

# A COMMON FRAMEWORK FOR THEORIES OF NORM COMPLIANCE\*

BY ADAM MORRIS AND FIERY CUSHMAN

*Abstract: Humans often comply with social norms, but the reasons why are disputed. Here, we unify a variety of influential explanations in a common decision framework, and identify the precise cognitive variables that norms might alter to induce compliance. Specifically, we situate current theories of norm compliance within the reinforcement learning framework, which is widely used to study value-guided learning and decision-making. This framework offers an appealingly precise language to distinguish between theories, highlights the various points of convergence and divergence, and suggests novel ways in which norms might penetrate our psychology.*

KEY WORDS: norms, reinforcement learning, compliance, cognitive psychology, value, decision-making

## I. INTRODUCTION

Social norms — behavioral prescriptions or proscriptions known by members of a group and informally socially enforced — are central to human societies.<sup>1</sup> They range from fashion trends to religious customs, from norms of honor in the American South<sup>2</sup> to norms of reciprocity in northeast Uganda,<sup>3</sup> from honest tax reporting in industrialized countries<sup>4</sup> to conflict resolution among rural farmers.<sup>5</sup> Thus, it is no surprise that virtually every corner of social and cognitive science has attempted to understand what norms are, when and why they emerge, and how they influence human behavior.<sup>6</sup>

\* We thank Jonathan Phillips and other members of the Moral Psychology Research Lab for their advice and assistance. This research was supported by Grant N00014-14-1-0800 from the Office of Naval Research.

<sup>1</sup> Christina Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (New York: Cambridge University Press, 2006); R. Cooter, "Do Good Laws Make Good Citizens? An Economic Analysis of Internalized Norms," *Virginia Law Review* 86, no. 8 (2000): 1577–1601. <https://doi.org/10.2307/1073825>; E. Fehr and U. Fischbacher, "Social Norms and Human Cooperation," *Trends in Cognitive Sciences* 8, no. 4 (2004): 185–90. <https://doi.org/10.1016/j.tics.2004.02.007>

<sup>2</sup> R. E. Nisbett and D. Cohen, *Culture of Honor: The Psychology of Violence in the South*, vol. xviii (Boulder, CO: Westview Press, 1996).

<sup>3</sup> C. M. Turnbull, *The Mountain People* (New York: Touchstone, 1987).

<sup>4</sup> Jon Elster, "Social Norms and Economic Theory," *The Journal of Economic Perspectives* 3, no. 4 (1989): 99–117.

<sup>5</sup> R. Ellickson, *Order without Law: How Neighbors Settle Disputes*, rev. ed. (Cambridge, MA: Harvard University Press, 1944).

<sup>6</sup> For much more nuanced definitions of norms and their kinds, see Bicchieri, *The Grammar of Society*.

It is equally unsurprising that not everybody agrees on the answers. Consider, for instance, the question of how norms influence behavior. People tend to comply with the norms they are exposed to. In many contexts, they choose to adopt the behaviors that the norms prescribe and avoid behaviors that the norms proscribe. Why? How does exposure to norms influence subsequent decision-making? Some prominent theories include: (1) People adopt the statistically dominant patterns of behavior they observe around them,<sup>7</sup> (2) people develop heuristic responses to social dilemmas based on their prior history of reward and punishment,<sup>8</sup> and (3) people learn to derive subjective utility from the absolute and relative welfare of others.<sup>9</sup> As these examples make apparent, contemporary theories of norm compliance are not just saying different things; they're speaking different languages.

Our goal is to propose a *lingua franca* for theories of norm compliance. Our point of entry is a computational framework for learning and decision-making that unites key ideas from machine learning, neuroscience, psychology, and economics. The framework, known as reinforcement learning (RL), identifies a set of cognitive variables involved in the decision process, any of which norms could intervene on. It offers a precise language in which to enumerate the ways that norms could influence decisions, and specifies exactly how those interventions would affect decision makers. In essence, RL is a broad framework that asks how people learn values and then make decisions based on those values, and offers a formal setting in which competing answers can be precisely articulated. It is an ideal framework for theories of norm compliance, which also involves learning and deciding based on representations of value.

We map various influential theories of norm compliance onto this framework, compare their commitments and predictions, and propose novel potential mechanisms of norm compliance. Our analysis suggests that norms influence many parts of the decision-making process, including ones not yet considered, and that norm compliance is grounded in similar computational principles as other forms of human decision-making.

In this endeavor, we focus on norms that are intuitively "moral": Norms for fairness, cooperation, reciprocity, and so forth, which instruct people to

<sup>7</sup> R. B. Cialdini and M. R. Trost, "Social Influence: Social Norms, Conformity and Compliance," in D. T. Gilbert, S. T. Fiske, and G. Lindzey, eds., *The Handbook of Social Psychology*, Vols. 1 and 2, 4th ed. (New York: McGraw-Hill, 1998), 151–92.

<sup>8</sup> D. G. Rand and Z. G. Epstein, "Risking Your Life without a Second Thought: Intuitive Decision-Making and Extreme Altruism," *PLOS ONE* 9, no. 10 (2014), e109687. <https://doi.org/10.1371/journal.pone.0109687>.

<sup>9</sup> E. Fehr and K. M. Schmidt, "The Economics of Fairness, Reciprocity and Altruism—Experimental Evidence and New Theories," in S. C. K. and J. M. Ythier, ed., *Handbook of the Economics of Giving, Altruism and Reciprocity* Vol. 1 (Amsterdam: Elsevier, 2006), 615–91. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1574071406010086>.

forgo personal gain to promote the well-being of others. (Think of norms for contributing to the public offering at church, or for recycling, or for offering fair monetary splits in an Ultimatum Game.<sup>10</sup>) Moral norms offer an ideal testing ground for theories of norm compliance, because their influence is strong enough to overpower other competing interests. Which parts of the decision process are altered by exposure to moral norms, and how do these alterations make people more likely to obey the norm?<sup>11</sup>

## II. A FRAMEWORK FOR MODELING VALUE-GUIDED DECISION-MAKING

To answer this, we need a cognitive framework of learning and decision-making. We take as our starting point the idea that humans often make decisions by putting their options on a common scale of subjective “goodness,” and biasing choice toward options with more of it. In other words, they choose according to each option’s value. The idea of value-based choice has a long history in both psychology<sup>12</sup> and economics,<sup>13</sup> and has recently assumed an equally prominent role in neuroscience.<sup>14</sup>

Value can be interpreted in at least two ways. One interpretation, common in economics, is that value is simply an abstract description of consistent choice, and not necessarily an actual variable computed by humans during the decision process. However, advances in cognitive and neural science in the last twenty years indicate that humans indeed compute a representation of the overall goodness of an option, and use it to guide their choices.<sup>15</sup> We therefore interpret value as a real cognitive variable computed during the decision process.<sup>16</sup>

Our aim is to ground current theories of norm compliance in the RL framework for value-guided decision-making in humans. Of course,

<sup>10</sup> R. H. Thaler, “Anomalies: The Ultimatum Game,” *Journal of Economic Perspectives* 2, no. 4 (1988): 195–206.

<sup>11</sup> We use the terms “obey,” “follow,” “comply with,” and so on, interchangeably to indicate situations where people do what a norm prescribes (or avoid doing what a norm forbids). Unlike some authors, we do not use the terms to differentiate hypotheses about the psychology underlying that behavior (H. C. Kelman, “Compliance, Identification, and Internalization: Three Processes of Attitude Change,” *The Journal of Conflict Resolution* 2, no. 1 [1958]: 51–60; H. H. Koh, “Why Do Nations Obey International Law?” *Yale Law Journal* 106, no. 8 [1997]: 2599–2659. <https://doi.org/10.2307/797228>).

<sup>12</sup> R. J. Dolan and P. Dayan, “Goals and Habits in the Brain,” *Neuron* 80, no. 2 (2013): 312–25. <https://doi.org/10.1016/j.neuron.2013.09.007>.

<sup>13</sup> P. A. Samuelson, “Foundations of Economic Analysis,” *Science and Society* 13, no. 1 (1948): 93–95; J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press, 1944).

<sup>14</sup> P. W. Glimcher and E. Fehr, *Neuroeconomics: Decision Making and the Brain* (London: Academic Press, 2013).

<sup>15</sup> *Ibid.*

<sup>16</sup> G. S. Becker, *Accounting for Tastes* (Cambridge, MA: Harvard University Press, 1996); Amartya Sen, *Choice, Welfare and Measurement* (Cambridge, MA: Harvard University Press, 1997); Cass R. Sunstein, “Social Norms and Social Roles,” *Columbia Law Review* 96, no. 4 (1996): 903–968. <https://doi.org/10.2307/1123430>

this presumes that decisions to comply with norms involve computing the value of options. But this is not a necessary truth; not all behavior is value-guided. One major counterexample is reflexive behavior, which is triggered by perceptual and cognitive processes that do not explicitly represent value in any meaningful sense.<sup>17</sup> Moreover, many nonreflexive choices appear to be driven by processes that do not place options on a common scale, and therefore are not value-guided.<sup>18</sup>

So there are viable alternatives to a value-guided theory of norm compliance; indeed, we will return to consider some reasons to take alternative approaches very seriously. One key aim of the computational framework we present below is to make precise what it actually means to say that decision-making is guided by value, such that the claim that norms influence behavior via such representations is both substantive and falsifiable.

