

THE COGNITIVE NEUROSCIENCES

Sixth Edition

David Poeppel, George R. Mangun,
and Michael S. Gazzaniga, *Editors-in-Chief*

Section Editors:

Danielle Bassett	John W. Krakauer
Marina Bedny	Nikolas Kriegeskorte
Sarah-Jayne Blakemore	Steven J. Luck
Alfonso Caramazza	Ulman Lindenberger
Maria Chait	Josh McDermott
Anjan Chatterjee	Elizabeth Phelps
Stanislas Dehaene	Liina Pykkänen
Mauricio Delgado	Charan Ranganath
Karen Emmorey	Adina Roskies
Kalanit Grill-Spector	Tomás J. Ryan
Richard B. Ivry	Wolfram Schultz
Sabine Kastner	Daphna Shohamy

THE MIT PRESS
CAMBRIDGE, MASSACHUSETTS
LONDON, ENGLAND

© 2020 The Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in ITC New Baskerville Std by Westchester Publishing Services. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Poeppel, David, editor. | Mangun, G. R. (George Ronald), 1956- editor. | Gazzaniga, Michael S., editor.

Title: The cognitive neurosciences / edited by David Poeppel, George R. Mangun, and Michael S. Gazzaniga.

Description: Sixth edition. | Cambridge, MA : The MIT Press, [2020] | Includes bibliographical references and index.

Identifiers: LCCN 2019007816 | ISBN 9780262043250
(hardcover : alk. paper)

Subjects: LCSH: Cognitive neuroscience.

Classification: LCC QP360.5 .C63952 2020 |
DDC 612.8/233--dc23 LC record available at
<https://lcn.loc.gov/2019007816>

10 9 8 7 6 5 4 3 2 1

1 Early Moral Cognition: A Principle-Based Approach

MELODY BUYUKOZER DAWKINS, FRANSISCA TING, MAAYAN STAVANS,
AND RENÉE BAILLARGEON

ABSTRACT There is considerable evidence that beginning early in life, abstract principles guide infants' reasoning about the displacements and interactions of objects (physical reasoning) and about the intentional actions of agents (psychological reasoning). Recently, developmental researchers have begun to explore whether early emerging principles also guide infants' reasoning about individuals' actions toward others (sociomoral reasoning). Investigations over the past few years suggest that at least four principles may guide early sociomoral reasoning: fairness, harm avoidance, in-group support, and authority. In this chapter, we review some of the evidence for these principles. In particular, we report findings that infants expect individuals to distribute windfall resources and rewards fairly; they expect individuals in a social group to help in-group members in need, to limit unprovoked and retaliatory harm toward in-group members, to prefer and align with in-group members, and to favor in-group members when distributing limited resources; and they expect an authority figure in a group to rectify transgressions among subordinate members of the group. Together, these findings support prior claims by a broad cross-section of social scientists that a small set of universal principles shapes the basic foundation of human moral cognition, a foundation that is then extensively revised by experience and culture.

Beginning in the first year of life, infants attempt to make sense of the world around them. How do they do so? A major hypothesis in developmental research has long been that in each core domain of causal reasoning, a skeletal framework of abstract principles and concepts guides how infants reason about events (Gelman, 1990; Leslie, 1995; Spelke, 1994). Initial investigations focused on infants' *physical reasoning* and found that principles of gravity, inertia, and persistence (with its corollaries of solidity, continuity, cohesion, boundedness, and unchangeableness) constrain early reasoning about objects' displacements and interactions (Baillargeon, 2008; Luo, Kaufman, & Baillargeon, 2009; Spelke, Phillips, & Woodward, 1995). Thus, even young infants realize that an inert object cannot remain suspended when released in midair (gravity); cannot spontaneously reverse course (inertia); cannot occupy the same space as another object (solidity); and cannot

spontaneously disappear (continuity), break apart (cohesion), fuse with another object (boundedness), or change into a different object (unchangeableness).

Next, researchers turned to infants' *psychological reasoning* (also referred to as mental-state reasoning or theory of mind). Investigations revealed that when infants observe an agent act in a scene, they attempt to infer the agent's mental states; these can include motivational states (e.g., intentions), epistemic states (e.g., ignorance), and counterfactual states (e.g., false beliefs) (Gergely, Nádasdy, Csibra, & Bíró, 1995; Luo & Baillargeon, 2007; Onishi & Baillargeon, 2005). Infants then use these mental states, together with a principle of rationality (and its corollaries of consistency and efficiency), to predict and interpret the agent's subsequent actions (Baillargeon, Scott, & Bian, 2016; Gergely et al., 1995; Woodward, 1998). Thus, if an agent wants a toy and sees someone place it in one of two containers, infants expect the agent to reach for the correct container (consistency) and to retrieve the toy without expending unnecessary effort (efficiency).

