

 COMMENTARY

# Moral status of accidents

Rebecca Saxe<sup>a,1</sup>

No one is naive enough to expect that all moral beliefs are universal. Today, some countries legally beat and imprison homosexuals, and others recognize gay marriage; in some places, killing a bull is a sport, and, in others, it is an abomination; in some places, corporal punishment is the obligation of a responsible parent and, in others, grounds for forced removal. Indeed, the burden of proof seems to be on the other side: Is there anything universal about human moral cognition? In PNAS, Barrett et al. (1) test one candidate for a universal principle of human morality: that an action's moral value depends not only on the action's consequences but on the person's intentions.

A cognitive universal is a way of thinking that does not have to be invented by an individual or a culture, and does not have to be explicitly transmitted to the next generation by formal pedagogy. In general, scientists take two approaches to searching for such universals. One approach is to measure cognition before cultural influences are likely to operate, in young children and infants. The other approach is to measure cognition across a wide range of cultures, with a special focus on people in remote groups who have been least affected by intergroup contact. Ideally, these two approaches converge: Features of universal cognition that are observed in preenculturation infants are also observed in adults across a wide range of cultures.

Cognitive universals can be contrasted with "cognitive technologies," ideas that evolve or are invented, in one or a few specific places and times, and are transmitted through explicit teaching and modeling (2). The prototypical example of a cognitive technology is the integer count system. Count lists emerged independently in some (but not all) human cultures; through contact between cultures, a small number of counting systems have become statistically dominant in contemporary humans. Nevertheless, count systems are clearly technologies, not a cognitive universal: Specific ways of counting exact set sizes do not emerge spontaneously in childhood and have to be transmitted explicitly through both pedagogy and modeling.

## Morality in Early Development

Is the idea that moral evaluations depend on intentions a universal or a technology? Conventional wisdom in developmental psychology is consistent with the cognitive technology view (3). In Europe and North America, 4-y-old children say that a boy who trips on a rock and accidentally knocks down a little girl is more naughty than a second boy who wants to hit the girl but trips and misses. Only by 7 y old do children clearly say the second boy was naughty because he intended harm (4). The developmental shift toward intent-based moral evaluation is explicitly promoted by parents and teachers in these cultures: When this scenario happens in a playground, adults instruct the boy to consider his sister's feelings (she feels sad, scared) and the girl to consider her brother's intentions (he didn't see her; it was an accident). Could the idea of moral evaluations based on intentions, rather than consequences, be a specific cultural invention, a cognitive technology like an integer count list?

On the other hand, recently, some developmental psychologists have challenged this consensus. Some aspects of very young children's spontaneous social behaviors appear sensitive to others' intentions. For example, 4-y-olds spontaneously share stickers with a puppet who accidentally knocked down their jointly built tower, but do not share stickers if the puppet knocked down the tower intentionally (5). Even more dramatic results come from studies of preverbal babies. Ten-month-olds reach for a puppet who was unknowingly unhelpful to a third character, but do not reach for a puppet who was knowingly unhelpful (6). Apparently, some sensitivity to intentions in moral evaluations emerges spontaneously early in development, before enculturation.

Still, it's unclear how to resolve the conflict between the traditional experiments (asking for children's moral evaluations) and the newer spontaneous measures used with infants and toddlers. Also, all of the children in all of these experiments share a cultural context: They come from families that are exclusively "WEIRD,"—Western, educated, industrialized, rich, and democratic (7). If the role of intent in moral evaluation is a cognitive

<sup>a</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139

Author contributions: R.S. wrote the paper.

The author declares no conflict of interest.

See companion article on page 4688.

<sup>1</sup>Email: saxe@mit.edu.

universal, it should be present in a wider, more representative sample of human social environments.

### Morality Across Cultural Contexts

Searching for cultural variability in our increasingly interconnected world requires heroic effort. Barrett et al. (1) assembled a group of 15 researchers from 11 different universities, collaborated to build a precisely shared protocol that would be culturally appropriate in all 10 cultural contexts, translated and back-translated the protocol into nine different languages, and then implemented these protocols at remote sites all over the globe. At each site, dozens of informants provided moral evaluations of hypothetical scenarios covering a range of immoral actions (e.g., theft, physical harm, poisoning the village well) and potentially mitigating factors (e.g., accident, self-defense). The resulting data are rich and fascinating, and raise many questions for future research.