With these caveats in mind, however, there are at least two major reasons to begin with the assumption that norm compliance operates in humans principally via processes of value-guided decision-making. First, people's choices to comply with norms show signatures of rational planning over a common currency objective. Their choices in economic games conform to the maxims of consistency,<sup>19</sup> and obey the law of demand: as the cost of being fair, reciprocal, and so on, increases, people correspondingly act in these ways less.<sup>20</sup> This is intuitively true for real-world examples of norm compliance also: People are more likely to cheat on their taxes when it will save them more money, or run a red light when they are in a rush. The ability to flexibly trade off compliance with other goods is a hallmark of value-guided decision-making.

<sup>17</sup> I. P. Pavlov and G. V. Anrep, *Conditioned Reflexes* (North Chelmsford, MA: Courier Corporation, 1927).

<sup>18</sup> S. Lichtenstein and P. Slovic, *The Construction of Preference* (New York: Cambridge University Press, 2006); I. Vlaev, N. Chater, N. Stewart, and G. D. A. Brown, "Does the Brain Calculate Value?" *Trends in Cognitive Sciences* 15, no. 11 (2011): 546–54. <https://doi.org/10.1016/j.tics.2011.09.00>

<sup>19</sup> J. Andreoni, M. Castillo, and R. Petrie, "What Do Bargainers' Preferences Look Like? Experiments with a Convex Ultimatum Game," *The American Economic Review* 93, no. 3 (2003): 672–85; J. Andreoni and J. Miller, "Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism," *Econometrica* 70, no. 2 (2002): 737–53. <https://doi.org/10.1111/1468-0262.00302>

<sup>20</sup> C. M. Anderson and L. Putterman, "Do Non-Strategic Sanctions Obey the Law of Demand? The Demand for Punishment in the Voluntary Contribution Mechanism," *Games and Economic Behavior* 54, no. 1 (2006): 1–24. <https://doi.org/10.1016/j.geb.2004.08.007>; J. Andreoni and L. Vesterlund "Which is the Fair Sex? Gender Differences in Altruism," *The Quarterly Journal of Economics* 116, no. 1 (2001): 293–312; V. Capraro, J. J. Jordan, and D. G. Rand, *Heuristics Guide the Implementation of Social Preferences in One-Shot Prisoner's Dilemma Experiments* (SSRN Scholarly Paper No. ID 2429862) (Rochester, NY: Social Science Research Network, 2014). Retrieved from <http://papers.ssrn.com/abstract=2429862>; J. P. Carpenter, "The Demand for Punishment," *Journal of Economic Behavior and Organization* 62, no. 4 (2007): 522–42. <https://doi.org/10.1016/j.jebo.2005.05.004>

Second, prototypically norm-compliant choices are accompanied by increased activity in brain regions involved in domain-general value computation.<sup>21</sup> When people choose to allocate money fairly, or punish someone who violated trust, or mutually cooperate in a Prisoner's Dilemma, they show patterns of activation in the ventromedial frontal and orbitofrontal cortex consistent with value-guided choice — even when doing so causes them personal financial loss.<sup>22</sup> These same regions appear to compute and store the value of options across a wide variety of other domains,<sup>23</sup> suggesting that people's decisions to comply with norms are guided by similar processes.

In summary, there is good reason to believe that norm compliance occurs at least in part by processes of value-guided decision-making. Next, we present a powerful formal computational framework for understanding the mechanics of such processes. It comprises two parts: a statement of the decision-making *problem*, which is formalized as a Markov Decision Process (MDP), and a description of the kinds of *algorithmic solutions* that have been pursued to solve that problem, which are called reinforcement learning (RL) approaches.

This framework will suggest five cognitive decision variables that could be influenced by norms. We will map four of these onto existing theories of norm compliance, and offer the other as a direction for future research.

### A. Defining the problem: Markov decision processes

Before defining MDPs in abstract terms, it is helpful to orient ourselves around the concrete details of an idealized example. Imagine a rat, Lenny, in a maze with cheese and electric shocks (Figure. 1). Lenny must decide how to traverse the maze in a way that achieves his objectives. As described above, we assume that the subjective goodness or badness of the cheese and electric shocks can be put on a common numeric scale, which in the MDP framework is called "reward." For example, Lenny's rewards might be: *CHEESE* = +1, *SHOCK* = -1. Lenny, then, has a single objective:

<sup>21</sup> C. C. Ruff and E. Fehr, "The Neurobiology of Rewards and Values in Social Decision Making," *Nature Reviews Neuroscience* 15, no. 8 (2014): 549–62. <https://doi.org/10.1038/nrn3776>

<sup>22</sup> D. J. F. de Quervain, U. Fischbacher, V. Treyer, M. Schellhammer, U. Schnyder, A. Buck, and E. Fehr, "The Neural Basis of Altruistic Punishment," *Science* 305, no. 5688 (2004): 1254–58. <https://doi.org/10.1126/science.1100735>; J. K. Rilling, D. A. Gutman, T. R. Zeh, G. Pagnoni, G. S. Berns, and C. D. Kilts, "A Neural Basis for Social Cooperation," *Neuron* 35, no. 2 (2002): 395–405. [https://doi.org/10.1016/S0896-6273\(02\)00755-9](https://doi.org/10.1016/S0896-6273(02)00755-9); G. Tabibnia and M. D. Lieberman, "Fairness and Cooperation Are Rewarding," *Annals of the New York Academy of Sciences* 1118, no. 1 (2007): 90–101. <https://doi.org/10.1196/annals.1412.001>; J. Zaki and J. P. Mitchell, "Equitable Decision Making is Associated with Neural Markers of Intrinsic Value," *Proceedings of the National Academy of Sciences* 108, no. 49 (2011): 19761–766. <https://doi.org/10.1073/pnas.1112324108>

<sup>23</sup> Glimcher and Fehr, *Neuroeconomics: Decision Making and the Brain*.

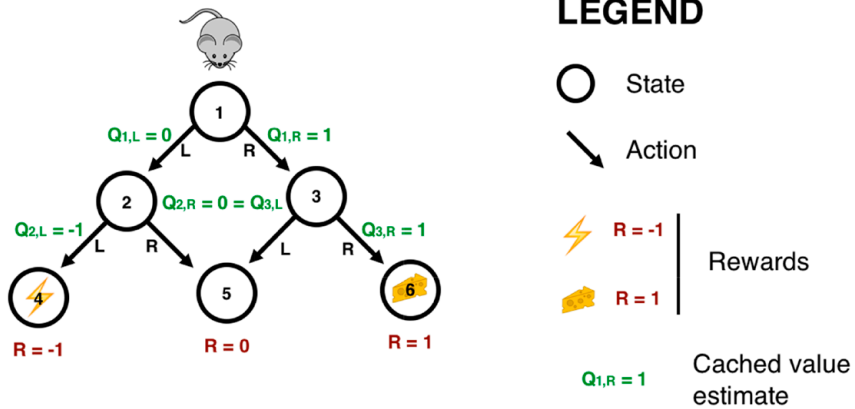


FIGURE 1. MDP representation of a simple decision problem. A rat must decide how to traverse a maze (represented as a set of 6 states, each with a set of potential actions) to maximize long-term accumulation of reward.  $R$ s represent the intrinsic rewards associated with each terminal state, and  $Q$ s represent the average value (i.e., long-term expected future reward) that the rat would learn to associate with each prior action. For instance, turning right at State 1 eventually leads to the cheese in State 6, and the rat would therefore learn that  $Q(\text{state } 1, \text{action } R) = +1$ .

maximize the long-term accumulation of reward. Lenny could then compute the amount of long-term reward gain expected from turning “left” or “right” at each fork in the maze, and simply choose the option with the highest amount.

This situation — an agent has a set of well-defined choices with well-defined consequences, and her goal is to maximize a single numerical objective — is what MDPs formally describe. MDPs carve the world into a set of discrete states  $S$ , which can be either decision points or terminal nodes. The decision points each have a set of actions  $A_s$ . (At terminal nodes, there are no more actions to be taken [e.g., the end of the maze], so  $A_s = \emptyset$ .) The decision points capture all the moments where the person could do meaningfully different things. For instance, Lenny could carve the maze up into the states in Figure 1, where each fork is a decision point with the actions “left” and “right.”

But note that this carving is subjective:  $S$  and  $A_s$  are not objective features of the external environment, but variables in the person’s head. Lenny could have incorporated the temperature of the room and made separate decision points for “initial fork in hot room” and “initial fork in cold room.” Or he could have codified each physical step as a decision point, with the actions “move left foot forward,” “move right foot forward,” “move left foot backward,” and so on. States and actions can also

be specified at various levels of abstraction.<sup>24</sup> Lenny could conceptualize his actions as “turn left” or “turn right,” but he could also conceptualize them as “turn body 27 degrees counterclockwise,” “move left foot 1.36 centimeters forward,” and so forth. These variables are subjective and, therefore, could be influenced by norms.

MDPs also formalize the consequences of choice. There are two questions to answer after making a choice: What happened, and how good or bad was it? In MDPs, after selecting an action at a decision point, an agent transitions to a new state (either another decision point, or a terminal node). This process is captured by a noisy transition function  $T(s, a, s')$ , which gives the probability of transitioning to any new state  $s'$  after choosing an action  $a$  at a decision point  $s$ . This is the answer to the question, “What happened?”

### RL Glossary

*Reward:* A subjective measure that the agent is attempting to maximize. In computer science this is defined by the programmer; in biological systems, it might be a consequence of biological or cultural evolution.

*State:* An agent’s representation of the present circumstances relevant to decision-making.

*Action:* A behavioral response available to an agent in its current state.

*Transition function:* A function describing the probability of attaining a particular subsequent state given the current state and action.