More recently, researchers have begun to study infants' *sociomoral reasoning*. Initially, it appeared as though the skeletal framework in this domain, unlike those in the previous two domains, might involve no principles. In particular, infants seemed to hold no expectations about whether individuals would refrain from harming others or would help others in need of assistance. In a series of experiments, infants ages 3–19 months were presented with various scenarios depicting interactions among nonhuman individuals (e.g., different blocks with eyes; Hamlin, 2013, 2014; Hamlin & Wynn, 2011; Hamlin, Wynn, & Bloom, 2007, 2010; Hamlin, Wynn, Bloom, & Mahajan, 2011). Each scenario involved two events: a *positive* event, in which a nice character acted positively toward a protagonist (e.g., rolled a dropped ball back to the protagonist or helped the protagonist reach the top of a steep hill), and a *negative* event, in which a mean character acted negatively toward the same protagonist (e.g., stole the ball or knocked the

protagonist down to the bottom of the hill). Across ages and scenarios, infants looked equally at the two events, suggesting that they detected no violations in the negative events and hence did not expect the mean character to either refrain from harming the protagonist or help it achieve its goal. These results did not stem from infants' inability to understand the scenarios presented: when encouraged to choose one of the two characters, infants 3–10 months old consistently preferred the nice one over the mean one (Hamlin et al., 2007, 2010; Hamlin & Wynn, 2011). Together, these results suggested that infants possess abstract concepts of welfare and harm, distinguish between positive and negative actions, and hold affiliative attitudes consistent with these valences. Nevertheless, infants seemed to lack principle-based expectations about individuals' actions toward others, suggesting that the skeletal framework for sociomoral reasoning included moral concepts but not moral principles (e.g., infants held no expectations as to whether the characters would harm or help the protagonist, but they did recognize harm or help when they saw it).

This characterization of early morality began to change, however, as researchers went on to explore other scenarios. It is now becoming clear that the skeletal framework that guides early sociomoral reasoning does include a small set of principles. However, because most of these principles apply only when specific preconditions are met, expectations related to the principles can be observed only with scenarios that satisfy these preconditions. For example, if infants view helping as expected only among in-group members, then they will expect an individual to aid another only when the two are clearly identified as members of the same social group.

Over the past few years, evidence has slowly been accumulating for at least four sociomoral principles (Baillargeon et al., 2015). The most general is *fairness*, which applies broadly to all individuals: all other things being equal, individuals are expected to treat others fairly, according to their just deserts. At the next level of generality is *harm avoidance*: when individuals belong to the same moral circle (e.g., humans), they are expected not to cause significant harm to each other. At the next level of generality is *in-group support*: when individuals in a moral circle belong to the same social group (e.g., teammates), additional expectations of in-group care and in-group loyalty are brought to bear. Finally, at the fourth and most specific level is *authority*: when individuals in a social group are identified as authority figures or subordinates, further expectations related to these group roles come into play (e.g., rectifying transgressions for the authority figures or obeying directives for the subordinates). Thus, each new structure in the

social landscape—moral circle, social group, group roles—brings forth new expectations about how individuals will act toward others.

This emerging characterization of early morality supports long-standing claims, by a broad cross-section of social scientists, that the basic structure of human moral cognition includes a small set of universal foundations or principles (Baumard, André, & Sperber, 2013; Brewer, 1999; Cosmides & Tooby, 2013; Dawes et al., 2007; Dupoux & Jacob, 2007; Graham et al., 2013; Jackendoff, 2007; Pinker, 2002; Rai & Fiske, 2011; Shweder, Much, Mahapatra, & Park, 1997; Tyler & Lind, 1992; Van Vugt, 2006). Although details about the nature and contents of these principles vary across accounts, it is commonly assumed that the principles evolved during the millions of years our ancestors lived in small groups of hunter-gatherers, where survival depended on cooperation within groups and, to a lesser extent, between groups; that the principles interact in various ways and must be rank ordered when they suggest distinct courses of action; and that different cultures implement, stress, and rank order the principles differently, resulting in the diverse moral landscape that exists in the world today. Graham et al. (2013) aptly described this view as “a theory about the universal first draft of the moral mind and about how that draft gets revised in variable ways across cultures” (p. 65).

