Representational exchange in human social learning: Learning policies, goals, and beliefs

Charley M. Wu[1], Natalia Vélez[2], & Fiery A. Cushman[2]
[1] Human and Machine Cognition Lab, University of Tübingen
[2] Department of Psychology, Harvard University

Abstract

What makes human social learning so powerful? Past accounts have approached this question by crediting a singular capacity—such as high-fidelity imitation or the ability to infer others' beliefs—for the whole success of human social learning. Here, we propose that the answer lies not within a single capacity, but in our ability to flexibly arbitrate between different computations and to integrate their outputs. In particular, learners can infer a demonstrator's model of the world (*belief inference*), assume their goals and values within a fixed world model (*value inference*), or directly copy their actions in the absence of causal insight (*policy imitation*). Each of these inferences trades off the cost of computation against the flexibility and compositionality of its outputs. Crucially, however, we have the capacity to arbitrate and exchange information between these representational formats. Human social learning, we suggest, is powerful not just because of the way it moves information between minds, but also because of the way it flexibly moves information within them.

Many animals can learn from each other, but not like us. If chimpanzee social learning were a simple tune carried by a single voice, ours would be the exuberant chorus of a 12-piece ragtime marching band. Naturally, anybody who wants to understand human learning and behavior must confront a central question: Why do they sing, while we *swing*?

Many researchers have approached this question by pinpointing a singular feature of human social cognition that distinguishes us from other animals. In essence, they have sought out the "small difference that […] made a big difference" (Tomasello et al., 2005). To some, human social learning is powerful because we are particularly disposed to high-fidelity imitation, granting us the capability to transmit and innovate upon cultural knowledge over multiple generations (Boyd & Richerson, 1988; Henrich, 2017; Tennie et al., 2009). When learning to bake a loaf of bread, for instance, we might imitate the exact motions and choices of a master baker. We might do this even if we cannot understand the rationale behind these motions and choices (Lyons et al., 2007)—indeed, the demonstrator might also be unaware of the rationale, having inherited some techniques through cultural transmission (Derex et al., 2019; Henrich, 2017).

To others, human social learning is powerful precisely because we are not limited to understanding others' behavior at the level of observable actions. Rather, we have the ability to draw rich social inferences about unobservable mental states, such as the goals, beliefs, and values that we impute to other people (Apperly, 2010; Gweon, 2021; Jara-Ettinger, 2019; Tomasello et al., 2005). These inferences allow us to copy not just actions, but also goals and beliefs that can be re-assembled productively into new behaviors. On this view, we rarely copy the concrete behaviors of each other with high fidelity; what we transmit instead is relatively abstract knowledge and values. When learning to bake a loaf of bread we would acquire knowledge such as the leavening power of yeast and extensibility of well-developed gluten, and emulate goals such as achieving airy expansion of the loaf in the oven.

These are both plausible candidates. Compared to other primates, humans are more disposed to high-fidelity imitation (Horner & Whiten, 2005; McGuigan et al., 2007) and have more sophisticated capacities for mental-state inference (Herrmann et al., 2007; Tomasello et al., 2005). Thus, the debate cannot possibly be about whether we're capable of one or the other, but rather must be about which carries the melody, and which merely adds embellishment.

We suggest this debate is ill posed. Might the power of human social learning come not from a single instrument, but from our ability to harmonize? We don't mean this in the banal sense that you can blow harder on two horns than one, but in the deeper sense that music resides not in individual notes, but in the ways we compose them together.

We argue that much of the power of human social learning depends on our ability to arbitrate between, and integrate across, distinct forms of social learning. This argument is structured around an analogy to *non*-social learning. Here, too, we have a variety of instruments at our disposal. We can use simple cached action values (i.e., repeating actions that have previously been rewarded), which are computationally cheap but lack generalizability. Or we can leverage a model of the world and engage in flexible, productive planning—a strategy that achieves generalizability and compositionality, but at increased computational costs. Decades of psychological theorizing were spent arguing over which of these things humans do, or which is more important (e.g., Skinner, 1950; Tolman, 1948).

Today, however, the most exciting work in learning and decision-making explores the way in which humans can integrate both of these methods within a single task, drawing on efficient heuristics when possible and flexible planning when necessary (Cushman & Morris, 2015; Huys et al., 2015; Keramati et al., 2016; Kool et al., 2018; Russek et al., 2017). A baker, for instance, relies at times on skills "in the hands"—a certain way of shaping a loaf, for instance, that has solidified into habit through countless hours of practice. At other times, she relies on knowledge and goals that allow her to adapt to variations in the ingredients, humidity, temperature, and so forth—or to plan out an entirely new recipe. A debate over whether “habit” or “planning” is more important misses the point: what is most remarkable is her ability to compose these elements into a whole that is both practiced and productive; both efficient and flexible.

We argue that this same structure arises in social learning. High-fidelity imitation is a close homolog to certain forms of cheap-but-inflexible learning and decision-making, such as habit. Meanwhile, mentalizing and emulation are a close homolog to goal-directed planning. The first part of this paper describes these homologies in detail. Just as a baker prepares her loaf in a way that integrates across diverse processes of decision-making, the baker’s apprentice faces the task of learning representations at multiple levels, from concrete actions to abstract knowledge and goals. To be successful, she must learn at multiple levels of representation, arbitrating between imitation and emulation for each part of the bread-making process. Next, she must integrate these learned representations with her own pre-existing knowledge, skills and goals, which requires integrating information across levels of representation. The heart of our paper addresses the processes of arbitration and integration, which we take to be essential to the power of human social learning.