*Policy:* The current likelihood with which the agent will decide to perform any given action in any given state — i.e., a complete probabilistic description of what the agent tends to do.

*Value:* The future reward that an agent expects to obtain, conditioned upon a given action and its current state and policy. That is, “if I do X, and given the actions I am likely to take subsequently, how much reward can I expect to obtain?”

The goal of reinforcement learning is to accurately estimate the value of actions (“learning”), and then to choose actions that tend to have the highest value (“decision-making”). More formally, the agent wishes to converge on a reward-maximizing policy.

Then, the agent receives a reward. In principle, how much the agent likes or dislikes the consequences of a choice could be based on all three variables: where the agent was ( $s$ ), what the agent chose ( $a$ ), and where

<sup>24</sup> Fiery Cushman, “From Moral Concern to Moral Constraint,” *Current Opinion in Behavioral Sciences* 3 (2015): 58–62. <https://doi.org/10.1016/j.cobeha.2015.01.006>; S. J. Gershman and Y. Niv “Learning Latent Structure: Carving Nature at its Joints,” *Current Opinion in Neurobiology* 20, no. 2 (2010): 251–56. <https://doi.org/10.1016/j.conb.2010.02.008>

the agent ended up ( $s'$ ). Think of a daredevil jumping a canyon on a motorcycle. Her reward from this experience will depend on where she was (at the edge of a canyon), what she did (ride a motorcycle forward), and where she happened to end up (on the other side, or at the bottom of the canyon). Thus, to relate the person's objective to the consequences of choice, we introduce a reward function  $R(s, a, s')$ , which gives the amount of reward a person receives for taking action  $a$  at decision point  $s$  and ending up at state  $s'$ . Note that, just like the previous representations, the reward function is subjective: Cheese and shocks are only good and bad in someone's head. Thus, the specification of subjective rewards could also be manipulated by norms.

So MDPs describe decision situations with four variables: the set of states  $S$ , the actions at each decision point  $A_s$ , the transition function  $T$ , and the reward function  $R$ . In plain English, the person needs to know what the states of the world are, what she can do at each state, what next state she will arrive at, and how good or bad the consequences are. All four of these variables could, in principle, be influenced by norms in a way that would promote compliance.<sup>25</sup>

### *B. Discovering solutions: Reinforcement learning*

Reinforcement learning is a branch of computer science that seeks algorithms for learning and deciding within MDPs. At the most abstract level, the optimal approach is for an agent to compute the *value* of each action available at each state, which has a precise definition: the expected sum of future rewards conditioned on choosing that action at the current decision point. Then, the person can choose the action with the highest value. This is analogous to computing the expected utility of a gamble in economic decision theory.<sup>26</sup>

Although value can be defined simply, calculating it is complicated. In a sequential setting such as the MDP, the long-term rewards following from an agent's present decision depend not only on that decision, but also on the *subsequent* decisions that the agent makes. In other words, the agent needs to know something about itself — for a chess algorithm to know which piece to move, it needs to consider what subsequent moves it is likely to implement as the game unfolds. This can be formalized via the

<sup>25</sup> There are two key details about MDPs that we've omitted for simplicity. First, the transition and reward functions, when conditioned on the current decision point, are assumed to be independent of past experience. This restriction is known as the Markov property, and is often attained by simply enhancing the representation of the current state to include all relevant prior factors (R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning* [Cambridge, MA: MIT Press, 1998]). Second, there is also a discount parameter, which controls the rate at which future rewards are discounted relative to current rewards (Sutton and Barto, *ibid.*).

<sup>26</sup> J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior* (Princeton, NJ: Princeton University Press, 1944).



concept of a “policy,” which is a complete probabilistic mapping from state to action; a description of the likelihood of the agent doing each possible thing in each possible set of circumstances. Thus, an agent’s assignment of value to an action must be conditioned not only on the agent’s present state but also on its own policy. From this vantage point, the agent’s overarching goal is to develop an optimal policy.

A consequence of this complexity is that rational decision-making is often intractable. Consequently, much work in reinforcement learning seeks to find computationally efficient mechanisms for approximating value representation. Broadly speaking, many algorithms can be positioned on a spectrum from those that are highly rational (achieving accuracy at the cost of computation) to those that are highly heuristic (sacrificing accuracy to lower computational demands). This distinction forms a major dividing line between rival theories of norm compliance, and so it worth our careful attention.

The standard economic approach defines the rational end of this spectrum: Use an explicit model of the situation to perform the exact calculations. For example, if Lenny knows that turning left twice leads him to cheese and turning right twice leads him to shock (that is, he has a representation of the transition function  $T$ ), he can compute that the expected value of turning left at the initial fork is +1. In MDPs, this is known as goal-directed planning, and captures the kind of forward thinking often associated with human rationality.<sup>27</sup> In the RL literature these methods are typically described as “model-based,” because they require an agent to learn and explicitly represent a model of the transition structure of the environment, which forms the cognitive basis for planning. This approach can be very precise — even perfectly so — but it can also entail major computational costs.

Alternatively, Lenny could approximate the value of turning left at the fork by keeping a running average of the reward he’s received after taking that action in the past. In this case, he wouldn’t turn left because he believed it would lead to cheese; he would turn left simply because it had been a good choice in the past. In other words, Lenny would be acting habitually. This backward-looking method is often associated with irrationality<sup>28</sup> — or, more charitably, with heuristic cognition<sup>29</sup> — and there is extensive evidence that people maintain and rely on these “cached” averages of past reward.<sup>30</sup> In the RL literature, these methods are typically

<sup>27</sup> Jon Elster, “Social Norms and Economic Theory,” *Journal of Economic Perspectives* 3, no. 4 (1989): 99–117; J. C. Harsanyi, “Morality and the Theory of Rational Behavior,” *Social Research* 44, no. 4 (1977): 623–56; D. Kahneman and R. H. Thaler, “Anomalies: Utility Maximization and Experienced Utility,” *The Journal of Economic Perspectives* 20, no. 1 (2006): 221–34. <https://doi.org/10.1257/089533006776526076>

<sup>28</sup> Elster, “Social Norms and Economic Theory,” 99–117.

<sup>29</sup> A. Tversky and D. Kahneman, “Judgment under Uncertainty: Heuristics and Biases,” *Science* 185, no. 4157 (1974): 1124–31. <https://doi.org/10.1126/science.185.4157.1124>

<sup>30</sup> R. J. Dolan and P. Dayan, “Goals and Habits in the Brain,” *Neuron* 80, no. 2 (2013): 312–25. <https://doi.org/10.1016/j.neuron.2013.09.007>

described as “model-free,” because they do not require learning, representing, or planning over a model of the transition structure of the environment; they just rely on averages of past reward. This approach often sacrifices precision but achieves major computational savings. We denote  $Q(s,a)$  as the average value an agent has obtained from choosing action  $a$  in decision point  $s$  in the past.  $Q$ 's are the fifth and final decision variable in the RL framework that could be influenced by norms.<sup>31</sup>

The computational distinction between “model-based” and “model-free” approaches — that is, between habit and planning — helped to crystallize a more fluid set of concepts that have long been fundamental to behavioral research. Psychologists since Thorndike<sup>32</sup> have debated the extent to which complex behavior is the result of planning over an internal representation of the causal structure of the world, or instead the accumulation of adaptive heuristics shaped under the prior history of reward and punishment. RL offers one means of describing the competition between rival cognitive “systems” that are characterized by different tradeoffs of accuracy against computational demand.<sup>33</sup>

The conflict between habit and planning is spelled out in a simple behavioral paradigm that mirrors our example of Lenny. A rat is trained to press a lever in order to obtain a food reward. Then, it is removed from the box and given free access to as much of its favorite foods as it wants, until it is completely satiated and shows no further interest in the food. Finally, the satiated rat is returned to the box with the lever. If the rat had received small amounts of training, it will not push the lever because it knows it doesn't want the food (that is, it engages in planning). However, if it was overtrained and formed a habit of pressing the lever, it will move to the lever and continually press it, ignoring the food that accumulates on the floor.<sup>34</sup> The irrational persistence in an action that was previously reinforced is the hallmark of habitual control of behavior. Reinforcement learning offers a means of making this distinction between habit and planning computationally precise, and assimilates a range of previously disparate decision-making models into a common framework.<sup>35</sup>

<sup>31</sup> This method of habit learning by reinforcing actions historically associated with reward emerged in the reinforcement learning literature in the 1980s and rapidly revolutionized the field, quickly enabling human-level proficiency in games like backgammon. Several computational signatures of these model-free RL algorithms were also discovered in dopaminergic neural circuits that implement value-guided learning and decision-making, catalyzing two decades of rapid theoretical and empirical advances (W. Schultz, P. Dayan, and P. R. Montague, “A Neural Substrate of Prediction and Reward,” *Science* 275, no. 5306 [1997]: 1593–99. <https://doi.org/10.1126/science.275.5306.1593>).

<sup>32</sup> E. L. Thorndike, “The Law of Effect,” *The American Journal of Psychology* 39 (1927): 212–22. <https://doi.org/10.2307/1415413>

<sup>33</sup> D. Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus, and Giroux, 2011).

<sup>34</sup> A. Dickinson, B. Balleine, A. Watt, F. Gonzalez, and R. A. Boakes, “Motivational Control after Extended Instrumental Training,” *Animal Learning and Behavior* 23, no. 2 (1995): 197–206. <https://doi.org/10.3758/BF03199935>

<sup>35</sup> Dolan and Dayan, “Goals and Habits in the Brain,” 312–25.