In the remainder of this chapter, we review some of the recent evidence that principles of fairness, harm avoidance, in-group support, and authority are included in the “first draft” of moral cognition.

Fairness

According to the principle of fairness, all other things being equal, individuals are expected to treat others fairly when allocating windfall resources, dispensing rewards, or meting out punishments (Baillargeon et al., 2015; Dawes et al., 2007; Graham et al., 2013; Rai & Fiske, 2011). Traditionally, investigations of fairness in preschoolers have used *first-party* tasks, in which the children tested are potential recipients, and *third-party* tasks, in which they are not. Perhaps not surprisingly, given young children's pervasive difficulty in curbing their self-interest, a concern for fairness has typically been observed only in third-party tasks (Baumard, Mascaro, & Chevallier, 2012; Olson & Spelke, 2008). Building on these results, investigations with infants have also used third-party tasks to examine early expectations about fairness.

Equality Do infants expect a distributor to divide windfall resources equally between similar recipients?

In a series of experiments (Buyukozer Dawkins, Sloane, & Baillargeon, 2019; Sloane, Baillargeon, & Premack, 2012), 4-, 9-, and 19-month-olds were tested using the violation-of-expectation method (this method takes advantage of infants' natural tendency to look longer at events that violate, as opposed to confirm, their expectations). Infants faced a puppet-stage apparatus and saw live events in which an experimenter brought in two identical items (e.g., two cookies) and divided them between two identical animated puppets (e.g., two penguins). In one event, the experimenter gave one item to each puppet (*equal* event); in the other, she gave both items to the same puppet (*unequal* event; figure 1.1A). At all ages, infants looked significantly longer if shown the unequal as opposed to the equal event, and this effect was eliminated if the puppets were inanimate (i.e., neither moved nor spoke).

Consistent with the claim that fairness applies broadly, positive results were also obtained when a monkey puppet divided items between two giraffe puppets (Bian, Sloane, & Baillargeon, 2018) and when an orange circle with eyes divided items between two yellow triangles with eyes (Meristo, Strid, & Surian, 2016). At the same time, however, other findings revealed that when the number

of items allocated was increased to four, infants under 12 months of age failed to detect a violation when one recipient was given three items and the other recipient was given one item (Schmidt & Sommerville, 2011; Ziv & Sommerville, 2017). Thus, while a concern for fairness emerges early in life, there are initially sharp limits to the fairness violations young infants can detect, for reasons that are currently being explored.

Equity The preceding findings demonstrate that even young infants possess an expectation of fairness. But how should this expectation be construed? Do infants possess a simple concept of *equality* and expect all individuals to be treated similarly, or do they possess a richer notion of *equity* and expect individuals to receive their just deserts? One way to examine this issue is to present infants with scenarios in which treating individuals the same way would violate fairness. For example, would infants expect a worker, but not a slacker, to receive a reward? To find out, 21-month-olds were shown events in which an experimenter asked two assistants to put away a pile of toys and then left; next to each assistant was a clear lidded box (Sloane et al., 2012). In the *both-help* event, each assistant placed about

Fairness Violations

A. Failing to distribute resources equally



B. Failing to dispense rewards equitably

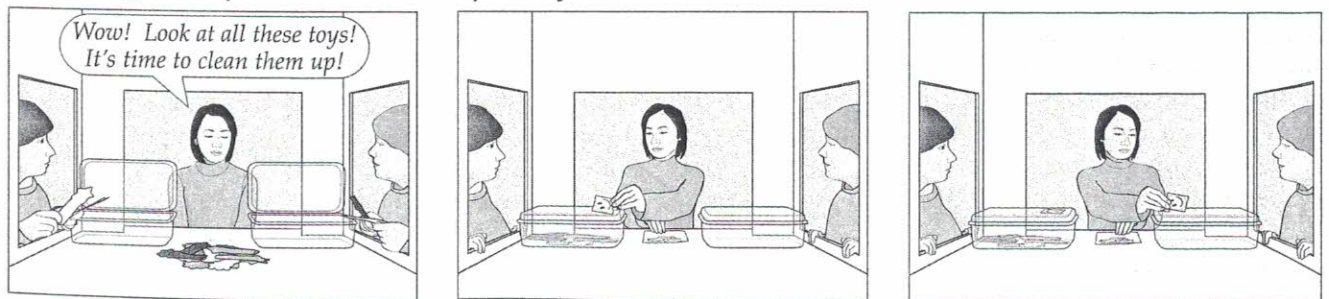


FIGURE 1.1 Infants detect a fairness violation when (A) an experimenter fails to divide windfall resources equally between two similar recipients or (B) fails to dispense rewards equitably

between a worker, who put away toys as instructed, and a slacker, who did no work.

