Cognition 167 (2017) 191–200

Contents lists available at ScienceDirect

Cognition

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

Original Articles

Moral learning as intuitive theory revision ☆

Marjorie Rhodes [a,*], Henry Wellman [b]

[a] New York University, United States
[b] The University of Michigan, United States

ARTICLE INFO

Article history:
Received 18 January 2016
Revised 29 June 2016
Accepted 24 August 2016
Available online 6 September 2016

Keywords:
Moral learning
Intuitive theories
Conceptual development

ABSTRACT

We argue that moral learning, like much of conceptual development more generally, involves development and change in children's intuitive theories of the world. Children's intuitive theories involve coherent and abstract representations of the world, which point to domain-specific, unobservable causal-explanatory entities. From this perspective, children rely on intuitive sociological theories (in particular, an abstract expectation that group memberships constrain people's obligations), and their intuitive psychological theories (including expectations that mental states motivate individual behavior) to predict, explain, and evaluate morally-relevant action. Thus, moral learning involves development and change in each of these theories of the world across childhood, as well as developmental change in how children integrate information from these two intuitive theories. This perspective is supported by a series of research studies on young children's moral reasoning and learning, and compared to other developmental approaches, including more traditional forms of constructivism and more recent nativist perspectives.

1. Moral learning as informed by children's developing theories of agents and groups

Imagine a childhood moral transgression: A child sneaks into her classroom while she is supposed to be at recess, takes a cookie belonging to another child that she eyed during snack time, puts it in her own backpack, and leaves the room to rejoin her class. Later, when the class is asked if anyone knows where the missing cookie might be, she remains silent.

Most would agree that the child's actions were morally wrong. Broadly, several types of information feed into this judgment. At the least: (a) that the other child was harmed (left sad and hungry, with certain property rights violated), and (b) that the agent's mental states (e.g., her knowledge that the cookie belonged to someone else and her intent to take it for herself) make her culpable for these outcomes.

What sort of learning and development does such a system of moral judgment require, enable, and manifest? We view moral development, like much of conceptual development more generally, as involving the development of children's intuitive theories of the world (Gopnik & Wellman, 2012; Wellman & Gelman, 1992). On this view, conceptual structures take the form of everyday theories (Murphy & Medin, 1985), and cognitive development may be understood as a process of theory revision. Thus, via processes of constructivist learning (Gopnik & Wellman, 2012; Xu, 2007) children acquire intuitive theories of the world, revise those theories in response to new evidence, and employ those theories to learn further information. Children's intuitive theories involve coherent and abstract representations of the world, which point to domain-specific, unobservable causal-explanatory entities (e.g., gravity in the case of intuitive physics, desires in the case of intuitive psychology). Children's theories are also hierarchical—specific theories of how things work (e.g., that cookies are more desirable than carrots and the child above desires cookies) are embedded in more abstracted "framework theories" of the relevant domain (e.g., that unobservable mental states such as desires generally motivate behavior; Carey, 2009; Wellman, 1990; Wellman & Gelman, 1992). Often, children first construct a framework theory of a domain—a broad view that human behavior relies on unobservable mental states in the case of intuitive psychology. These framework theories underlie more specific theories within the domain—e.g., that desirable cookies cause specific sorts of behavior.

Development and learning can involve change in both of these levels of children's knowledge. For example, change in children's framework theories could (and does, see e.g., Wellman, 2014) include a change from the theory that desires motivate behavior to a more complex theory that the influence of desire on behavior

☆ Preparation of this article was supported by National Science Foundation grant BCS -1226942 to Rhodes and a grant from the Varieties of Understanding Project at Fordham University and the John Templeton Foundation to Wellman.
* Corresponding author at: 6 Washington Place, 301, New York, NY 10003, United States.
E-mail address: marjorie.rhodes@nyu.edu (M. Rhodes).

http://dx.doi.org/10.1016/j.cognition.2016.08.013
0010-0277/

is moderated by beliefs and knowledge. Change in a child's more specific theory might include learning that both cookies and carrots could be desirable for different reasons (for taste or health), and that an agent's choice might reflect multiple concerns.

Children's theories serve specific cognitive functions—they enable children to predict, explain, and evaluate events in their environment (Gopnik & Wellman, 1994, 2012). For example, in the case of intuitive psychology, children can use a desire-based theory to predict that an agent will reach for the snack they desire, to then explain the agent's action (e.g., he took that one *because* he wanted it) and to evaluate whether an observed outcome was consistent or inconsistent with one's expectation. This type of evaluation allows theories to be dynamic as well—intuitive theories can change in response to observed evidence. This change often happens in a gradual and progressive manner, instead of all-at-once. For example, in the case of intuitive psychology, children move from a fairly rudimentary desire-based theory to a full-fledged representational theory of mind by passing through levels (Wellman & Liu, 2004) where they come to understand desires as moderated by simple aspects of perception, then to understand knowledge and ignorance, and then finally to understand fully the representational nature of belief (including that beliefs can be inconsistent with reality). Viewing cognitive development as a process of intuitive-theory-change encourages researchers to examine several types of development and learning. In particular, it is important to examine both how intuitive theories develop and change over extended periods of time, as well as how intuitive theories direct attention and memory to shape learning in children's day-to-day interactions. Moreover, an intuitive theories perspective encourages researchers to examine important progressions in children's conceptual development where earlier understandings within a progression set the stage for and constrain acquisition of later conceptual understandings.