### D. Summary

Together, MDPs and RL offer a powerful model of decision situations, with five variables ( $S$ ,  $A_s$ ,  $T$ ,  $R$ , and  $Q$ ) that could be influenced by norms (Figure 2). This framework will form the backbone of our discussion. But it is important to emphasize again that this is not the only viable model of decisions. There are, for example, many models that forgo the notions of reward and value, and don't assume all objectives can be mapped onto a common currency.<sup>36</sup> We adopt a reward-based framework because it is widespread in both cognitive decision theory and theories of norms, but future work might pursue these alternatives. There are also models of norm compliance that forgo any notion of mental representations at all.<sup>37</sup> While nonrepresentational models have much to contribute, our goal is to see how much traction cognitive models can gain.

## III. A COMMON LANGUAGE FOR THEORIES OF NORM COMPLIANCE

We use the framework of MDPs and RL algorithms to couch four major theories of norm compliance in a common language (Figure 2).

### A. Compliance by planning: A folk theory of norm influence

The simplest theory of norm influence was stated by Glaucon in Plato's *Republic*: People "are only diverted into the path of justice by . . . force."<sup>38</sup> Paraphrasing: People follow norms because they fear reprisal, or a loss of reputation. Although Glaucon's specific theory focused on punishment, we are concerned not with the *content* of people's expectations ("I'll go to jail"), but rather the *form* of the influence upon decision-making. The claim is that people obey norms because they explicitly represent the likely outcomes of compliance versus noncompliance, and conclude that the likely outcomes of compliance hold a greater prospect of maximizing long-term reward. Thus, this theory encompasses a wide range of specific motivations: Avoiding jail, earning respect, winning friends, maximizing profit, and so on. This view is common enough to be called the folk theory of norm influence, and is the launching point for our discussion.

<sup>36</sup> S. Lichtenstein and P. Slovic, *The Construction of Preference* (New York: Cambridge University Press, 2006); I. Vlaev, N. Chater, N. Stewart, and G. D. A. Brown, "Does the Brain Calculate Value? *Trends in Cognitive Sciences* 15, no. 11 (2011): 546–54. <https://doi.org/10.1016/j.tics.2011.09.008>

<sup>37</sup> J. L. McClelland, D. E. Rumelhart, and G. E. Hinton, "Parallel Distributed Processing: Explorations in the Microstructure of Cognition," Vol. 1, in D. E. Rumelhart, J. L. McClelland, and C. PDP Research Group, eds., (Cambridge, MA: MIT Press, 1986), 3–44. Retrieved from <http://dl.acm.org/citation.cfm?id=104279.104284>

<sup>38</sup> B. Jowett, "The Dialogues of Plato," *Journal of Hellenic Studies* 45, no. 4 (1925): 274.

| <u>Variables</u>                   | <u>Theories</u>                      |
|------------------------------------|--------------------------------------|
| Transition function<br>$T(s,a,s')$ | <p>Folk Theory</p>                   |
| Cached value estimates<br>$Q(s,a)$ | <p>Habit Theory</p>                  |
| Reward function<br>$R(s,a,s')$     | <p>Internalization Theory</p>        |
| Action set<br>$A_s$                | <p>Choice set restriction theory</p> |
| State space<br>$S$                 | No theory yet                        |

FIGURE 2. Five cognitive decision variables that norms could influence, and the theory of norm compliance that corresponds to each variable. On the folk model, norms change the decision maker's internal causal model of the world ("if I cheat on my taxes, I'll go to jail"). On the habit model, norms change the stored values of options via day-to-day reward and punishment of obedience or disobedience ("every time I cheated on things like taxes in the past, bad things happened"). On the internalization model, norms affect the intrinsic reward that people assign to different outcomes ("it's wrong to not pay your fair share of taxes"). And on the unthinkable action model, norms change which actions are even considered ("I didn't even think to cheat on my taxes").

Construed within the RL framework, the folk theory is comprised of two key claims:

1. The existence of the norm changes the agent's *expectation* that future reward will be maximized via norm compliance, by altering the agent's internal representation of the transition function  $T$ .
2. The existence of the norm does not change the agent's *assignment* of reward (that is, the reward function  $R$ ). In other words, norm compliance has value to the agent exclusively because it is linked to other rewards that the agent already holds (such as food, pain, and so forth).

In the sections that follow, we will review alternative models of norm compliance that depart from the folk theory in two different ways. One claims that norms influence behavior not by planning over an explicit model of future reward, but rather by heuristics and habits established by the prior history of reward (that is, norms affect  $Q$ ). The other claims that norms change not just our expectation of reward but rather the very set of events that we find rewarding at all (that is, norms affect  $R$ , for example, by "internalization" of the norm as an intrinsic good). Thus, the folk theory stands apart from these alternatives principally in its claims of *planning* over an *unmodified* specification of reward.

It is uncontroversial that the folk theory captures at least part of the manner in which norms influence human behavior. At least some of the time, people pay taxes because they fear an audit; vote because they want to influence the outcome; tip because they want good service; recycle because they want to burnish their social image; and so on. But it is equally uncontroversial that the folk theory is incomplete. The problem is that people routinely obey norms even when they know there is little chance of reprisal. They report their taxes honestly when the risk of an audit is incredibly low; vote in elections that they have no chance of influencing; tip in restaurants they will never revisit; recycle, even when living alone; and so on.<sup>39</sup>

The persistence of norm compliance in the absence of explicit expectations of enforcement has been most carefully demonstrated with laboratory economic games.<sup>40</sup> In these games, people are given an endowment of bonus money and asked to make decisions that will affect both their own payment and the payment of other subjects. For example, people are sometimes asked whether they want to contribute to a public pool that will benefit everyone else but cost them personally. Other times they are asked to split the endowment with another subject. Time after time, most

<sup>39</sup> Sunstein, "Social Norms and Social Roles," 903–968.

<sup>40</sup> C. Camerer, *Behavioral Game Theory: Experiments in Strategic Interaction* (Princeton, NJ: Princeton University Press, 2003).

participants at least sometimes obey norms prescribing fairness, cooperation, reciprocity, or honesty. Importantly, they follow the norms even when the game is one-shot and anonymous — in other words, when non-compliance is guaranteed to go unpunished.<sup>41</sup>

We call this phenomenon — people’s dogged persistence in obeying norms — the *persistence problem*. People don’t always persist in norm compliance, but they do it often enough to make us question whether explicit expectations of reward are truly the driving force. The folk theory can be thought of as a null hypothesis, and the persistence problem the data that force us to consider the more complex alternatives that follow.

There are a few concerns with this analysis that will be useful to address before moving to more complex theories. First, is the persistence problem really a problem, or just an ordinary mistake? People could simply be failing to notice that they are in a situation where punishment doesn’t apply, or fail to incorporate this knowledge into their representation of the transition function.<sup>42</sup> Though this objection is difficult to rule out in real-world examples, the simplicity and transparency of countless laboratory experiments makes it unlikely that subjects are not representing the contextual one-shot anonymity.<sup>43</sup> Moreover, people *do* often incorporate this information into their planning, increasing compliance when people are observing them or can reciprocate.<sup>44</sup> The interesting fact is that their levels of norm compliance are far from zero even in the one-shot anonymous case.

Second, we have emphasized examples like tax reporting, recycling, and one-shot economic games to draw inferences about norm compliance. But how do we know that these are actually examples of norm compliance? People might have performed these behaviors in the absence of exposure to any norm (for example, because of evolved instinct). Again, while this is difficult to rule out in the real world, numerous experiments show that norms drive people’s behavior in anonymous situations. People exhibit enormous cross-cultural variation in one-shot economic games, and this variation is partially explained by differences in cultural norms.<sup>45</sup>

<sup>41</sup> Fehr and Schmidt, “The Economics of Fairness, Reciprocity and Altruism,” 615–91.

<sup>42</sup> A. W. Delton, M. M. Krasnow, L. Cosmides, and J. Tooby, “Evolution of Direct Reciprocity under Uncertainty Can Explain Human Generosity in One-Shot Encounters,” *Proceedings of the National Academy of Sciences* 108, no. 32 (2011), 13335–40. <https://doi.org/10.1073/pnas.1102131108>

<sup>43</sup> J. Andreoni, “Cooperation in Public-Goods Experiments: Kindness or Confusion?” *American Economic Review* 85, no. 4 (1995): 891–904; Fehr and Schmidt, “The Economics of Fairness, Reciprocity and Altruism”; Thaler “Anomalies: The Ultimatum Game,” 195–206.

<sup>44</sup> Fehr and Schmidt, “The Economics of Fairness, Reciprocity and Altruism.”

<sup>45</sup> J. Henrich, “Does Culture Matter in Economic Behavior? Ultimatum Game Bargaining among the Machiguenga of the Peruvian Amazon,” *American Economic Review* 90, no. 4 (2000): 973–79; J. Henrich, R. McElreath, A. Barr, J. Ensminger, C. Barrett, A. Bolyanatz, J. Ziker, “Costly Punishment Across Human Societies,” *Science* 312, no. 5781 (2006): 1767–70. <https://doi.org/10.1126/science.1127333>; J. Henrich, J. Ensminger, R. McElreath, A. Barr, C. Barrett, A. Bolyanatz, J. Ziker, “Markets, Religion, Community Size, and the Evolution of Fairness and Punishment,” *Science* 327, no. 5972 (2010): 1480–84. <https://doi.org/10.1126/science.1182238>

And manipulating norms changes behavior. People walking alone rarely litter after seeing anti-littering signs, but litter quite often after seeing piles of trash on the ground.<sup>46</sup> People reuse hotel towels more often, and give away more money in economic games, when told it is the norm.<sup>47</sup> All this suggests that people are influenced by norms, even when they won't be sanctioned for disobedience — and thus the folk theory of norm compliance is incomplete.