half of the toys in her box and then closed it. The experimenter then returned, inspected both boxes, and rewarded each assistant with a sticker. The *one-helps* event was similar except that one assistant put away all the toys in her box while the other assistant continued to play. Nevertheless, as before, the experimenter gave each assistant a reward (figure 1.1B). Infants looked significantly longer if shown the *one-helps* as opposed to the *both-help* event. This effect was eliminated if the boxes were opaque so the experimenter could no longer determine who had worked in her absence.

Additional experiments indicated that 10-month-olds detected a violation when an experimenter praised two assistants equally even though she could see that only one had performed the assigned task (Buyukozer Dawkins, Sloane, & Baillargeon, 2017); 21-month-olds detected a violation when an experimenter punished two assistants equally even though she could see that only one had not performed the assigned task (Buyukozer Dawkins et al., 2017); and 17-month-olds detected a violation when two workers shared a resource in a manner inconsistent with their respective efforts in obtaining this resource (Wang & Henderson, 2018).

Together, the preceding results suggest that infants' concern for fairness is equity-based: infants expect individuals to get their just deserts, be it an equal share of a windfall resource, a reward commensurate with their efforts, or a punishment that befits their actions.

In-Group Support

According to the principle of in-group support, members of a social group are expected to act in ways that sustain the group (Baillargeon et al., 2015; Brewer, 1999; Graham et al., 2013; Rai & Fiske, 2011; Shweder et al., 1997). The principle has two corollaries, *in-group care* and *in-group loyalty*, each of which carries a rich set of expectations. With respect to in-group care, for example, one is expected (a) to provide help and comfort to in-group members in need and (b) to limit harm to in-group members by refraining from unprovoked harm and by curbing retaliation. Similarly, with respect to in-group loyalty, one is expected (c) to prefer in-group members over out-group members and (d) to reserve limited resources for the in-group. Below, we report evidence that infants already hold these expectations.

Helping the in-group Do infants view helping as expected with an in-group individual but as optional otherwise? In one experiment, 17-month-olds watched events involving three female experimenters, E1–E3, who sat around three sides of an apparatus and announced their group memberships via novel labels (Jin & Baillargeon, 2017).

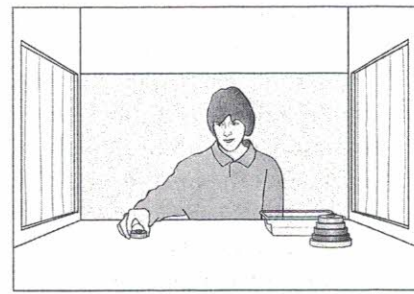
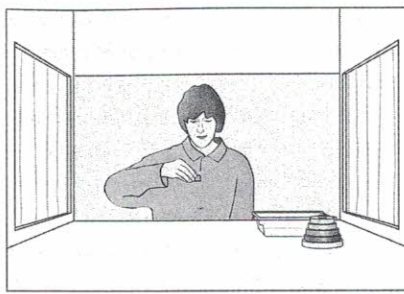
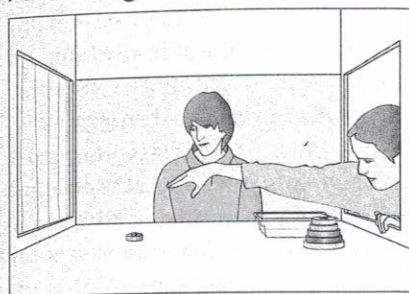
In the *in-group* condition, E1 (on the right) and E2 (in back) belonged to the same group (e.g., “I’m a bem!”; “I’m a bem too!”), while E3 (on the left) belonged to a different group (“I’m a tig!”). In the *out-group* condition, E2 belonged to the same group as E3 instead of E1 (E1: “I’m a bem!”; E2: “I’m a tig!”; E3: “I’m a tig too!”). Finally, in the *no-group* condition, the Es used phrases that provided incidental information about objects they had seen, rather than inherent information about their social groups (E1: “I saw a bem!”; E2: “I saw a bem too!”; E3: “I saw a tig!”). In the test trial, E3 was absent (her main role was to help establish group affiliations), and while E2 watched, E1 selected discs of decreasing sizes from a clear box and stacked them on a base. The final, smallest disc rested across the apparatus from E1, out of her reach (but within E2’s reach). E1 tried in vain to reach the disc until a bell rang; at that point, E1 said, “Oh, I have to go. I’ll be back!” and left. E2 then picked up the smallest disc, inspected it, and either placed it in E1’s box so that she could complete her stack when she returned (*help* event) or returned it to its same position on the apparatus floor, out of E1’s reach (*ignore* event; figure 1.2A).

