Cognition 167 (2017) 91–106

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Social is special: A normative framework for teaching with and learning from evaluative feedback

Mark K. Ho [a,*], James MacGlashan [b], Michael L. Littman [b], Fiery Cushman [c]

[a] Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Box 1821, Providence, RI 02912, United States
[b] Department of Computer Science, Brown University, 115 Waterman St, Providence, RI 02906, United States
[c] Department of Psychology, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 31 March 2016
Revised 13 December 2016
Accepted 9 March 2017
Available online 22 March 2017

Keywords:
Reward
Punishment
Theory of mind
Social learning
Evaluative feedback
Teaching

ABSTRACT

Humans often attempt to influence one another's behavior using rewards and punishments. How does this work? Psychologists have often assumed that "evaluative feedback" influences behavior via standard learning mechanisms that learn from environmental contingencies. On this view, teaching with evaluative feedback involves leveraging learning systems designed to maximize an organism's positive outcomes. Yet, despite its parsimony, programs of research predicated on this assumption, such as ones in developmental psychology, animal behavior, and human-robot interaction, have had limited success. We offer an explanation by analyzing the logic of evaluative feedback and show that specialized learning mechanisms are uniquely favored in the case of evaluative feedback from a social partner. Specifically, evaluative feedback works best when it is treated as communicating information about the value of an action rather than as a form of reward to be maximized. This account suggests that human learning from evaluative feedback depends on inferences about communicative intent, goals and other mental states—much like learning from other sources, such as demonstration, observation and instruction. Because these abilities are especially developed in humans, the present account also explains why evaluative feedback is far more widespread in humans than non-human animals.

1. Introduction

Parents scold; teachers grade; lovers pout; bosses bonus; colleagues grouse; nations sanction; citizens protest; eyes smile and mouths frown. In short, people rarely forgo an opportunity for *evaluative feedback*: reward or punishment of another person in a manner designed to change their future behavior. Although teaching by evaluative feedback is sometimes costly, the potential benefit is obvious: We can exploit the capacity of social partners to learn from reward and punishment to shape their future behavior to profit ourselves, our kin and our allies. In many instances, such as parenting, long-run benefits accrue not only to the teacher (e.g., a parent) but also to the learner (the child) as they learn more adaptive patterns of behavior. The ubiquity of evaluative feedback is unremarkable because it is so effective. Dozens of laboratory (Balliet, Mulder, & Van Lange, 2011; Fehr & Gächter, 2002) and field (Owen, Slep, & Heyman, 2012) studies show that humans can effectively shape the behavior of other humans through the use of selective reward and punishment. Our goal is to understand how.

More precisely, we ask whether there is anything special about learning from social rewards and punishments, as compared to ordinary environmental rewards and punishments. Evaluative feedback from social others take on many forms. For instance, a social other may redirect naturally occurring stimuli in order to inflict pleasure or pain on a learner; giving or withholding food, comfort, poison, and painful experiences all fall under this category. Evaluative feedback may also depend on uniquely human and intrinsically social signals such as verbal praise or reprimands, or a smile or scowl. Although these forms of evaluative feedback differ in many ways, they all involve (1) a social agent causing (2) a rewarding or aversive experience in (3) another social agent, and (4) in a manner ultimately designed to cause learning and behavioral change. What are the cognitive mechanisms that support this form of social teaching and learning in humans? Are they specially adapted to the social domain? Should they be?

At first blush, the answer seems obvious. The tendency of organisms to repeat what is positive and to avoid what is negative is fundamental to psychological theory, akin to gravity in physics

* Corresponding author.
E-mail addresses: mark_ho@brown.edu (M.K. Ho), jmacglashan@gmail.com (J. MacGlashan), mlittman@cs.brown.edu (M.L. Littman), cushman@fas.harvard.edu (F. Cushman).

http://dx.doi.org/10.1016/j.cognition.2017.03.006
0010-0277/

or natural selection in biology. The power of these rewards and punishments to shape human behavior is entirely unsurprising because rewards and punishments exert a gravitational force on the behaviors of non-human animals from the sea-slug (Cook & Carew, 1986) to the chimpanzee (Randolph & Brooks, 1967), and every lab rat (Guttman, 1953), cat (Populin & Yin, 1998; Thorndike, 1898) and pigeon (Skinner, 1948) in between. Moreover, brain imaging studies have confirmed that material rewards and inherently social rewards like facial expressions are processed in similar regions (Lin, Adolphs, & Rangel, 2012). Here, then, is a simple premise that has inspired much prior research: *Social rewards and punishment shape behavior by exploiting the same learning mechanisms that process environmental rewards and punishments*. This claim does not commit to any particular form of the learning (associative, causal, Bayesian, etc.). Rather, the key claim is that however we learn from rewards and punishments of non-social origin, we learn the same way from rewards and punishments originating from social partners. That is, we learn from the sting of criticism just as we would from the prick of a thorn.