To motivate our exploration of the theory space, we will repeatedly draw on these kinds of examples as genuine instances of norm-following. The key assumption in each instance is that the aggregate behavior — a tendency to offer fair splits of a monetary endowment, contribute to a public good, return an investment, and so forth — is in part driven by prior exposure to norms. Thus, from patterns in the behavior, we can infer features of how norms influence us. In many cases, there is experimental (and anecdotal) evidence to support this assumption. But in other cases, it has not been explicitly tested.<sup>48</sup> We will highlight claims that demand more rigorous experimental support from future work.

Finally, it is important to note that our review of literature will be dramatically incomplete. We focus on studies that we believe offer the most precision and insight into the cognitive underpinnings of compliance, and, in doing so, pass over much important work in sociology and other fields. We hope to offer a cognitive perspective that naturally complements these alternative approaches.

<sup>46</sup> R. B. Cialdini, C. A. Kallgren, and R. R. Reno, "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior," in M. P. Zanna ed., *Advances in Experimental Social Psychology*, vol. 24 (Waltham, MA: Academic Press, 1991), 201–234. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0065260108603305>; R. B. Cialdini, R. Reno, and C. A. Kallgren, "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places," *Journal of Personality and Social Psychology* 58, no. 6 (1990): 1015–26. <https://doi.org/10.1037/0022-3514.58.6.1015>

<sup>47</sup> N. J. Goldstein, R. B. Cialdini, and V. Griskevicius, "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels," *Journal of Consumer Research* 35, no. 3 (2008): 472–82. <https://doi.org/10.1086/586910>; N. J. Raihani and K. McAuliffe, "Dictator Game Giving: The Importance of Descriptive versus Injunctive Norms," *PLOS ONE* 9, no. 12 (2014), e113826. <https://doi.org/10.1371/journal.pone.0113826>

<sup>48</sup> Many studies show that manipulating a variable (e.g., time available to make a decision) makes people more or less cooperative, fair, generous, and so on, in economic games. These choices are likely influenced by the norms to which people have been exposed (Fehr and Fischbacher, "Social Norms and Human Cooperation"; D. G. Rand, A. Peysakhovich, G. T. Kraft-Todd, G. E. Newman, O. Wurzbacher, M. A. Nowak, and J. D. Greene, "Social Heuristics Shape Intuitive Cooperation," *Nature Communications* 5, no. 3677 [2014]. <https://doi.org/10.1038/ncomms4677>). But ideally, to show that the specific variable at hand affects norm compliance, the study would simultaneously manipulate whether the relevant norm is present, and show that the variable only has an effect when the norm is present. Unfortunately, the ideal study has often not yet been run. We cite the imperfect studies with the hope that future work will fill in the gaps.

*B. Compliance by habit: Heuristic approaches to value approximation*

The folk theory assumes that people are representing their environment's transition function ("if I turn right at the fork, it will lead to cheese"), and selecting actions via forward planning ("I want cheese, therefore I should turn right"). They know the consequences of their choices, and act on that knowledge. This assumption is what causes trouble for the folk theory: People seem to follow norms even when they know the typical motivating consequences are suspended (for instance, in one-shot anonymous settings).

There are two ways out of this bind. Perhaps the consequences that motivate norm compliance are not sanctions or reputation, but an intrinsic "social preference," such as promoting fairness or being a good person. This solution to the persistence problem depends on altering the reward function  $R$ , and it is pursued in the next section.

First, though, there is a simpler alternative. As described above, people don't always plan ahead; they often act out of habit. More formally, instead of representing the transition function  $T$  (that is, the explicit consequences of actions), they store the average value  $Q$  of past actions, and simply choose actions with higher  $Q$ s — actions that accumulated higher reward in the past. A habitual Lenny turns left in the maze, not because he knows it will lead to cheese, but because he remembers that it led to an average goodness of  $+1$  in the past. If you told him that the cheese had moved, it wouldn't affect his immediate decision-making, because he's not even considering cheese when making his decision.

Similarly, a habitual human might propose a fair monetary split in an economic game, not because she knows it will avoid sanctions and improve her reputation, but because she merely remembers that being fair led to an "average goodness" of  $+1$  in the past. If you told her that the game was one-shot/anonymous and she couldn't be sanctioned, it wouldn't affect her immediate decision-making, because she's not even considering sanctions when making her decision.

We call this the habit theory of norm compliance. In everyday life, most situations are not one-shot or anonymous, and complying with norms is typically beneficial (to avoid sanctions, accrue reputation, and so on). This experience produces high-cached  $Q$  values for norm-compliant actions. Then, when people are in one-shot anonymous settings where the typical benefits are removed, they choose based on those  $Q$  values, which are not immediately affected by the knowledge of anonymity. Thus, people persist in norm compliance.<sup>49</sup>

<sup>49</sup> D. G. Rand and Z. G. Epstein, "Risking Your Life without a Second Thought: Intuitive Decision-Making and Extreme Altruism," *PLOS ONE* 9, no. 10 (2014): e109687. <https://doi.org/10.1371/journal.pone.0109687>; D. G. Rand, J. D. Greene, and M. A. Nowak, "Spontaneous Giving and Calculated Greed," *Nature* 489, no. 7416 (2012): 427–30. <https://doi.org/10.1038/nature11467>; Rand, Peysakhovich, Kraft-Todd, Newman, Wurzbacher, Nowak, and Greene, "Social Heuristics Shape Intuitive Cooperation."



One point in favor of the habit theory is that people appear habitually averse to causing direct personal harm, even in contexts where it maximizes welfare.<sup>50</sup> In contemporary societies, there are strong norms against causing typical forms of harm: pushing, hitting, shooting, stabbing, and so on. The habit theory predicts that people will therefore have developed a habitized aversion — that is, a low Q value — to performing those behaviors. In support of this, people are unwilling to perform these actions (for example, pushing someone in front of a runaway trolley), even in the presence of a causal model linking such an action to positive consequences (for example, saving five other people<sup>51</sup>). Moreover, people exhibit physiological signs of aversion to pretend actions that resemble acts of physical harm, such as pulling the trigger of a fake replica handgun while pointing it at a person's head.<sup>52</sup> Like the rat who keeps pushing its food lever even when full, people seem to be relying on a cached value that they have assigned to certain norm-violating behaviors — even when the typical consequences of those behaviors are suspended.<sup>53</sup>

The habit theory of norm compliance is also supported by a suite of experimental results motivated by the fact that acting on habits requires less computation time/resources than planning. (Planning requires the agent to integrate over a potentially complex transition function, while acting on habits just requires a comparison among pre-computed Q values.) Thus, according to the habit theory, people with less time or mental resources should be more compliant.

Indeed, when people in economic games are put under time pressure or cognitive load, they become more cooperative,<sup>54</sup> fair,<sup>55</sup> reciprocal,<sup>56</sup>

<sup>50</sup> M. J. Crockett, "Models of Morality," *Trends in Cognitive Sciences* 17, no. 8 (2013): 363–66. <https://doi.org/10.1016/j.tics.2013.06.005>; Fiery Cushman, "Action, Outcome, and Value: A Dual-System Framework for Morality," *Personality and Social Psychology Review* 17, no. 3 (2013): 273–92. <https://doi.org/10.1177/1088868313495594>

<sup>51</sup> J. D. Greene, R. B. Somerville, L. E. Nystrom, J. M. Darley, and J. D. Cohen, "An fMRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293, no. 5537 (2001): 2105–2108. <https://doi.org/10.1126/science.1062872>

<sup>52</sup> F. Cushman, K. Gray, A. Gaffey, and W. B. Mendes, "Simulating Murder: The Aversion to Harmful Action," *Emotion* 12, no. 1 (2012): 2–7. <https://doi.org/10.1037/a0025071>

<sup>53</sup> Fiery Cushman, "From Moral Concern to Moral Constraint," *Current Opinion in Behavioral Sciences* 3 (2015): 58–62. <https://doi.org/10.1016/j.cobeha.2015.01.006>

<sup>54</sup> D. G. Rand, "Cooperation, Fast and Slow Meta-Analytic Evidence for a Theory of Social Heuristics and Self-Interested Deliberation," *Psychological Science* (2016), 0956797616654455. <https://doi.org/10.1177/0956797616654455>

<sup>55</sup> D. Cappelletti, W. Güth, and M. Ploner, "Being of Two Minds: Ultimatum Offers under Cognitive Constraints," *Journal of Economic Psychology* 32, no. 6 (2011): 940–50. <https://doi.org/10.1016/j.joep.2011.08.001>; E. Halali, Y. Bereby-Meyer, and A. Ockenfels, "Is It All about the Self? The Effect of Self-Control Depletion on Ultimatum Game Proposers," *Frontiers in Human Neuroscience* 7 (2013). <https://doi.org/10.3389/fnhum.2013.00240>

<sup>56</sup> C. Anderson and D. L. Dickinson, "Bargaining and Trust: The Effects of 36-H Total Sleep Deprivation on Socially Interactive Decisions," *Journal of Sleep Research* 19 (2010): 54–63. <https://doi.org/10.1111/j.1365-2869.2009.00767.x>; V. Grimm, and F. Mengel, "Let Me Sleep on It: Delay Reduces Rejection Rates in Ultimatum Games," *Economics Letters* 111, no. 2 (2011): 113–15. <https://doi.org/10.1016/j.econlet.2011.01.025>; E. Halali, Y. Bereby-Meyer,

and—sometimes—more generous.<sup>57</sup> This finding resonates with real-world reports of extreme prosociality: People who risk their lives to help others almost always describe their decision process as quick and intuitive (like acting on a habit), not slow and reflective (like planning).<sup>58</sup>