For simplicity, we focus on observational learning, where we directly perceive a person’s action and the setting in which they are performing it, but without direct access to the underlying causes of their action. We will also assume a relatively naïve observer is watching a relatively experienced expert, and attempts to learn how to act in more expert ways themselves (like baker and her apprentice). Clearly human social learning involves much more than this specific kind of observational learning, however. It also involves teaching, talking, rewarding and punishing, and much more. Ideally, our case study of observational learning might inform theories of human social learning more broadly.

Mechanisms of social learning and decision-making

Theories of observational social learning have been broadly divided into two approaches: Imitation and emulation. With imitation, the learner copies the observed behavior more-or-less directly (Bandura, 1962; Heyes, 2001; N. E. Miller & Dollard, 1941). With emulation, the learner decomposes the observed behavior into a set of primitives, such as the beliefs, goals or preferences of the other individual, and then integrates these into her own subsequent planning (Tomasello et al., 1987; Whiten & Ham, 1992).

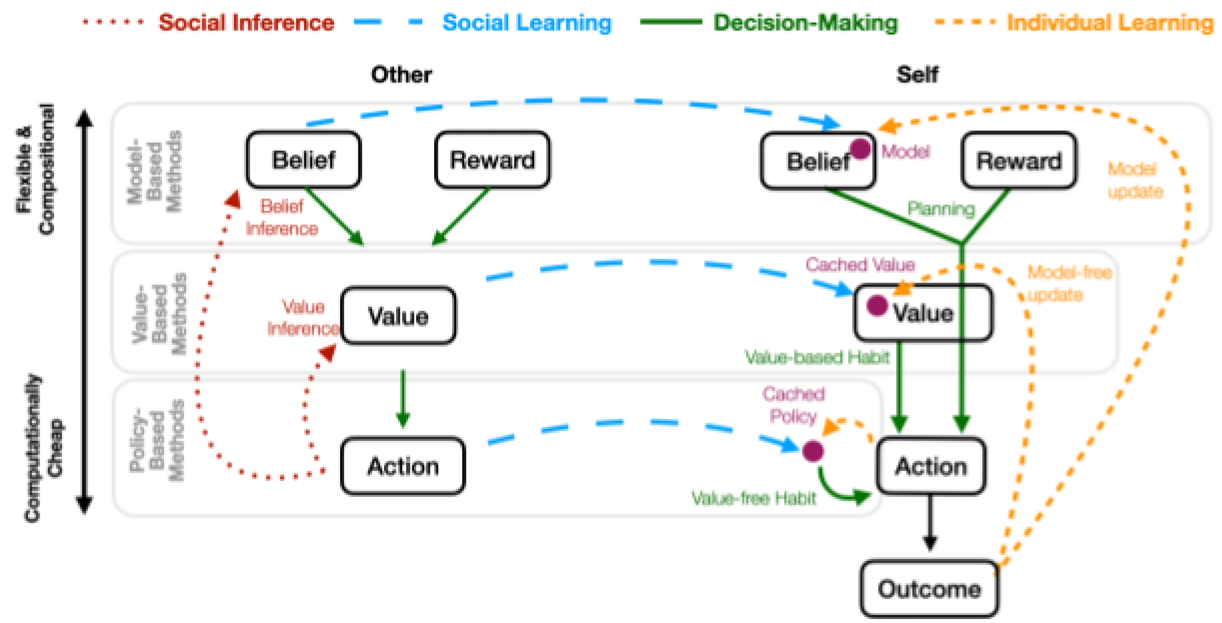

Figure 1. Different forms of social and individual learning share a similar trade-off involving computational complexity and flexibility, which both increase as we ascend the decision-making hierarchy (green solid arrows). In individual learning (orange short dashed arrows), we can update a cached policy, a cached value representation, or our model of the environment. Similarly, different forms of social learning (blue long dashed arrows) draw upon different components of observed behavior: Directly adopting the policy of another agent, adopting their inferred value representations, or adopting their inferred beliefs about the environment. Value-based and model-based methods of social learning require inferring hidden mental states from observed actions (red dotted arrows), which incur added computational costs. However, they also afford increased flexibility, since learning from these primitives of the decision-making process allow us to deploy them in a adaptive and compositional fashion.

The choice between imitation and emulation resembles a more general choice that arises in individual learning and decision-making. When deciding how to act in an non-social situation, a person can rely on any of several strategies: Repeating actions they tended to perform in the past (i.e., drawing on a *cached policy*), repeating the actions that tended to be rewarded in the past (i.e., drawing on a *cached value*), or selecting the actions they expect to produce the best outcomes based on a model of the environment (i.e., *model-based* planning).

These three forms of individual learning are illustrated on the right hand side of Figure 1. After performing an action and observing the outcome (i.e., reward), a person can update any (or all) of these representations. She can update a cached representation of her policy, a cached representation of the value of the action (on the basis of any reinforcement she receives), or her model of the environment and its dynamics (based on the outcomes she observes following her action). Updates higher up in the decision-making process afford greater flexibility, but incur more computational costs.

Similarly, we can think of different social learning strategies as spanning an analogous trade-off involving flexibility and complexity. When observing another's actions, one can learn from and represent, among other things, (1) the action itself, (2) the inferred goal and value representations underlying the action, and (3) the inferred beliefs of the other. Cognitive demands increase along this spectrum, but so do the properties of flexibility and compositionality.