Viewing cognitive development in this way contrasts with other theoretical proposals. In what follows we will distinguish our position in particular from nativist accounts, which emphasize early (evolved) understandings rather than the processes that underlie change over human development. We also distinguish the type of learning that we describe from other constructivist or social learning accounts, which describe change as motivated solely by children's responses to their own actions or from direct instruction. Finally, we distinguish our account from Social Domain Theory, which describes the cognitive domains relevant to moral judgment much differently than we propose here.

An intuitive theories perspective has been fruitfully applied to multiple conceptual domains, including intuitive physics (Kushnir & Gopnik, 2007), biology (e.g., Carey, 1985; Gelman, 2003; Inagaki & Hatano, 2002), psychology (Gopnik & Meltzoff, 1997; Wellman, 2014), and sociology (Hirschfeld, 1996; Rhodes, 2012). We view moral judgment, and thus moral learning, as resting on the interplay of intuitive psychology and sociology. Intuitive psychology shapes children's beliefs about how individuals' mental states (e.g., beliefs, desires, knowledge, traits, and so on) predict and explain behavior, whereas intuitive sociology shapes their understanding of how people relate to one another. Because moral judgments integrate information represented by both of these intuitive theories, moral learning can entail change in the relevant components of the theories that compose children's intuitive psychology or sociology, as well as in how children integrate information from these two theories. We predominantly focus on children's explicit conceptual understandings in these domains, as we will clarify in what follows.

1.1. Intuitive psychology

Return to the scenario at the beginning of this paper, but with an entirely different set of mental states. In this new account, the girl did not know the cookie belonged to another child, but simply sees it sitting on a table and thinks that it is for anyone in the class. Perhaps also, the girl doesn't realize when the teacher asks about the other child's cookie that it is the same cookie she has taken. Given these different beliefs, the agent might not be judged as morally culpable (or at least not nearly to the same extent)—her actions still caused harm, but she didn't hold the mental states necessary to make her responsible for their outcome. Or consider this scenario—the girl sees her friend with a cookie at snack time which she wants for herself, returns to the classroom to take it, but in her absence does not know that her friend in fact ate her own cookie, and what is now on the desk is a cookie that is available for anyone in the class. In this case, she had *malicious intent*, even though her mistaken beliefs and knowledge were such that her actions did not actually cause harm or infringe on anyone's property rights. Most would agree that her actions in this case were morally suspect, despite the lack of a harmful outcome.

These examples illustrate that an agent's mental states matter a great deal in everyday moral judgments. Indeed, the role of mental states in determining moral culpabilities is reflected in the legal system (e.g., in the difference between murder and manslaughter, Hart, 1968; Mikhail, 2007), and is readily recognized by adult participants in psychological studies. Adults view agents who cause harm intentionally or who intend to cause harm but fail to (because of mistaken knowledge or beliefs) as more morally culpable than those who cause harm accidentally (while trying to do good; Cushman, 2008; Knobe, 2005; Singer, Kiebel, Winston, Dolan, & Frith, 2004; Young, Cushman, Hauser, & Saxe, 2007). Further, disruption to brain regions that support reasoning about others' psychological states disrupts this pattern of judgment, leading people in this case to hold others more responsible for actions that they do not bring about on purpose (Young et al., 2007).

Clearly then, one key candidate for important developmental change and learning in moral judgments concerns children's intuitive psychologies—if there is important developmental change in children's abilities to represent and track things like knowledge, intent, and beliefs, then this will correspond to developmental change in moral judgment. Indeed, substantial developmental change occurs in the extent to which children incorporate concepts like *beliefs* and *knowledge* into their intuitive psychological theories across childhood (Wellman, Cross, & Watson, 2001; Wellman & Liu, 2004).

Detailed empirical findings provide evidence for the hypothesized process by which this type of theory-change occurs. For intuitive psychology, Rhodes and Wellman (2013) combined developmental scaling and experimental, microgenetic methods to examine the processes underlying the acquisition of a representational theory of mind within a progression of conceptual development. In employing scaling methods, they first assessed children's initial psychological theories for the extent to which children understood that (a) people have unique desires, (b) people have unique beliefs, (c) people only know what they have access to, and (d) people can believe things that are false (inconsistent with reality). Such conceptions exhibit a developmental progression in intuitive psychological understanding, proceeding from (a) to (d) as revealed in children's task performance on a well-validated developmental scale (e.g., Wellman, Fang, & Peterson, 2011; Wellman & Liu, 2004). Note that in this scale, and within the wider literature on children's intuitive psychology, an understanding of false beliefs constitutes a milestone achievement. In particular, the ability to pass an explicit false belief tasks (exemplified on the left side of Fig. 1) has often been taken to indicate the development of a full-fledged representational theory of mind (see the meta-analysis by Wellman et al., 2001).

In Rhodes and Wellman (2013), only children who initially failed explicit measures of false belief understanding, and thus