by computing (and continuously updating) a posterior distribution over RL state transitions models, reward probabilities, and value functions, based on sensory input and the history of past state-action pairs. Algorithms that maintain a distribution over state transitions and the reward structure of the environment are model-based Bayesian RL algorithms. Algorithms that maintain a distribution over mappings from state to actions, or over state-values are model-free Bayesian RL algorithms that do not store any map of the environment. Courtesy of these two types of algorithms, uncertainty over the ingredients of RL computing are fully captured, which could facilitate agents to make more informed decisions, while learning social norms more quickly.

Consider model-based Bayesian RL. Agents start with a prior distribution over different state transition models (i.e. structures) of the environment, $T(z, a, z')$. Examples of different state transition models of a social environment like the one described in the previous section are: *restaurant* ($s1$)-*take table on the left* (a_L)-*excellent service* ($s2$), and *restaurant*

(*s*1)-*take table on the left* (a_L)-*bad service* (*s*5). The prior represents the initial belief of the learner about the structure of the social environment (Gershman and Niv 2010; see also Tenenbaum et al. 2011 on Bayesian inference over structures). Agents update this belief by implementing Bayesian inference based on two signals: sensory input, and observed state-action-state' triples. Beliefs about states of the world are updated given sensory input. Beliefs about state transitions of an environment are updated given state-action-state' sequences.

What could it mean that reward modulates Bayesian inference at all levels of sensory processing? The idea is that representations of social states are always imbued with reward value, which is supported by evidence on how the acquisition and representation of incoming sensory information in the human visual cortex is influenced by e.g. the reward history of a state (e.g. Serences 2008). This reward-modulation of sensory processing can dramatically constrain the dimensionality of the space of *relevant* hypotheses about states and structures of the environment. It is unnecessarily complex and costly if our neurocomputational systems stored and computed a lot of information of little relevance to the agent. So, Bayesian and RL systems would learn and make inference only over reward-relevant representations, representations relevant to adaptive social interaction.

A study investigating the responses of auditory neurons in grasshoppers underwrites this conclusion (Machens et al. 2005). It was found that primary auditory receptors in grasshoppers are not equally sensitive to different auditory stimuli equally frequent in grasshoppers' natural environment. These neurons seem to maximize the information gained about specific, but much less frequent stimuli, namely mating signals stimuli. Hence, the processes carried out by sensory neurons might not be always matched to the statistics of the environment. More plausibly, these processes might be tuned to a "weighted ensemble of natural stimuli, where the different behavioral relevance [i.e. reward value] of stimuli determines their relative weight in the ensemble" (Ibid., p. 454). The different social

relevance of different stimuli—determined by reward-based RL computing—appears to constrain the workings of Bayesian inference. So, our perception of, and expectations about, the social world are channelled by our conative processes or states about it, which are molded, in turn, by the inferences we make based on sensory input.

Conclusions

Our acquisition of the grammar that governs social situations can be driven by minimization of three types of prediction-errors computed by our nervous system. First, a sensory prediction-error that is produced and minimized by Bayesian algorithms, which give rise to social representations, running on hierarchically organized cerebral cortex. Second, a state prediction-error that is produced and minimized by model-based RL algorithms. Third, a social reward prediction-error that is produced and minimized by model-free RL algorithms running on midbrain, dopamine-based circuits. These two RL strategies enable us to act on our social representations so that we comply with social norms. By working in concert, such Bayesian-RL neurocomputational system ensures that our predictions about people's behavior become self-fulfilling prophecies. Our complying with norms is one trick to make these predictions come true in social environments. It ensures that our prior expectations about social sensory input are met and social uncertainty avoided.

Acknowledgements

References

Baker CL., Saxe RR, Tenenbaum, JB (2011) Bayesian theory of mind: Modeling joint belief–desire attribution. Proceedings of the Thirty-Third Annual Conference of the Cognitive Science Society, pp 2469–2474

Behrens TE, Hunt LT, Rushworth MF (2009) The computation of social behavior. Science 324:1160-1164

Berridge KC (2007) The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology 191:391-431

Bicchieri C (2006) The Grammar of Society: The Nature and Dynamics of Social Norms. Cambridge University Press, New York

Binmore K (1994). Game Theory and the Social Contract, Vol. I. Playing Fair. MIT Press, Cambridge MA

Botvnick MM, Niv Y, Barto A (2009) Hierarchically organized behavior and its neural foundations: A reinforcement-learning perspective. Cognition 113:262-280

Boyd R, Richerson P (2001). Norms and bounded rationality. In *Bounded rationality: the adaptive toolbox* Gigerenzer G, Selten R 2001pp. 281–296. Eds. Cambridge, MA:MIT Press.

Bowers JS, Davis, CJ (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin,* 138, 389-414.

Clark A (1997) *Being There*. Cambridge, MA: MIT Press.

Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci,* 36, 181–253

Camerer, C., Teck Hua Ho, T., and Chong, K. (2004). A Cognitive Hierarchy Model of One-Shot Games. *Quarterly Journal of Economics*, 119, 861-898.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors." *Neuron*, 69, 1204–1215

Daw, N.D., Niv, Y., and Dayan, P. (2005). "Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control." *Nature Neuroscience*, 8, 1704–1711.

Dayan P (2008). The role of value systems in decision making. In Engel C & Singer W, editors, *Better than Conscious? Decision Making, the Human Mind, and Implications for Institutions* Frankfurt, Germany: MIT Press, 51-70.

Dickinson A, Balleine BW (2002). The role of learning in the operation of motivational systems. In CR Gallistel (Ed.), *Learning, motivation and emotion Vol. 3*. New York: John Wiley & Sons, 497-533.