Policy imitation

In the simplest case, observing a behavior by another individual will often make us more likely to also adopt it. In other words, social learning often involves action imitation (Heyes, 2002; Hoppitt & Laland, 2013; Legare & Nielsen, 2015; Whiten & Ham, 1992). You might simply choose a restaurant based on its popularity (conformity-biased imitation; Boyd & Richerson, 1988), or you might be more selective and follow the recommendations of a prestigious food critic, deferring to their expertise (prestige-biased imitation; Henrich & Gil-White, 2001; Jiménez & Mesoudi, 2019).

Imitation is computationally cheaper and simpler compared to other forms of social learning, since no inference is needed about the other person's goals or intentions. Rather, their behavioral policy can simply be copied verbatim. Thus, a useful analogy can be made to individual learning, where we often deploy a "cached policy" in the form of habitual behaviors (Cushman & Morris, 2015; Daw et al., 2005; Gershman, 2020). For instance, we can take our usual route to work in the morning or order our usual meal at our favorite restaurant, without having to iterate over possible plans or weigh the benefits of different options each time. Sometimes referred to as "amortization" (Dasgupta et al., 2018), we can simply avoid costly computations by caching and redeploying solutions that have worked in the past. Paired with an appropriate rate of even random variability (akin to "mutations") in the copying process, imitation has been proposed to be a key driver of human cultural evolution (Henrich, 2017; Heyes, 2018; Tennie et al., 2009).

In exchange for its simplicity, pure imitation lacks flexibility. The behavioral policies of other people can be informed by different skills, preferences, or goals, imposing limitations on how adaptive an imitated policy will be in different settings. Copying the moves of a professional skier might get you killed or severely injured. Following a meat eater's restaurant recommendation might leave you hungry if you're a vegetarian. And following a random car when you are lost certainly won't help you navigate to your destination. Thus, compared to more sophisticated forms of social learning, imitation generalizes poorly and lacks compositionality, since one cannot pluck an action out of its original context and expect it to be useful when placed in a new situation.

Value Inference

Another form of observational learning is to update one's value representation in response to social information. This is sometimes known as "value shaping" (Galef, 1988; Ho et al., 2017; Najar et al., 2020), where socially observed actions enhance our value representations of a choice or stimulus. This is a more sophisticated and computationally complex strategy than policy imitation, but affords increased flexibility and compositionality.

Imagine you see someone eat a novel food called a "dax". Policy imitation dictates you should also eat daxes in the future. But value inference is more flexible. If you observed someone eating daxes when they could have chosen beluga caviar or wagyu beef, you might infer that daxes are very high in value. But not so if you see a very hungry person choose daxes over stale bread or a dirty hotdog. Thus, by assigning value to daxes (rather than simply imitating dax-eating), you can make more flexible decisions about whether and when to eat one yourself. Social value inference also allows you to learn by watching somebody *not* choose a "dax" in the presence of high- or low-value alternatives, or by overhearing somebody describe how much they like or dislike daxes. This information can be used to update a value representation, which can be deployed later during decision-making.

Inferring values from social observation is closely related to the well-established use of cached value representations or value-based habits in individual learning (Daw et al., 2005; Keramati et al., 2016; Kool et al., 2018; Solway & Botvinick, 2012). Like the use of a cached policy, caching value representations allows an agent to re-use costly computations, although this process takes place one step higher up the decision-making hierarchy (Fig. 1). While caching value representations is not as cheap as caching policies (K. J. Miller et al., 2019), it offers superior generalizability and flexibility. This is because, as we have described, representations of value allow us to make comparisons across novel sets of choice alternatives.

Importantly, cached value can also be assigned to goal representations in a nested hierarchy (Botvinick & Weinstein, 2014; Cushman & Morris, 2015; Keramati et al., 2016; Maisto et al., 2019). For instance, the superordinate goal "make coffee in the morning" can be assigned value, but can also induce a set of value representations over subordinate goals: When valuing morning coffee, one should assign value to grinding beans, heating water, and so forth. Specifying value at the level of a goal or subgoal has the advantage that the particular method used to *attain* the goal can be computed ad hoc, depending on the specific circumstances (e.g., the coffee is already ground, or the water has to be retrieved from the well). Since value assignments of this kind are bundled hierarchically, the result is an abstract schema, or program, for accomplishing a goal. This kind of cognitive architecture is naturally suited to "goal emulation" (Tomasello, 1996), in which the observer imputes goals to the actor and then adopts those goals, but uses novel planning to derive her own policy for attaining the goal.

We can formally characterize social value inference using *inverse reinforcement learning* (IRL; Jara-Ettinger, 2019; Ng et al., 2000). Standard reinforcement learning (RL) models (Sutton & Barto, 2018) are used to understand how agents (biological or artificial) learn through interactions with the environment in non-social settings. As the agent interacts with the environment, they receive rewards, which update the value representations for different states and actions, allowing it to implement a policy for selecting future actions. IRL inverts this approach using Bayes inference to model inference about the hidden value representations that gave rise to observed actions.