Infants in the in-group condition looked significantly longer if shown the ignore as opposed to the help event, whereas infants in the out-group and no-group conditions looked equally at the events. Thus, in accordance with the principle of in-group care, infants detected a violation when E2 chose not to help in-group E1. In additional experiments (Jin, Houston, Baillargeon, Groh, & Roisman, 2018), 4-, 8-, and 12-month-olds were shown videotaped events in which a woman was performing a household chore when a baby (who presumably belonged to the same group as the woman) began to cry. The woman either attempted to comfort the baby (*comfort* event) or ignored the baby and continued her work (*ignore* event). At all ages, infants detected a violation in the ignore event, and this effect was eliminated if the baby laughed instead.

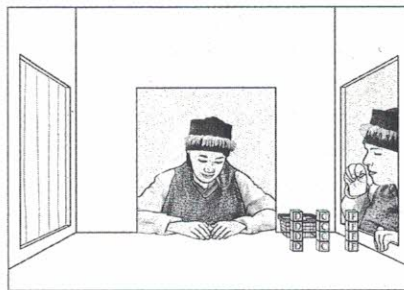
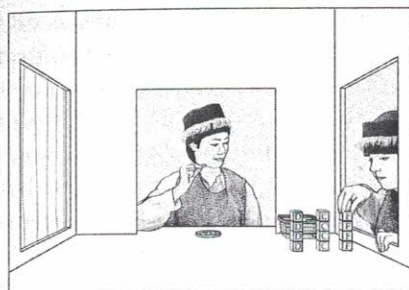
Limiting harm toward the in-group If infants’ sense of in-group care modulates their expectations about harm avoidance, they might expect individuals to direct *less* unprovoked and retaliatory harm at in-group members than at out-group members. To examine these predictions, 18-month-olds were first tested in a baseline *out-group* experiment (Ting, He, & Baillargeon, 2019a). Three female experimenters, E1–E3, sat around three sides of an apparatus, and their group memberships were marked by salient outfits: E1 (on the right) wore one outfit, while E2 (in back) and E3 (on the left) wore a different outfit. While E2 and E3 watched, E1 used small blocks to build two towers of four blocks. In the next trial, E3 was absent and E2 ate crackers from a

In-group—Support Violations

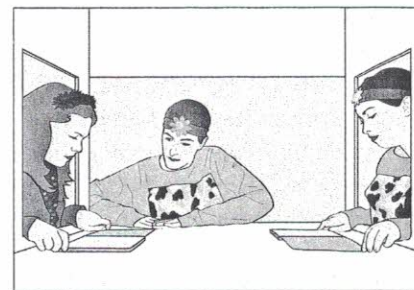
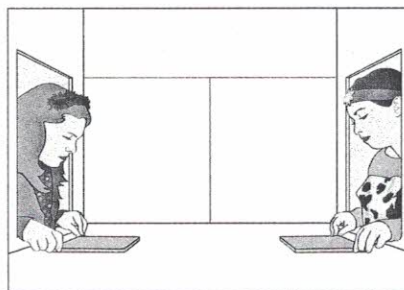
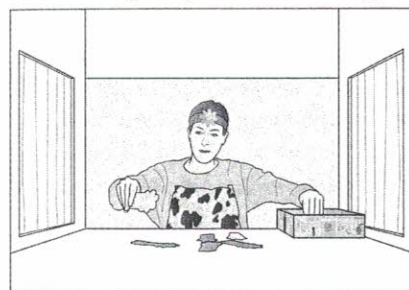
A. Failing to help the in-group