Moreover, the effect of time pressure/cognitive load in economic games is dependent on prior experience: It is strongest for people who report positive cooperative experiences in their daily life,<sup>59</sup> and disappears with extensive experience in one-shot anonymous economic games.<sup>60</sup> Both of these findings are explained by the habit theory. Only people who were rewarded for complying with norms in the past should form a habit of it. And, while  $Q$  values are not *immediately* affected by contextual anonymity, they will be in the long run: repeated experience in a context where compliance is not beneficial will drive down the  $Q$  values until compliance is no longer a habit. Imagine we move Lenny's cheese. If Lenny keeps habitually turning left in the maze, eventually the average reward for turning left will plummet, and Lenny will stop turning left. Similarly, if people

---

and N. Meiran, *When Rationality and Fairness Conflict: The Role of Cognitive-Control in the Ultimatum Game* (SSRN Scholarly Paper No. ID 1868852) (Rochester, NY: Social Science Research Network, 2011). Retrieved from <http://papers.ssrn.com.ezp-prod1.hul.harvard.edu/abstract=1868852>; E. Halali, Y. Bereby-Meyer, and N. Meiran, "Between Self-Interest and Reciprocity: The Social Bright Side of Self-Control Failure," *Journal of Experimental Psychology: General* 143, no. 2 (2014): 745–54. <https://doi.org/10.1037/a0033824>; W. S. Neo, M. Yu, R. A. Weber, and C. Gonzalez "The Effects of Time Delay in Reciprocity Games," *Journal of Economic Psychology* 34 (2013): 20–35. <https://doi.org/10.1016/j.joep.2012.11.001>; M. Sutter, M. Kocher, and S. Straub, "Bargaining under Time Pressure in an Experimental Ultimatum Game," *Economics Letters* 81, no. 3 (2003): 341–47. [https://doi.org/10.1016/S0165-1765\(03\)00215-5](https://doi.org/10.1016/S0165-1765(03)00215-5)

<sup>57</sup> J. F. Schulz, U. Fischbacher, C. Thöni, and V. Utikal, "Affect and Fairness: Dictator Games under Cognitive Load," *Journal of Economic Psychology* 41 (2014): 77–87. <https://doi.org/10.1016/j.joep.2012.08.007>. The effect of time pressure and cognitive depletion on generosity is mixed. Some studies report that they induce more giving (G. Cornelissen, S. Dewitte, and L. Warlop, "Are Social Value Orientations Expressed Automatically? Decision Making in the Dictator Game," *Personality and Social Psychology Bulletin*, 0146167211405996 [2011]. <https://doi.org/10.1177/0146167211405996>; Schulz, Fischbacher, Thöni, and Utikal, "Affect and Fairness," 77–87), while others report a null or (rarely) reversed effect (K. E. Hauge, K. A. Brekke, L. O. Johansson, O. Johansson-Stenman, and H. Svedsäter, "Keeping Others in Our Mind or in Our Heart? Distribution Games under Cognitive Load," *Experimental Economics* 19, no. 3 [2015], 562–76. <https://doi.org/10.1007/s10683-015-9454-z>; Halali, Bereby-Meyer, and Ockenfels, "Is It All about the Self?"). Interestingly, the game used in these studies to measure generosity has notoriously fickle norms (Bicchieri, *The Grammar of Society*; Fehr and Schmidt, *The Economics of Fairness, Reciprocity, and Altruism*; E. L. Krupka and R. A. Weber, "Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary?" *Journal of the European Economic Association* 11, no. 3 (2013): 495–524. <https://doi.org/10.1111/jeea.12006>);. Perhaps the ambiguous effects of time pressure on generosity can be explained by differences in norm perception across studies.

<sup>58</sup> Rand and Epstein, "Risking Your Life without a Second Thought," e109687.

<sup>59</sup> *Ibid.*; Rand, Greene, and Nowak, "Spontaneous Giving and Calculated Greed," 427–30; Rand, Peysakhovich, Kraft-Todd, Newman, Wurzbacher, Nowak, and Greene, "Social Heuristics Shape Intuitive Cooperation.

<sup>60</sup> Rand and Epstein, "Risking Your Life without a Second Thought: Intuitive Decision-Making and Extreme Altruism," e109687; Rand, Peysakhovich, Kraft-Todd, Newman, Wurzbacher, Nowak, and Greene, "Social Heuristics Shape Intuitive Cooperation."

keep playing one-shot anonymous economic games, their average stored reward for complying with norms will drop, they will stop complying out of habit, and thus the effect of time pressure will disappear. These experimental results therefore support the habit theory of norm compliance.<sup>61</sup>

In sum, habit theory likely explains some part of norm compliance. But it is also incomplete. The problem is that, even after people have enough experience with one-shot anonymous settings to lose their habit for compliance, they continue to comply. In the real world, experienced travelers still tip in foreign cities; experienced solo-hikers still avoid littering when alone in public parks; hotel guests still reuse towels on the tenth day of their stay. In all these cases, if people were acting out of a habit developed in non-anonymous settings, experience with the anonymous setting should eliminate compliance — but it doesn't.

This phenomenon is most precisely demonstrated with laboratory games. As described above, people who participate in many economic games seem to lose their habit for compliance; they stop showing an effect of time pressure on compliance choices. Yet they continue to contribute to public goods, propose fair monetary splits, return investments, and so on, at almost the same rate<sup>62</sup> — it just no longer matters whether they have a short or a long amount of time to do so. The persistence of norm-compliant behavior even in the face of massive experience with one-shot, anonymous contexts is difficult to explain under the habit theory.

In sum, the habit theory appears only to explain “shallow” persistence: people comply with norms in the first few anonymous situations they encounter. But it does not naturally account for “deep” persistence: people continue to comply with norms, even after they have enough experience with anonymous situations to kick the habit. Norms seem to do more than just instill habits.

### C. *Compliance by internalization: Norms as sources of intrinsic reward*

A popular idea, hinted at before, is that norm compliance may sometimes be valued intrinsically, rather than as an instrumental means of avoiding

<sup>61</sup> It is possible that, under time pressure, people do not become more compliant; they simply become more prosocial (Rand, Greene, and Nowak, “Spontaneous Giving and Calculated Greed”). The fact that people are also more-negatively reciprocal under time pressure suggests the former interpretation. But it is an open question. One study appeared to show that, even after being instilled with a norm for competition instead of cooperation, people were still more cooperative under time pressure (J. Cone, and D. G. Rand, “Time Pressure Increases Cooperation in Competitively Framed Social Dilemmas,” *PLOS ONE* 9, no. 12 [2014]: e115756. <https://doi.org/10.1371/journal.pone.0115756>.) But people in the competitive condition did not, on average, contribute less than people in the cooperative condition (the difference was around one cent out of an endowment of forty), suggesting that the norm manipulation was insufficient.

<sup>62</sup> Rand, “Cooperation, Fast and Slow Meta-Analytic Evidence for a Theory of Social Heuristics and Self-Interested Deliberation”; Rand, Greene, and Nowak, “Spontaneous Giving and Calculated Greed,” *Nature*, 427–30.

punishment or attaining reward. Through socialization, people internalize the norm and come to care about the things that norms prescribe. People are fair, not because they know unfairness would be punished, but because they care about producing fair outcomes.<sup>63</sup> They contribute to public goods, not because their contributions are being externally tracked, but because they believe it's the right thing to do.<sup>64</sup> If this were true, then of course people would continue to comply with norms in anonymous settings — anonymity doesn't remove the consequences they care about.

This concept can be made precise within the reinforcement-learning framework by positing that exposure to norms changes people's reward functions,  $R$ .<sup>65</sup> Recall that the function of reinforcement learning is to estimate the instrumental *value* of an action based on the long-term expectation of intrinsic *reward* in subsequent states. Whereas the folk theory and habit theory posit that norms affect behavior by altering the way we perceive the instrumental value of compliance, the internalization theory posits that norms affect behavior by altering the very set of states or actions we find intrinsically rewarding.

This theory naturally accounts for deep persistence without requiring us to abandon the possibility of rational planning.<sup>66</sup> In the language of MDPs, the dictates of the norm ("split goods fairly") have just been internalized as another part of people's reward functions (for example, +1 for proposing a fair split in an economic game), which can be flexibly and rationally traded off against other parts of the reward function during planning. Internalization theory also explains the role of guilt (and other negative emotions) in preventing norm violations: guilt is part of the internalized negative reward for violating norms.<sup>67</sup>

Although the internalization theory can be stated in terms of MDPs, it is important to note that it falls outside the scope of the RL algorithms typically pursued in machine learning contexts. Ordinarily, a reinforcement-learning agent operates with an unchangeable reward function defined by the programmer. It should be obvious why: If the programmer specifies that winning chess is the reward for an agent, it would not want that agent to later redefine rewards around, say, making pretty patterns with the pawns. Yet this, more or less, is what internalization theory posits: that

<sup>63</sup> E. Fehr and K. M. Schmidt, "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics* 114, no. 3 (1999): 817–68.