While IRL models have been successfully applied to model how humans reason about others' preferences, goals, and beliefs (Baker et al., 2017; Collette et al., 2017; Jern et al., 2017), it is very computationally costly. For most interesting problems IRL is computationally intractable (Jara-Ettinger et al., 2016; Vélez & Gweon, 2020). Nevertheless, as a rational framework it can be used to uncover inductive biases that simplify the required computations. For instance, the *principle of efficient action* (Colomer et al., 2020; Gergely & Csibra, 2003; Jara-Ettinger et al., 2015; Scott & Baillargeon, 2013) assumes that other agents are acting in the most efficient manner towards achieving their aims, greatly constraining the hypothesis space for IRL inference. An intriguing possibility is that this may explain why very young infants are relatively more adept at inferring others goals or values than at inferring their beliefs (Gergely & Csibra, 2003).

Belief Inference and Model-based Planning

A final method of social learning is to update one's own model of the world (i.e., beliefs) when observing another person's actions. This affords maximal flexibility. Copying another person's actions or values may not be useful to you, because those actions and values may depend on their actor's unique circumstances, abilities, or desires. In contrast, true facts about the world are equally true for all of us. Learned beliefs can be seamlessly composed with existing ones, and then novel actions can be selected by planning in a way that reflects the specific circumstances, abilities and desires of the actor.

How do we learn facts by observing actions? Sometimes observing a person's actions will directly impart new knowledge about the world. For instance, you might learn that there is milk in a coconut the first time you see one cracked. But we also have the ability to learn facts about the world by inferring another person's beliefs. For instance, you might infer that a pond is well stocked with fish by observing an expert fisherman casting on it, even if you never actually observed her catch a fish. Here, our focus will be on this latter case: social learning based on the beliefs imputed to an expert actor.

Just as with values, an actor's unobserved beliefs can be inferred by Bayes' rule, given observations of their actions (Baker et al., 2009, 2017; Jara-Ettinger et al., 2016; Pantelis et al., 2014; Rafferty et al., 2015; Shafto et al., 2012). However, this may be computationally costlier than inferring the more direct linkage between actions and values. It is often possible to infer a person's values from their action without considering their beliefs; for instance, concluding that they enjoy eating apples when you see them bite into one. But it is much harder to infer a person's beliefs from their action without performing a joint inference over their values (or rewards) as well. If a person opens the cabinet, you cannot know what they *believe* is in the cabinet without a hypothesis about what they *want* to retrieve from it.

We can relate model-based social inference to analogous mechanisms of model-based planning in individual learning (K. J. Miller et al., 2017; Vikbladh et al., 2019). People often make decisions by computing the expected values of candidate actions given a model of their likely outcomes. This is more computationally demanding than working from cached value or cached policy, but it affords greater behavioral flexibility.

Interim summary

Every socially observed action contains information about rewards and the environment—information that also implies facts about value and the beliefs of the agent that produced them. This information is implicitly "packed" in any adaptive action. It is computationally cheap to copy the "packed" information in the format we observe them (i.e., as actions). Through inference, we can ascend the model (Fig. 1 left) and "unpack" the implicit, hidden structure that generated the action: the other individual's representations about value and their model of the environment. In this unpacked form, information can be used more flexibly, and can be productively composed with unique features of our own circumstances (our situation, abilities, other beliefs and values, etc.) Then, by re-descending the model during value-guided decision-making (Fig. right), we can translate socially learned information back into adaptive actions. Integration at higher levels affords more flexibility and generalization, but comes at increased computational costs. This sets up two difficult problems: Deciding when to rely on different social learning strategies, and integrating information learned at different "levels" into a coherent and adaptive set of behaviors.

Arbitration

How do humans arbitrate between social learning strategies? Here, again, we take inspiration from research on non-social learning and decision-making (Marewski & Schooler, 2011; Payne et al., 1988; Rieskamp & Otto, 2006). A common thread that runs through this literature is that meta-cognitive strategy selection balances the cost of a computation against the relative effectiveness of a strategy (Gershman et al., 2015; Lieder & Griffiths, 2020). For instance, we engage in simpler model-free learning when there is little added value to using more sophisticated and computationally expensive model-based planning—but if the extra computation greatly improves accuracy, then the cost-benefit trade-off may tip the scales in the other direction (Kool et al., 2017). We speculate that similar principles arbitrate between social learning strategies.

Yet this is still a relatively understudied problem in the social learning literature. Much past work has focused on a slightly different question: How humans and other animals arbitrate between social and non-social learning—specifically, whether to copy others or rely on firsthand experience (Kendal et al., 2018; Laland, 2004). This problem can be modeled as an evolutionary game, where each individual can choose either individual learning, which is assumed to always yield the same payoff, or imitation, where payoffs depend on the frequency of other imitators in the population (Rogers, 1988). Game theoretic models predict that imitation is profitable when other imitators are sparse but fails when they are too abundant; too many imitators can get stuck copying each other, instead of striking out and discovering better solutions (Toyokawa et al., 2019; Tump et al., 2020). Accordingly, humans adapt their strategies to the structure of their social environment. People self-organize into balanced mixtures of individual learners and imitators (Kameda & Nakanishi, 2002), they rely more on imitation when solving harder problems or when placed in larger groups (Toyokawa et al., 2019), and they selectively imitate individual learners (Wu et al., 2021).

In contrast, less is known about how humans arbitrate different forms of social learning—for example, when do we infer others' values, and when does it suffice to merely copy their actions? Recent behavioral and neuroimaging evidence suggests that when comparing choice imitation against value inference (called goal emulation in the paper), people preferred the strategy that produced the most reliable predictions about outcomes (Charpentier et al., 2020). A neural signal of reliability was found for the value inference strategy, but not for imitation. The reliability signal was continuously updated on a trial-by-trial basis, corresponding to neural activity in the bilateral temporoparietal junction and right ventrolateral prefrontal cortex, which is hypothesized to provide a control signal.