Doll BB, Jacobs WJ, Sanfey AG, Frank MJ (2009) Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. *Brain Res* 1299:74–94.

Douglas M (1986). *How Institutions Think*. New York: Syracuse University Press.

Elster, J. (1989). "Social norms and economic theory." *Journal of Economic Perspectives*, 3, 99-117.

Friston, K. (2010). "The free-energy principle: a unified brain theory?" *Nature Review Neuroscience*, 11, 127-138.

Friston, K. (2008). "Hierarchical models in the brain." *PLOS Computational Biology*, 4, e1000211.

Gershman, S.J., and Daw, N.D. (2012). "Perception, action and utility: the tangled skein." In M. Rabinovich, K. Friston, and P. Varona (Eds.), *Principles of Brain Dynamics: Global State Interactions*, pp. 293-312. Cambridge, MA: MIT Press.

Gershman SJ, and Niv Y (2010) Learning latent structure: Carving nature at its joints - *Current Opinion in Neurobiology* 20(2), 251-256.

Gintis, H. (2010). "Social norms as choreography." *Politics, Philosophy and Economics*, 9, 251-264.

Gläscher, J., Daw, N., Dayan, P., and O'Doherty, J.P. (2010). "States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning." *Neuron*, 66, 585–595.

Glimcher, P.W. (2011). “Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis.” *Proceeding of the National Academy of Science USA*, 108, 15647–15654.

Hamlin, J.K, Ullman, T.D., Tenenbaum, J.B., Goodman, N.D., & Baker C.L. (2013). “The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model.” *Developmental Science*, 16:2, 209-226.

Kwisthout, J. and van Rooij, I. (2013). “Bridging the gap between theory and practice of approximate Bayesian inference.” *Cognitive Systems Research*, 24, 2-8.

Lee D. (2008). “Game theory and neural basis of social decision making.” *Nature Neuroscience*, 11, 404–409.

Lee, M.D. (2011). “How cognitive modeling can benefit from hierarchical Bayesian models.” *Journal of Mathematical Psychology*, 55, 1-7.

Lee TS, Mumford, D (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America, A*, 20, 1434-1448.

Li J, Delgado MR, Phelps EA (2011) How instructed knowledge modulates the neural systems of reward learning. Proc Natl Acad Sci USA 108:55–60

Lewis, D.K. (1969). *Convention: A Philosophical Study*. Cambridge MA: Harvard University Press.

Millikan, RG (2005). *Language: A Biological Model*. New York: Oxford University Press.

Machens C, Gollisch T, Kolesnikova, O, and Herz, A. (2005). Testing the efficiency of sensory coding with optimal stimulus ensembles. *Neuron*, 47(3):447-456.

Montague PR, Lohrenz T (2007) To detect and correct: norm violations and their enforcement. *Neuron* 56:14–18.

Niv, Y. (2009). Reinforcement Learning in the brain. *Journal of Mathematical Psychology*, 53, 139–154.

Niv, Y., and Schoenbaum, G. (2008). “Dialogues on prediction errors.” *Trends in Cognitive Science*, 12, 265–272.

Pettit, P. (1990). *Virtus Normativa:* Rational Choice Perspectives, *Ethics* 100, 725-55.

Rao, R.P.N., Olshausen, B. and Lewicki M. (eds.) (2002) *Probabilistic Models of the Brain: Perception and Neural Function*. Cambridge, MA: MIT Press.

Rosati, A.G. & Hare, B. (2010). “Social cognition: From behavior-reading to mind-reading.” In: *The Encyclopedia of Behavioral Neuroscience*, G. Koob, R. F. Thompson, and M. Le Moal (Eds.). Elsevier pp. 263-268.

Ross, D. (2005). *Economic Theory and Cognitive Science: Microexplanation*. Cambridge, MA: MIT Press.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). “Rational approximations to rational models: Alternative algorithms for category learning.” *Psychological Review*, 117(4), 1144–1167.

Schotter, A. (1981). *The economic theory of social institutions*. Cambridge, MA: Cambridge University Press.

Schultz, W. (2007b). “Reward.” *Scholarpedia*, 2(3):1652. URL = <http://www.scholarpedia.org/article/Reward>.

Serences, J. (2008). Value-based modulations in human visual cortex. *Neuron*, 60(6):1169-1181.

Seymour B, Yoshida W, Dolan R (2009) Altruistic learning. Frontiers in Behavioral Neuroscience **3**:23. doi: 10.3389/neuro.08.023.2009

Smith V (2007) Rationality in Economics: Constructivist and Ecological Forms. Cambridge University Press, New York.

Sripada CS, Stich S (2007) A framework for the psychology of moral norms. In: Carruthers P, Laurence S, Stich S (eds) Innateness and the structure of the mind, vol II. Oxford University Press, London, pp 280–302

Sterelny K (2003). Thought in a Hostile World: The Evolution of Human Cognition. Oxford: Blackwell.

Sugden R. (1986) *The Economics of Rights, Cooperation and Welfare*. Oxford: Blackwell.

Sutton, R.S., and Barto, A.G. (1998). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Tenenbaum JB, Kemp, C, Griffiths TL, and Goodman ND (2011). How to grow a mind: statistics, structure and abstraction. *Science*, 331, 1279-1285

Tricomi E, Balleine B, O'Doherty J (2009) A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci 29:2225–2232

Ullmann-Margalit E (1977) The Emergence of Norms. Oxford University Press, Oxford

Wolpert DM, Doya K, Kawato M (2003) A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 358:593-602

Yoshida W, Seymour B, Friston KJ, Dolan RJ (2010) Neural mechanisms of belief inference during cooperative games. J Neurosci 30:10744-10751