<sup>64</sup> R. B. Cialdini and M. R. Trost, "Social Influence: Social Norms, Conformity and Compliance," in D. T. Gilbert, S. T. Fiske, and G. Lindzey, eds., *The Handbook of Social Psychology*, Vols. 1 and 2, 4th ed. (New York: McGraw-Hill, 1998), 151–92; Talcott Parsons and Edward Shils, *Toward a General Theory of Action* (Charleston, SC: Nabu Press, 2011), sec. 1.

<sup>65</sup> M. K. Ho, J. MacGlashan, M. L. Littman, and F. Cushman, "Social Is Special: A Normative Framework for Teaching with and Learning from Evaluative Feedback," *Cognition* (2017). <https://doi.org/10.1016/j.cognition.2017.03.006>

<sup>66</sup> E. Fehr and U. Fischbacher, "The Nature of Human Altruism," *Nature* 425, no. 6960 (2003): 785–91. <https://doi.org/10.1038/nature02043>

<sup>67</sup> Fehr and Schmidt, "The Economics of Fairness, Reciprocity and Altruism."

humans initially designed to maximize one thing (for instance, maximum food for themselves) might later autonomously decide to try to maximize something else (for instance, equal food for all). An interesting area of contemporary research seeks to understand why natural selection might favor such an architecture, in which reward itself is subject to evaluation and change.<sup>68</sup>

Although the internalization theory does some violence to traditional assumptions of reinforcement learning, its commitments can also be made much more precise within the RL framework. For instance, there is disagreement in the literature over *what*, precisely, gets internalized during norm learning. Some theories posit that humans learn to intrinsically value certain states of affairs, such as an equitable division of resources among people.<sup>69</sup> Other theories posit that humans learn to intrinsically value certain actions, such as the act of generating an equitable division.<sup>70</sup> These have been regarded as competitors, yet often the precise commitments of various theories are not transparent.

A theory of norm internalization couched in the language of MDPs is naturally committed to precise answers to all of these questions, simply by specifying a reward function. The reward function  $R$  takes as input the current state  $s$ , the chosen action  $a$ , and the subsequent state  $s'$ . If norm internalization made the outcome of compliance rewarding (us receiving an equal amount of money), then  $R$  would be positive for an  $s'$  with equal splits, for any  $a$ .<sup>71</sup> If it made the act rewarding, then  $R$  would be positive for an  $a$  of proposing an equal split, for any  $s'$ .<sup>72</sup> If the reward was for specific norms, then  $R$  would be defined over a set of concrete actions  $a$ ; if the reward was for following norms in the abstract, then  $R$  would be defined over an abstract action space of “follow the norm” or “don’t follow the norm.”<sup>73</sup> If context mattered, then  $R$  would be positive for only certain states  $s$ . MDPs can precisely model all of these situations,

<sup>68</sup> David Ackley and Michael Littman, “Interactions Between Learning and Evolution,” in *Artificial Life II, SFI Studies in the Sciences of Complexity*, vol. X, ed. C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen (London: Addison-Wesley, 1991); Ho, MacGlashan, Littman, and Cushman, “Social Is Special”; Satinder Singh, Richard Lewis, and Andrew Barto, “Where Do Rewards Come From?” *Proceedings of the Annual Conference of the Cognitive Science Society* (2009): 2601-2606.

<sup>69</sup> Fehr and Schmidt “A Theory of Fairness, Competition, and Cooperation,” 817–68.

<sup>70</sup> J. Andreoni, “Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving,” *The Economic Journal* 100, no. 401 (1990): 464–77. <https://doi.org/10.2307/2234133>

<sup>71</sup> Fehr and Schmidt, “A Theory of Fairness, Competition, and Cooperation.”

<sup>72</sup> Andreoni, “Impure Altruism and Donations to Public Goods.”

<sup>73</sup> J. Dana, D. M. Cain, and R. M. Dawes, “What You Don’t Know Won’t Hurt Me: Costly (But Quiet) Exit in Dictator Games,” *Organizational Behavior and Human Decision Processes* 100, no. 2 (2006): 193–201. <https://doi.org/10.1016/j.obhdp.2005.10.001>; Krupka and Weber “Identifying Social Norms Using Coordination Games”; R. López-Pérez, “Aversion to Norm-Breaking: A Model,” *Games and Economic Behavior* 64, no. 1 (2008): 237–67. <https://doi.org/10.1016/j.geb.2007.10.009>

showcasing the breadth of ideas that fall under the umbrella of the reward theory of compliance.

#### *D. A challenge to value-guided theories*

Each of the theories that we have considered so far shares a common assumption: norms intervene on decision-making by altering the value of actions under consideration, as they compete against alternatives in a common currency. This would appear to be a basic commitment of any theory of norm compliance grounded in value-guided decision-making.

An important challenge to this perspective is that people sometimes disregard, or are highly insensitive to, the costs and benefits of complying with or violating these norms. For example, many people say there is no amount of benefit that would justify selling organs, auctioning off brides, or mandating abortions to limit population growth.<sup>74</sup> If a hospital director is deciding whether to perform a surgery to save a child's life, he should not consider the surgery's cost;<sup>75</sup> if a soldier is deciding whether to obey orders, she should not consider the benefits of disobedience.<sup>76</sup> These thought experiments about "taboo tradeoffs" or "incommensurate values" are so compelling that some philosophers consider the inability to trade off compliance with other goods to be the essence of moral norms.<sup>77</sup> This suggests that the motivation to comply with moral norms is not always of a common currency with other goods, and is not simply incorporated into people's reward functions.

Two objections to this line of argument immediately come to mind. First, perhaps people place arbitrarily large negative reward on the outcome of certain norm violations, accounting for the apparent "unthinkable" nature of noncompliance within a value-guided framework. For example, they might place infinitely negative reward on a state in which abortions are mandated, or organs are sold on markets. Thus, any benefit of violating the norm would be outweighed by the cost, and people would be unwilling to violate it. This interpretation appears to reconcile protected values with the reward theory of norm compliance, obviating the need for an alternative theory.

<sup>74</sup> J. Baron and M. Spranca "Protected Values," *Organizational Behavior and Human Decision Processes* 70, no. 1 (1997): 1–16. <https://doi.org/10.1006/obhd.1997.2690>

<sup>75</sup> P. E. Tetlock, O. V. Kristel, S. Beth, M. C. Green, J. S. Lerner, "The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals," *Journal of Personality and Social Psychology* 78, no. 5 (2000): 853–70. <https://doi.org/10.1037/0022-3514.78.5.853>

<sup>76</sup> Raz Joseph, *Practical Reason and Norms* (Oxford: Oxford University Press, 1999).

<sup>77</sup> Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, 1974); Joseph Raz, *Practical Reason and Norms* (Oxford: Oxford University Press, 1999); Frederick Schauer, *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life* (Oxford: Clarendon Press, 1991).

The consilience, however, is misleading. Rewards are common-currency, numerical representations of the subjective, intrinsic goodness in certain states or actions. An essential feature of such representations is that they can be traded off with each other.<sup>78</sup> Saying that a reward is too large to ever be outweighed is equivalent to saying that it is not a reward.

Second, even though people *say* they are unwilling to trade off compliance with other goods, perhaps they still would if the right situation arose. Perhaps their talk of taboos is just signaling (“I would *never* do that!”), or they are not imaginative enough to consider situations with truly enormous benefits of disobedience.<sup>79</sup> After all, when prompted, juries are able to put a price on something as sacred as a human life.<sup>80</sup> More work needs to be done to test whether, in their actual decisions, people are unwilling to put norm compliance on a common scale with other goods.

There is reason, however, to think that people genuinely treat compliance differently from other goods. People often go to absurd lengths to comply with trivial situational norms, far beyond what a reasonable cost-benefit, common-currency analysis would recommend. For instance, subjects in the Milgram experiment chose to fatally electrocute another human rather than disobey norms of proper conduct.<sup>81</sup> Do people value norms of proper conduct more than human life? More likely, the norms were influencing behavior through some non-value-guided channel, thus circumventing this comparison. What is this alternative channel?

#### *E. Compliance by choice set construction: A model of “unthinkable” action*

Here, we suggest one possibility. The basic insight of the “taboo trade-offs” literature is that people sometimes take norm violation off the table completely as an option. When people are unwilling to consider the benefits of organ markets, they are effectively eliminating selling their kidney from the set of choices they consider. On this view, moral norms aren’t just another factor in the expected value of an option; they influence which options even get their expected value calculated in the first place. They change which options get entertained (Figure 2).

This theory can be naturally formulated in the language of MDPs. A crucial variable in an MDP is the set of available actions  $A_s$  at each decision point  $s$ . This variable is subjective, and must be constructed by the decision-maker. It corresponds to what economists sometimes call the “choice set” that a decision-maker constructs prior to choice, narrowing

<sup>78</sup> G. S. Becker, *Accounting for Tastes* (Cambridge, MA: Harvard University Press, 1996).

<sup>79</sup> Tetlock, Kristel, Beth, Green, and Lerner, “The Psychology of the Unthinkable.”

<sup>80</sup> P. A. Ubel, *Pricing Life: Why It’s Time for Health Care Rationing* (Cambridge, MA: MIT Press, 2001).

<sup>81</sup> S. Milgram, “Behavioral Study of Obedience,” *Journal of Abnormal and Social Psychology* 67, no. 4 (1963): 371–78. <https://doi.org/10.1037/h0040525>

from all conceivable actions to a restricted set of considered alternatives. The idea is that norms influence choices in part by excluding norm-violating actions (for instance, selling your kidney) from  $A_s$ , for some broad set of decision points  $s$ . Thus, after being exposed to a norm, people become less likely to violate it.