To return to an example above: If your friend orders "dax" instead of a salad, you can confidently infer that "dax" has higher value and use it to decide between new choice options—for example, you might order "dax" over the salad, but pass it up in favor of a delicious pizza. However, learners do not always have sufficient evidence to infer value. If your friend introduces you to an entirely new cuisine and orders the "dax" instead of the "fep," you might have no idea of how valuable "dax" is, or whether it is tastier than a "zog." In this case, the safest option would be to imitate your friend's policy and stick to ordering "dax."

Research on the arbitration between value-based and model-based forms of social learning is even sparser, and we again turn to theories from individual learning for inspiration. It has been observed that hippocampal replay/preplay (hypothesized to support model-based planning) prioritize the retrieval of memories according to two statistics: their Need and their Gain (Mattar & Daw, 2018). Need describes the relevance of a specific state (i.e., how often it will be visited under the current policy), while Gain describes the expected improvement in reward based on changes in policy at that state. These results suggest that the neural mechanisms underlying model-based planning are sensitive to what types of planning is more useful, with priority given to the most relevant and consequential aspects. Similar computations could provide a control signal for the general effectiveness of model-based planning, which could facilitate the cost-benefit arbitration of value-based and model-based behavior observed in individual learning (Kool et al., 2017). We can further speculate that similar processes perhaps also arbitrate between value-based and model-based forms of social learning, though there is much fertile ground for future empirical research in this direction.

Integration

So far we have emphasised three "horizontal pathways" for social learning (Fig. 1): The direct copying of policy, the copying of imputed values, and the copying of imputed beliefs. We now turn to consider "vertical pathways" for exchanging information between representational formats (Cushman, 2019). This includes ascending from observed actions to imputed mental states via inference, and also descending from mental states into actions via value-guided decision-making. A key part of the success of human social learning, we suggest, is our ability to productively exchange information between these levels of representation.

Descending via decision-making and amortization

The simplest representational exchange occurs during decision-making. Planning, for instance, involves the compression of one's model of the environment into a value representation, and then implementing a policy to select actions on the basis of value (Sutton & Barto, 2018). Each of these steps creates new representations that can be stored or "cached" for reuse in similar future episodes (Cushman & Morris, 2015; Keramati et al., 2016; Maisto et al., 2019), such that the computational burden can be distributed over time (i.e., amortization; Dasgupta et al., 2018).

Amortization is also useful because precompiled representations can be deployed rapidly for computationally efficient actions. Imagine a downhill skier who must navigate a slalom course at high speeds, executing incredibly precise movements. In this setting, computationally efficiency is imperative for peak performance. To learn such swift maneuvers, the skier could repeatedly practice at slower speeds, gradually caching value or policy representations through habitization. Or, the skier could form a mental model of the course and acquire cached representations

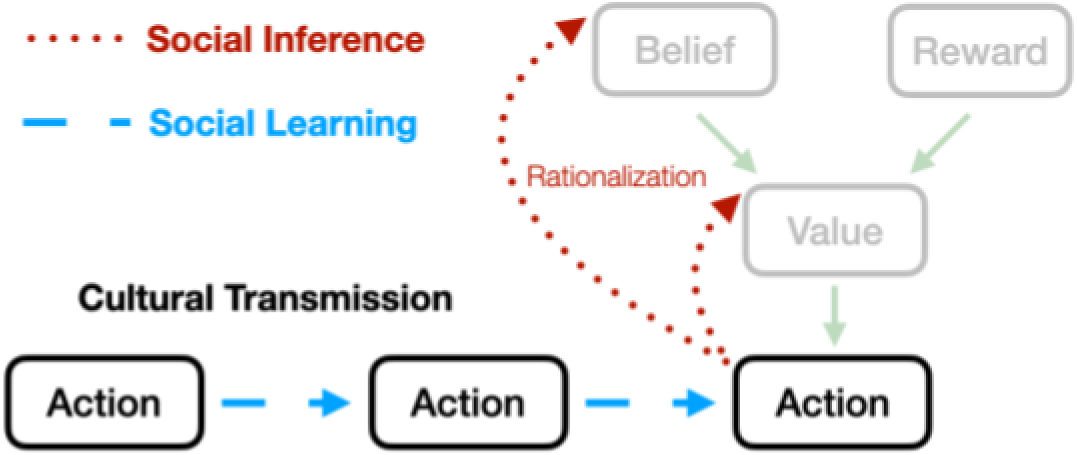

Figure 2. "Rationalization" of cultural practices. Behaviors may be transmitted through cultural norms, but because observed actions are rich in implicit information, they can be "unpacked" into imputed mental states. These mental states may not belong to the other individual (who directly acquired the behavior through imitation), yet they provide a useful fiction, exchanging an inflexible cultural norm for flexible and compositional representations to solve new problems.

through mental simulations. Both methods generate new, compressed representations that facilitate rapid performance on the course. Current evidence suggests that offline amortization occurs during periods of both wakeful rest (Eldar et al., 2020; Momennejad et al., 2018) and sleep (Foster, 2017; Ólafsdóttir et al., 2018; Pavlides & Winson, 1989).