In the two “Western” cultures, English-speaking adults in public areas of a large American city, and Ukrainian-speaking adults in the rural mountain village of Storozhnitsa, moral evaluations are highly sensitive to intentions. For example, theft (Alex taking Bob’s bag) is judged to be very bad if done intentionally (Alex noticed where Bob placed his bag) or with motivation (Alex saw expensive jewelry in Bob’s bag), but not if the “theft” was done accidentally (Alex meant to pick up his own very similar bag) or without motivation (Alex saw some cheap lettuce in Bob’s bag).

By contrast, Himba adults in the Omuhonga Basin of Namibia report that, even if done accidentally or without motivation, taking the bag is almost as bad as motivated intentional theft. Sursurung-speaking adults from coastal villages in New Ireland, Papua New Guinea, report that taking the bag should be punished equally, whether intentional or not. The most extreme difference from the “Western” groups, in this sample, is observed in the adults from Yasawa Island in Fiji. With striking consistency, individuals from this group treat intentions as irrelevant: Theft, punching someone in the face, and poisoning the village well are all wrong, bad for one’s reputation, and punishable, regardless of the knowledge or motivations of the person.

The use of intentions in moral evaluations clearly varies across cultures, but do these results show that moral evaluation based on intentions is a cognitive technology? A deeper look at the same data suggests the opposite.

Consider the adults from Yasawa Island. Their moral evaluations of accidents and intentional harms are the same. The cognitive technology view suggests that the Yasawa have not evolved or acquired the idea of moral evaluation based on intentions. However, there are many good reasons to reject this interpretation. First, the Yasawa informants unambiguously identify the intentional versus accidental actions, when asked directly. Second, many cultural groups, including the Yasawa, have an explicitly transmitted cultural norm that it is inappropriate, an invasion of privacy, and even a punishable offense to speculate about another person’s thoughts and feelings. Yasawa adults’ public evaluations of these scenarios are thus congruent with their explicit norm about appropriate public discourse, but may diverge from their private cognitions (8).

Third, Yasawa adults do evaluate actions based on the person’s knowledge, as is apparent by contrasting another pair of scenarios in this experiment’s amazingly rich protocol. First, consider Clark, who believes that David is attacking his (Clark’s) son; racing over, Clark punches David in the face, knocking him down, only to discover that David and the son were just play-fighting. Compare that story with Ethan, who sees a dangerous

fire start in the market; racing to put the fire out, Ethan punches Fred (who is blocking the way), knocking him down, but, by the time he arrives, someone else has put out the fire. [Barrett et al. (1) call Ethan a case of “necessity,” but, in the scenario, the harm was actually unnecessary; cases like this are often called “foreseen side-effects” (9).] Clark’s and Ethan’s actions seem elegantly parallel, but, in nearly every society, including the Yasawa and accepting only Los Angeles, Clark’s action is seen as much more morally wrong than Ethan’s. Why?

Clark’s punch was a mistake based on a false belief (his son was not in danger). By contrast, Ethan’s punch was based on a true belief [there was a fire (10)] and reflects a deliberate choice to do a harmful action as a means to a greater good (11). Relatedly, Clark

### In PNAS, Barrett et al. test one candidate for a universal principle of human morality: that an action’s moral value depends not only on the action’s consequences but on the person’s intentions.

and Ethan differ in what they should have known: Clark probably should have known that David was just pretending, but it’s less obvious that Ethan should have known that someone else would put out the fire. If this kind of reasoning explains why the Yasawa (and adults from most cultural groups) evaluate Clark’s punch as punishable but Ethan’s punch as praiseworthy, then it seems clear that Yasawa adults, too, engage in subtle and sophisticated consideration of a person’s beliefs and knowledge when making moral evaluations.