As before, formulating the theory in the language of MDPs allows us to clarify some of its ambiguities. It would be foolish to claim that moral norms always eliminate norm-violating actions from  $A_s$  — people sometimes do sell their kidneys, and soldiers sometimes do disobey orders. The theory needs some kind of hedge to account for these exceptions. But where exactly should it hedge? One possibility is that the process that generates the action set  $A_s$  is probabilistic, and norms make it less likely (but not impossible) for norm-violating actions to make the cut. Alternatively, being exposed to a norm could cause you to deterministically exclude certain actions from  $A_s$ , but only in certain states  $s$  (for instance, never consider selling your kidney, unless you are in a state of desperation). MDPs can capture both possibilities.

There is also ambiguity in what it means to exclude an action from  $A_s$ . It could mean that the decision maker is aware of the action, but deems it irrelevant. This view is bolstered by the fact that people often deem norm-violating options irrelevant when judging others. For example, consider a captain on a sinking ship who throws either some cargo, or his wife, off the ship. In the latter case, people think the captain acted freely, because he could have thrown the cargo off instead. But in the former case, people think the captain was forced, because there was nothing else he could have done.<sup>82</sup> This belief persists even if people are reminded that he could have thrown his wife off the ship instead; they say that option is irrelevant. This tendency to deem norm-violating options irrelevant seems to appear whenever people judge others' choices.<sup>83</sup> Perhaps people apply the same tendency to their own choices, and deem norm-violating options irrelevant to the decision-making process.

Alternatively, excluding an action from  $A_s$  could mean that the decision-maker is never aware of the action — it doesn't even come to mind. (More precisely, the representation of a norm-violating action ["I should sell my kidney..."] is never tokened in the person's mind.) While this seems unlikely in laboratory experiments with explicit, constrained choice sets (such as "cooperate" or "defect"), real-world decisions typically have choice sets that are unbounded and unstructured (think of all the things you could conceivably do on a given Saturday afternoon). People never

<sup>82</sup> J. Phillips and J. Knobe, "Moral Judgments and Intuitions About Freedom," *Psychological Inquiry* 20, no. 1 (2009): 30–36. <https://doi.org/10.1080/10478400902744279>

<sup>83</sup> J. Phillips, J. B. Luguri, and J. Knobe, "Unifying Morality's Influence on Non-Moral Judgments: The Relevance of Alternative Possibilities," *Cognition* 145 (2015): 30–42. <https://doi.org/10.1016/j.cognition.2015.08.001>



represent the vast majority of the choices they could make, and norms might influence which ones make the cut. The cognitive mechanisms behind choice set construction are extremely unclear, and so it is unclear how norms would produce this effect. But the idea is ripe for future research.

Why would a decision-maker ever exclude actions from consideration? Why not consider all of them? In laboratory experiments with constrained choice sets, it seems implausible that people wouldn't evaluate every option. But again, in the real world, choice sets are enormous. People must have mechanisms for narrowing down their options to a small subset.<sup>84</sup> Since norm-violating options are typically unwise, people would obtain a practical benefit from excluding them.

Moreover, there is a strategic benefit to not even think about violating norms. If other people notice that you considered violating a norm (for example, you took two minutes before deciding to cooperate in a repeated economic game), they will infer that you *would* violate the norm if the incentives were strong enough, and thus terminate productive relationships with you. In other words, in repeated games where your partners know which options you consider, it is strategically beneficial to obey norms "without looking."<sup>85</sup> These considerations further motivate the theory that norms cause people to ignore norm-violating options.

The possibility that norms exclude options from consideration stands out as an important new frontier for research. Some preliminary evidence comes from a study showing that, by default, people tend to treat immoral actions as if they were actually impossible.<sup>86</sup> Only after a little time and thought do they acknowledge that immoral actions are possible, albeit wrong. Although this pattern of response becomes most pronounced in adults subject to time pressure, it is promiscuous in the judgments of young children. For instance, at four years old, a majority of children claim that it is impossible to steal a candy bar from a store — and, moreover, that to succeed would require magic to do so.<sup>87</sup>

<sup>84</sup> Fiery Cushman and Adam Morris, "Habitual Control of Goal Selection in Humans," *Proceedings of the National Academy of Sciences* 112, no. 45 (2015): 13817–13822. <https://doi.org/10.1073/pnas.1506367112>; Q. J. M. Huys, N. Eshel, E. O'Nions, L. Sheridan, P. Dayan, and J. P. Roiser, "Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees," *PLOS Comput Biol* 8, no. 3 (2012): e1002410. <https://doi.org/10.1371/journal.pcbi.1002410>

<sup>85</sup> M. Hoffman, E. Yoeli, and M. A. Nowak, (2015). "Cooperate without Looking: Why We Care What People Think and Not Just What They Do," *Proceedings of the National Academy of Sciences* 112, no. 6 (2015): 1727–32. <https://doi.org/10.1073/pnas.1417904112>; J. J. Jordan, M. Hoffman, M. A. Nowak, and D. G. Rand, "Uncalculating Cooperation Is Used to Signal Trustworthiness," *Proceedings of the National Academy of Sciences* 113, no. 31 (2016): 8658–63. <https://doi.org/10.1073/pnas.1601280113>

<sup>86</sup> Jonathan Phillips, and Fiery Cushman, "Morality Constrains the Default Representation of What Is Possible," *Proceedings of the National Academy of Sciences* 114, no. 18 (2017): 4649–54.

<sup>87</sup> Jonathan Phillips and P. Bloom, "Do Children Believe Immoral Events Are Magical?" (unpublished manuscript, available at <https://osf.io/en7ut/>).

## IV. CONCLUSION

We identified four viable theories of how norms influence human behavior (Figure 2). Each theory specified a representation involved in decision-making that norms could alter. Norms could change the representation of the transition function  $T$  (“if I cheat on my taxes, then I’ll go to jail”); the average values of past actions  $Q$  (“I’ve learned to associate cheating on things like taxes with badness”); the reward function  $R$  (“I would feel guilty for not paying my fair share”); or the action set  $A_s$  (“I never even considered cheating on my taxes”).

Construing theories of norm compliance within a common framework has several benefits. We briefly highlight three of these. First, by formalizing norm compliance within the MDP setting, we can identify new potential mechanisms — ones that are currently poorly represented in the literature. For instance, one of these is the definition of the state space,  $S$ . Consider, for instance, the norm of impartial beneficence — that is., the claim that welfare benefits should be distributed identically to all people without respect to their personal identity or group membership. This norm might naturally be captured by dictating that the state representations implicated in welfare decisions should also be insensitive to personal identity or group membership. This formalizes the ordinary notion of “blind justice,” in the sense that rules are applied in a manner deliberately isolated from a source of potential information. Despite its intuitive appeal, we know of little current research on the interface between norm learning and state representation in the psychology, neuroscience or economics literatures.

Second, drawing on the MDP formalization facilitates comparison between norm compliance and other forms of decision-making. The MDP formalization and several RL algorithms defined within it have enabled notable advances in psychology and neuroscience. This research has mostly focused on the principles of learning in nonsocial environments. There is considerable interest in determining whether the influence of morality — and, more broadly, social norms — operates via overlapping or even identical mechanisms.<sup>88</sup> Asking this question requires us to map theories of social and nonsocial decision-making into a common conceptual space, as we have attempted here.

Third, the framework draws out certain unexplored points of divergence between the rival theories. For instance, by positing that norms affect different decision variables, the theories implicitly make different

<sup>88</sup> Fiery Cushman, “Action, Outcome, and Value: A Dual-System Framework for Morality,” *Personality and Social Psychology Review* 17, no. 3 (2013): 273–92. <https://doi.org/10.1177/1088868313495594>; C. C. Ruff and E. Fehr, “The Neurobiology of Rewards and Values in Social Decision Making,” *Nature Reviews Neuroscience* 15, no. 8 (2014): 549–62. <https://doi.org/10.1038/nrn3776>

claims about how easy it is to change people's norm-compliance tendencies. The transition function  $T$  is relatively pliable; it flexibly incorporates many sources of information (like verbal instructions<sup>89</sup>). Thus, according to the folk theory, it should be easy to change people's norm-compliance tendencies. In contrast, the cached values  $Q$  are thought to primarily reflect direct experience with rewards and punishments; and the reward function  $R$  is often assumed to be nearly immutable. Thus, if norms affect  $Q$  (habit theory) or  $R$  (internalization theory), their influence is less easily undone. This highlights just one of several ways in which the MDP formalization can draw out new insights about the nature of rival theories of norm compliance.<sup>90</sup>

As should be apparent, by exploring the space of possible cognitive mechanisms underlying norm compliance, we have raised more questions than we answered. Nonetheless, we hope that this review will encourage the kind of cross-pollination necessary to progress our understanding of this vital topic. Norms touch every aspect of our external lives; one day, we may know which parts of our internal lives they touch as well.

*Psychology, Harvard University*

<sup>89</sup> N. D. Daw, S. J. Gershman, B. Seymour, P. Dayan, and R. J. Dolan, "Model-Based Influences on Humans' Choices and Striatal Prediction Errors," *Neuron* 69, no. 6 (2011): 1204–1215. <https://doi.org/10.1016/j.neuron.2011.02.027>

<sup>90</sup> An alternative formulation of this idea is that the decision variables have different levels of informational encapsulation (J. A. Fodor, *The Modularity of Mind: An Essay on Faculty Psychology* [Cambridge, MA: MIT Press, 1983]):  $Q$  and  $R$  are more encapsulated than  $T$ , and there are therefore fewer types of experience that can change them. At the present, it is unknown how encapsulated the action set  $A_s$  or the state space  $S$  are.