Still another possibility is to learn by observing other skillful skiers. Yet, what representations should be learned? If we adopted the view that "imitation" and "emulation" are exclusively competing strategies, then our skier would face a difficult choice. Should she copy the actions of the expert, and hope that the individual differences are not too large? Or, should she fully decompose the observed behavior into its causal primitives—the beliefs, rewards and values that shaped the expert's performance? The latter offers more flexibility, but requires a great deal of thought. It's difficult to imagine she could make it down the mountain at any reasonable speed.

Offline amortization rescues the "emulation" strategy. Rather than having to recompute her plan while on the slope, she can integrate whatever she has learned from the expert with her own knowledge through mental simulation in the safety of her hotel room. Still, the required mental effort is daunting. Presumably, real skiers don't always "unpack" observed behavior into their most elemental components and then fully re-plan a new policy. Rather, they likely employ a mixture of different social strategies, guided by the principles of arbitration: using cheap methods whenever possible and more costly methods when more reliable. Re-planning can then be used to re-assemble these components—together with the skier's own beliefs, values and policy—into a new and better trajectory down the hill.

Ascending via inference and rationalization

When we observe another person’s actions, we can infer the hidden mental states that gave rise to those actions, including their beliefs, reward function, and computations of instrumental value. This is a basic form of “ascent” in our model (Figure 1). And, in the social cognition literature, this kind of thing is said so frequently that it can be easy to overlook it’s hidden premise: That, when we observe another person’s actions, those actions really were caused by their beliefs, reward function and computations of instrument value—that is, by rational planning.

But, often, they aren’t. People’s actions are also the product of habit (i.e., when they draw upon cached value or cached policy representations), instinct (i.e., innate behavioral responses), or conformity to cultural norms (i.e., when they imitate what other people do without reasoning about why). In these cases, what would it mean if an observer imputed beliefs, a reward function, and computations of instrumental value to the actor? It would mean that the observer is constructing a fiction—a “rationalization” of behavior (Cushman 2019)—which we will argue, can be quite useful. This fiction furnishes a key method of representational exchange, extracting implicit information from the cached policies or values of other people.

Suppose, for instance, that an aspiring baker wishes to improve the flavor of her loafs. She notices that in her culture people let their dough rise overnight, and she imputes the belief that this is a superior method—perhaps because the cool temperatures allow flavor to build during a longer fermentation. So, she tries putting her loaves in especially cool spots, even during the daytime. Now, it might be the case that in her culture, nobody knows why they let the bread rise overnight, they just do it because “that’s how it’s done”. Nevertheless, by imputing beliefs, values and rational choice, the aspiring baker might learn useful information from cultural practices that she can generalize to new settings.

The rationalization of cultural practices is ubiquitous. Its basic structure is depicted in Figure 2: A cultural norm is transmitted at the level of cached policy, but then later it is “rationalized”, yielding putative representations of values and beliefs. Parents and schoolteachers constantly find themselves attempting to “make sense” out of cultural practices for inquisitive children. Sometimes we are honest with ourselves, acknowledging that neither we nor perhaps anybody else has ever considered why we do certain things the way we do, but that there may be some implicit wisdom that we can extract nevertheless. But just as often, we are less honest with ourselves (and with children), acting as if either we had chosen or some forgotten Ur-designer had constructed, this cultural practice for precisely the reasons we articulate to the child. In this case we are constructing a fiction, but nevertheless, the resulting representations afford flexible and compositional thinking, where formerly an inflexible cultural norm prevailed.

Of course, we rationalize not only other's behaviors, and "culture" more generally, but also our own behavior. Insofar as our own behavior is caused by non-rational processes (e.g., habits, instincts or norms), we can falsely impute reasoning processes to ourselves, assigning the beliefs, values and rewards in light of which the behaviors we perform would have been rational. Insofar as our behaviors are often guided by social learning, this is another pathway by which socially acquired information can propagate across levels of representation within our own mind (Cushman, 2019). For instance, having learned the motor routines involved in bread-making (a form of policy imitation), we can rationalize our own actions and thus extract useful values or beliefs about the bread-making process for further innovation.

Conclusion

Much attention has been given to the problem of how humans learn from others. Rather than trying to pick out a single distinguishing feature of what makes humans special, we have argued the distinction is in how the different mechanisms of social learning are harmonized.

We draw inspiration from the literature on non-social decision-making, where there is emerging consensus that human intelligence arises from the productive cooperation of diverse strategies for learning and decision-making (Huys et al., 2015; Kool et al., 2018; Solway & Botvinick, 2015). Just as individual learning consists of multiple strategies, social learning can deploy a range of tools: We can directly copy observed behavior (*policy imitation*), or by leveraging the implications of rational actions, we can infer hidden mental states, such as value representations (*value inference*) or beliefs about the world (*belief inference*) imputed to the demonstrator. While it is computationally cheap to directly copy behavior, the more we unpack it into the primitive components of decision-making, the more flexible and compositional it becomes. Arbitration between different social learning mechanisms seems to balance computational costs against effectiveness, although the exact cognitive and neural mechanisms are only beginning to be understood.

Moreover, learning from others is just the first step in a series of cognitive challenges. We must next integrate what we have learned with our pre-existing policies, values and beliefs. To do this, we exchange information between different formats of representation. This involves "descending" pathways (moving from more flexible and compositional representations towards compressed policy-relevant representations), as well as "ascending" pathways (extracting information implicit in policy-relevant representations into more flexible and compositional elements).