### Convergence of Two Approaches

Taken together, developmental and cross-cultural evidence appear to be converging on a striking picture of universal moral cognition. No matter where you go in the world, from fishing villages of Fiji to mountain towns in Western Ukraine, human adults make moral evaluations of one another’s actions, judging some actions to be wrong, punishable, and bad for one’s reputation, and judging other actions to be neutral or praiseworthy. Nowhere in the world are these evaluations based exclusively on the harm caused to the victim. Adults in every culture recognize the importance of an action’s context: Punching and injuring the victim is very wrong if unprovoked but not wrong at all if the punch occurred in self-defense against the victim’s knife attack. The same punch is also not blameworthy if it was committed as a means to achieve a larger good (putting out a dangerous fire), even if the good outcome did not actually occur. In addition, people from all cultures recognize that some actions’ consequences were not desired or intended, including accidents (tripping) and mistakes (false beliefs). Even infants and young children spontaneously distinguish between accidents, mistakes, and intentionally harmful acts. Recognizing and evaluating actions in terms of a person’s desires, goals, knowledge, and control is a stunningly sophisticated human cognitive universal.

At the same time, the moral status of accidents and mistakes varies substantially, both across cultures and within cultures across action contexts [e.g., theft, ingesting taboo substances, sex with forbidden partners (12)]. At one extreme, a justified mistake may be evaluated as entirely blameless; at the opposite extreme, a false belief or uncontrolled movement may itself be blameworthy. One reason for this variation may be that individuals and groups have different views about what a person

could control or ought to know. Barrett et al. (1) acknowledge that they cannot explain the observed pattern of variation in the status of actions and mistakes, across cultures and contexts; devising and testing potential explanations will be a key challenge for future research.

Even if developmental and cross-cultural approaches converge on cognitive universals, however, we must be wary of conflating the two. When children differ from adults, it is usually because they (children) have not developed or learned a concept or technology that we (adults) have mastered. When adults from

different cultures differ from one another, however, we (Western scientists) must resist slipping into an analogous interpretation. Indeed, an intriguing hypothesis is that the most impressive cognitive technology revealed by this experiment is one we (Westerners) conspicuously lack: the ability to resist evaluating actions based on intentions. The best strategy for shaping future social behavior may be to react only to consequences (13). Blaming a boy for knocking his sister down, whether or not he intended to do it, may be the most effective way to teach him to be more careful the next time.

- 
- 1 Barrett HC, et al. (2016) Small-scale societies exhibit fundamental variation in the role of intentions in moral judgment. *Proc Natl Acad Sci USA* 113:4688–4693.
  - 2 Frank MC, Everett DL, Fedorenko E, Gibson E (2008) Number as a cognitive technology: Evidence from Pirahã language and cognition. *Cognition* 108(3): 819–824.
  - 3 Piaget J (1965) *The Moral Judgment of the Child* (Free Press, New York).
  - 4 Cushman F, Sheketoff R, Wharton S, Carey S (2013) The development of intent-based moral judgment. *Cognition* 127(1):6–21.
  - 5 Chernyak N, Sobel DM (2016) “But he didn’t mean to do it”: Preschoolers correct punishments imposed on accidental transgressors. *Cogn Dev* 39:13–20.
  - 6 Kiley Hamlin J, Ullman T, Tenenbaum J, Goodman N, Baker C (2013) The mentalistic basis of core social cognition: Experiments in preverbal infants and a computational model. *Dev Sci* 16(2):209–226.
  - 7 Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? *Behav Brain Sci* 33(2-3):61–83, discussion 83–135.
  - 8 Astuti R (2015) Implicit and explicit theory of mind. *Anthropol This Century* 13:636–650.
  - 9 Mikhail J (2007) Universal moral grammar: Theory, evidence and the future. *Trends Cogn Sci* 11(4):143–152.
  - 10 Young L, Nichols S, Saxe R (2010) Investigating the neural and cognitive basis of moral luck: It’s not what you do but what you know. *Rev Phil Psych* 1(3):333–349.
  - 11 Cushman F (2015) Deconstructing intent to reconstruct morality. *Curr Opin Psychol* 6:97–103.
  - 12 Young L, Saxe R (2011) When ignorance is no excuse: Different roles for intent across moral domains. *Cognition* 120(2):202–214.
  - 13 Cushman F (2013) The role of learning in punishment, prosociality, and human uniqueness. *Cooperation and Its Evolution*, eds Sterelny K, Calcott B, Fraser B (MIT Press, Cambridge, MA), pp 333–372.