Although parsimonious, this premise is closely associated with several unfulfilled programs of research. In the 1950s and 1960s, buoyed by decades of progress in animal learning, researchers began to apply principles of operant conditioning discovered in non-social learning tasks to the socialization of children (Aronfreed, 1968; Bryan & London, 1970; Sears, Maccoby, & Levin, 1957). There were some later successes in showing that behaviors like altruism could be reinforced (Gelfand, Hartmann, Cromer, Smith, & Page, 1975; Grusec & Redler, 1980). But as operant conditioning as a theory of social learning in humans lost adherents, the field eventually moved on to alternative models of social learning—for instance, by observation, instruction, or attribution—rather than learning by reinforcement as such (Grusec, 1997; Maccoby, 1992). There is something unsatisfying about this resolution: Humans obviously *do* reward and punish each other, so why can't our best models explain how this contributes to learning?

Similarly, buoyed by theoretical models that predicted the evolution of cooperation through punishment (Clutton-Brock & Parker, 1995) and reciprocal rewards (Trivers, 1971), biologists sought to document their prevalence among non-human animals. Again, these attempts yielded surprisingly few empirical successes (Hammerstein, 2003; Raihani, Thornton, & Bshary, 2012; Stevens, Cushman, & Hauser, 2005; Stevens & Hauser, 2004), and attention turned to alternative means of explaining non-human prosociality (West, Griffin, & Gardner, 2007). Again, something has been left unresolved: Given that animals are proficient at learning from environmental rewards and punishments, why don't they reward and punish *each other* more often?

In more recent decades, computer scientists have developed mathematical tools to build agents that embody the basic principles of non-human and human reward learning (e.g. Sutton & Barto, 1998). Yet, when they allow actual human participants to train these agents through reward and punishment, the results are spectacularly disappointing. Machines will often unlearn their initial training or even acquire unintended behaviors that human trainers fail to detect (Isbell, Shelton, Kearns, Singh, & Stone, 2001). Here, again, there is something left unfulfilled. Humans are happy to reward and punish agents employing artificial intelligence in order to improve their behavior. But if the agents are designed to *maximize* those rewards (and minimize punishment), they fail to learn what the humans are trying to teach. Where is the bug in the system?

Collectively, this evidence suggests that there is something special about the way that *human* learners respond to *social* rewards and punishments—and something correspondingly special about how human teachers structure those rewards and punishments. By understanding what that "special something" is, we will be in a better position to understand what human evaluative feedback is good for, why non-human animals are relatively less prone to use it, and how to build artificial intelligence that benefits from it.

Our approach to this problem leverages basic concepts borrowed from reinforcement learning, a framework that formalizes the problem of learning and decision-making based on reward and punishment (Dayan & Niv, 2008; Kaelbling, Littman, & Moore, 1996; Sutton & Barto, 1998). We provide a normative analysis of how teaching and learning from social evaluative feedback should be structured, contrast features of this approach to learning from non-social reinforcement, and compare each of these models against extant findings.

2. Adapting to non-social rewards and punishments

Like most animals, humans learn the value of actions as they experience positive and negative outcomes in the environment. For instance, a rat learns the value of pushing a lever when it experiences contingent food rewards (Guttman, 1953). A major goal of contemporary learning theory is to provide a formal account of the cognitive operations that enable this form of learning (Dayan & Niv, 2008). Many diverse answers to this problem have been proposed, but virtually all of them share a few key features. By summarizing these features, we can state with greater precision the potential similarities or dissimilarities between "traditional" reward learning (in non-social settings) and evaluative feedback (i.e. reward and punishment in a social setting).

2.1. The problem of learning value from reward

The central challenge of decision making for organisms is to choose the right behavior in any situation that arises. If the optimal, fitness-enhancing behavior were sufficiently consistent across time and individuals, then it could be specified entirely innately. For instance, koalas, an arboreal marsupial, mainly consume toxic eucalypts that are not difficult to find or competitively consumed by other species. In part due to the natural invariance of their main food source, koalas will only consume eucalyptus leaves that are attached to branches and not ones that have been plucked and placed on a flat surface (Tyndale-Biscoe, 2005). Reflexes, fixed action patterns, or unconditioned responses all fall into this category of innate stimulus-response mappings.

Of course, this approach is generally impractical: Many features of the world are not predictable from birth and stable across generations. Consider, for instance, the challenge of foraging for food. The timescale at which forests burn, herds migrate, ponds dry, and so forth, means that the most effective behaviors for obtaining food undergoes large changes within (and certainly between) generations. Thus, organisms must have an adaptive mechanism for altering their behavior in response to variable circumstances.

One solution to the problem of adapting behavior to partially predictable environments consists of two interacting representations: innate rewards and learned value (Littman & Ackley, 1991). First, an innate system designates the experience of certain actions, stimuli, or states of affairs as intrinsically rewarding or aversive because they are reliable indicators of fitness improvement or decline. Honey, for instance, could be experienced by an organism as intrinsically rewarding because of its high caloric content. Conversely, bee stings could be intrinsically aversive because they lead to swelling and potential infections.

Second, as an organism acts and undergoes different rewarding, aversive, and neutral experiences, a learning process flexibly updates a representation that predicts the contingencies of actions and experiences. For example, if an organism experiences eating