We study social learning because we want to understand real human behaviors—the kinds we perform every day, such as baking a loaf of bread or choosing what to eat. One way in which these behaviors are paradigmatic of human intelligence is that we are able to blend strategies we've learned from others with strategies we've developed ourselves. Another part is the way that they involve skills and representations spanning from very specific, concrete behaviors to very abstract, general principles. Our ability to build harmony across these levels is essential to the virtuosic music of the human mind.

References

Apperly, I. (2010). *Mindreaders: The Cognitive Basis of "Theory of Mind."* Psychology Press.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 1–10.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Bandura, A. (1962). Social learning through imitation. *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation*, *330*, 211–274.

Botvinick, M., & Weinstein, A. (2014). Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *369*(1655). https://doi.org/10.1098/rstb.2013.0480

Boyd, R., & Richerson, P. J. (1988). *Culture and the Evolutionary Process*. University of Chicago Press.

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.

Collette, S., Pauli, W. M., Bossaerts, P., & O'Doherty, J. (2017). Neural computations underlying inverse reinforcement learning in the human brain. *eLife*, *6*. https://doi.org/10.7554/eLife.29718

Colomer, M., Bas, J., & Sebastian-Galles, N. (2020). Efficiency as a principle for social preferences in infancy. *Journal of Experimental Child Psychology*, *194*, 104823.

Cushman, F. (2019). Rationalization is rational. *The Behavioral and Brain Sciences*, *43*, e28.

Cushman, F., & Morris, A. (2015). Habitual control of goal selection in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(45), 13817–13822.

Dasgupta, I., Schulz, E., Goodman, N. D., & Gershman, S. J. (2018). Remembrance of inferences past: Amortization in human hypothesis generation. *Cognition*, *178*, 67–81.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*(12), 1704–1711.

Derex, M., Bonnefon, J.-F., Boyd, R., & Mesoudi, A. (2019). Causal understanding is not necessary for the improvement of culturally evolving technology. *Nature Human Behaviour*, *3*(5), 446–452.

Eldar, E., Lièvre, G., Dayan, P., & Dolan, R. J. (2020). The roles of online and offline replay in planning. *eLife*, *9*. https://doi.org/10.7554/eLife.56911

Foster, D. J. (2017). Replay Comes of Age. *Annual Review of Neuroscience*, *40*, 581–602.

Galef, B. G., Jr. (1988). Imitation in animals: History, definition and interpretation of the data from the psychological laboratory. In T. R. Zentall & B. G. Galef Jr (Eds.), *Social learning: Psychological and Biological Perspectives* (pp. 15–40). Psychology Press.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naıve theory of rational action. In *Trends in Cognitive Sciences* (Vol. 7, Issue 7, pp. 287–292). https://doi.org/10.1016/s1364-6613(03)00128-1

Gershman, S. J. (2020). Origin of perseveration in the trade-off between reward and complexity. *Cognition*, *204*, 104394.

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278.

Gweon, H. (2021). *Inferential Social Learning: How humans learn from others and help others learn*. https://doi.org/10.31234/osf.io/8n34t

Henrich, J. (2017). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton University Press.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior: Official Journal of the Human Behavior and Evolution Society*, *22*(3), 165–196.

Herrmann, E., Call, J., Hernàndez-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science*, *317*(5843), 1360–1366.

Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, *5*(6), 253–261.

Heyes, C. (2002). Transformational and associative theories of imitation. *Imitation in Animals and Artifacts.*, *607*, 501–523.

Heyes, C. (2018). *Cognitive Gadgets: The Cultural Evolution of Thinking*. Harvard University Press.

Ho, M. K., MacGlashan, J., Littman, M. L., & Cushman, F. (2017). Social is special: A normative framework for teaching with and learning from evaluative feedback. *Cognition*, *167*, 91–106.

Hoppitt, W., & Laland, K. N. (2013). *Social Learning: An Introduction to Mechanisms, Methods, and Models*. Princeton University Press.

Horner, V., & Whiten, A. (2005). Causal knowledge and imitation/emulation switching in chimpanzees (Pan troglodytes) and children (Homo sapiens). *Animal Cognition*, *8*(3), 164–181.

Huys, Q. J. M., Lally, N., Faulkner, P., Eshel, N., Seifritz, E., Gershman, S. J., Dayan, P., & Roiser, J. P. (2015). Interplay of approximate planning strategies. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(10), 3098–3103.

Jara-Ettinger, J. (2019). Theory of mind as inverse reinforcement learning. *Current Opinion in Behavioral Sciences*, *29*, 105–110.

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The Naïve Utility Calculus: Computational Principles Underlying Commonsense Psychology. *Trends in Cognitive Sciences*, *20*(8), 589–604.

Jara-Ettinger, J., Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2015). Children's understanding of the costs and rewards underlying rational action. *Cognition*, *140*, 14–23.

Jern, A., Lucas, C. G., & Kemp, C. (2017). People learn other people's preferences through inverse decision-making. *Cognition*, *168*, 46–64.

Jiménez, Á. V., & Mesoudi, A. (2019). Prestige-biased social learning: current evidence and outstanding questions. *Palgrave Communications*, *5*(1), 20.

Keramati, M., Smittenaar, P., Dolan, R. J., & Dayan, P. (2016). Adaptive integration of habits into depth-limited planning defines a habitual-goal-directed spectrum. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(45), 12868–12873.

Kool, W., Cushman, F. A., & Gershman, S. J. (2018). Chapter 7 - Competition and Cooperation Between Multiple Reinforcement Learning Systems. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-Directed Decision Making* (pp. 153–178). Academic Press.

Kool, W., Gershman, S. J., & Cushman, F. A. (2017). Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. *Psychological Science*, *28*(9), 1321–1333.

Legare, C. H., & Nielsen, M. (2015). Imitation and Innovation: The Dual Engines of Cultural Learning. *Trends in Cognitive Sciences*, *19*(11), 688–699.

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. In *Behavioral and Brain Sciences* (Vol. 43). https://doi.org/10.1017/s0140525x1900061x

Lyons, D. E., Young, A. G., & Keil, F. C. (2007). The hidden structure of overimitation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(50), 19751–19756.

Maisto, D., Friston, K., & Pezzulo, G. (2019). Caching mechanisms for habit formation in Active Inference. *Neurocomputing*, *359*, 298–314.

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: an ecological model of strategy selection. *Psychological Review*, *118*(3), 393–437.

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, *21*(11), 1609–1617.

McGuigan, N., Whiten, A., Flynn, E., & Horner, V. (2007). Imitation of causally opaque versus causally transparent tool use by 3- and 5-year-old children. *Cognitive Development*, *22*(3), 353–364.

Miller, K. J., Botvinick, M. M., & Brody, C. D. (2017). Dorsal hippocampus contributes to model-based planning. *Nature Neuroscience*. https://doi.org/10.1101/096594

Miller, K. J., Shenhav, A., & Ludvig, E. A. (2019). Habits without values. *Psychological Review*, *126*(2), 292–311.

Miller, N. E., & Dollard, J. (1941). *Social Learning and Imitation* (Vol. 55). Yale University Press.

Momennejad, I., Otto, A. R., Daw, N. D., & Norman, K. A. (2018). Offline replay supports planning in human reinforcement learning. *eLife*, *7*. https://doi.org/10.7554/eLife.32548

Najar, A., Bonnet, E., Bahrami, B., & Palminteri, S. (2020). The actions of others act as a pseudo-reward to drive imitation in the context of social reinforcement learning. *PLoS Biology*, *18*(12), e3001028.

Ng, A. Y., Russell, S. J., & Others. (2000). Algorithms for inverse reinforcement learning. *Icml*, *1*, 2.

Ólafsdóttir, H. F., Bush, D., & Barry, C. (2018). The Role of Hippocampal Replay in Memory and Planning. *Current Biology: CB*, *28*(1), R37–R50.

Pantelis, P. C., Baker, C. L., Cholewiak, S. A., Sanik, K., Weinstein, A., Wu, C.-C., Tenenbaum, J. B., & Feldman, J. (2014). Inferring the intentional states of autonomous virtual agents. *Cognition*, *130*(3), 360–379.

Pavlides, C., & Winson, J. (1989). Influences of hippocampal place cell firing in the awake state on the activity of these cells during subsequent sleep episodes. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *9*(8), 2907–2918.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(3), 534–552.

Rafferty, A. N., LaMar, M. M., & Griffiths, T. L. (2015). Inferring learners' knowledge from their actions. *Cognitive Science*, *39*(3), 584–618.

Rieskamp, J., & Otto, P. E. (2006). SSL: a theory of how people learn to select strategies. *Journal of Experimental Psychology. General*, *135*(2), 207–236.

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS Computational Biology*, *13*(9), e1005768.

Scott, R. M., & Baillargeon, R. (2013). Do infants really expect agents to act efficiently? A critical test of the rationality principle. *Psychological Science*, *24*(4), 466–474.

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning From Others: The Consequences of Psychological Reasoning for Human Learning. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *7*(4), 341–351.

Skinner, B. F. (1950). Are theories of learning necessary? *Psychological Review*, *57*(4), 193–216.

Solway, A., & Botvinick, M. M. (2012). Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. *Psychological Review*, *119*(1), 120–154.

Solway, A., & Botvinick, M. M. (2015). Evidence integration in model-based tree search. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(37), 11708–11713.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning, second edition: An Introduction*. MIT Press.

Tennie, C., Call, J., & Tomasello, M. (2009). Ratcheting up the ratchet: on the evolution of cumulative culture. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1528), 2405–2415.

Tolman, E. C. (1948). Cognitive maps in rats and men. In *Psychological Review* (Vol. 55, Issue 4, pp. 189–208). https://doi.org/10.1037/h0061626

Tomasello, M. (1996). Do apes ape. *Social Learning in Animals: The Roots of Culture*, 319–346.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *The Behavioral and Brain Sciences*, *28*(5), 675–691; discussion 691–735.

Tomasello, M., Davis-Dasilva, M., Camak, L., & Bard, K. (1987). Observational learning of tool-use by young chimpanzees. *Human Evolution*, *2*(2), 175–183.

Vélez, N., & Gweon, H. (2020). *Learning from other minds: An optimistic critique of reinforcement learning models of social learning*. https://doi.org/10.31234/osf.io/q4bxr

Vikbladh, O. M., Meager, M. R., King, J., Blackmon, K., Devinsky, O., Shohamy, D., Burgess, N., & Daw, N. D. (2019). Hippocampal Contributions to Model-Based Planning and Spatial Memory. *Neuron*, *102*(3), 683–693.e4.

Whiten, A., & Ham, R. (1992). Kingdom: reappraisal of a century of research. *Advances in the Study of Behavior*, *21*, 239.

Wu, C. M., Ho, M., Kahl, B., Leuker, C., Meder, B., & Kurvers, R. (2021). Specialization and selective social attention establishes the balance between individual and social learning. In *bioRxiv*.