B. Failing to curb retaliation against the in-group



C. Failing to prefer the in-group



D. Failing to favor the in-group when dividing limited resources

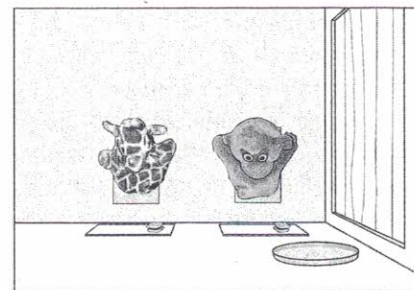
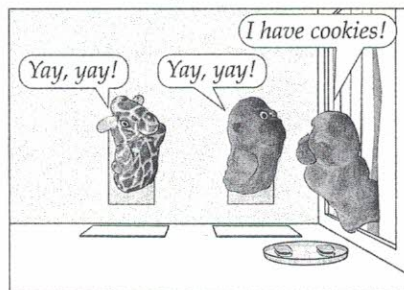
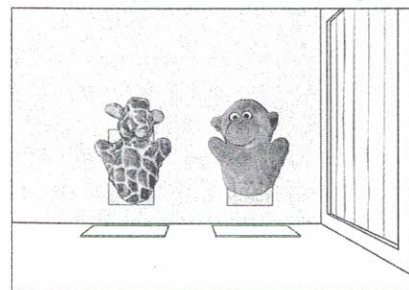


FIGURE 1.2 Infants detect an in-group—support violation when (A) an individual fails to help an in-group member in need of assistance, (B) fails to curb retaliation against an

in-group member who stole and ate a cracker, (C) fails to prefer an in-group member over an out-group member, and (D) fails to favor the in-group when distributing limited resources.

small plate in front of her while watching E1 build a third tower. After completing this tower, E1 either simply left the scene (*no-provocation* condition) or first stole a cracker from E2 and then left the scene (*provocation* condition). In both conditions, E2 then knocked down one block from one tower (*one-block* event), one tower (*one-tower* event), or two towers (*two-tower* event). In the no-provocation condition, infants looked significantly longer if shown the one- or two-tower event as opposed to the one-block event; in the provocation event, in contrast, infants looked significantly longer if shown the one-block or one-tower event as opposed to the two-tower event. Thus, when no provocation had occurred, infants detected a violation in all but the one-block event: mild unprovoked harm to out-group E1 was acceptable, but more significant harm was not. Following provocation, however, infants detected a violation in all but the two-tower event, suggesting that they viewed knocking down at least two of out-group E1's towers as an appropriate retaliatory response for her theft of one cracker (perhaps in a sort of "two-for-one" accounting).

Would infants show similar expectations if E1 and E2 belonged to the same group, or would considerations of in-group care modulate these expectations, leading infants to expect both less unprovoked harm and less retaliatory harm? To find out, infants were tested in an *in-group* experiment identical to that above except that E2 wore the same outfit as E1 and hence belonged to the same group. Across conditions, infants now detected a violation in all but the one-block event of the provocation condition. Thus, when no provocation had occurred, infants expected E2 to refrain from knocking down *any* of in-group E1's blocks; following provocation, knocking down one block became permissible in retaliation for in-group E1's theft—but no more than one block and certainly not two towers, as in the out-group experiment (figure 1.2B).

Together, the preceding results make clear that from an early age, considerations of in-group care modulate expectations about harm avoidance: infants expect stricter limits on unprovoked and retaliatory harm when directed at in-group members. In line with these results, recent experiments have found that infants also expect individuals to *punish* harm to in-group members, at least indirectly, through the withholding of help (Ting, He, & Baillargeon, 2019b). When a bystander saw a wrongdoer harm a victim, and the wrongdoer subsequently needed help to complete a task, 13-month-olds expected the bystander to refrain from providing help if the victim was an in-group member, but not if she was an out-group member. Infants' concern for in-group

care thus leads them to expect individuals both to limit harm to in-group members and to punish such harm, at least indirectly, when perpetrated by others.

Preferring the in-group Do infants expect individuals in a group to prefer in-group members over out-group members, in accordance with the principle of in-group loyalty? In one experiment (Bian & Baillargeon, 2016), 12-month-olds again saw events involving three female experimenters, E1–E3, whose group memberships were marked by salient outfits. In one familiarization trial, E2 (in back) sat alone; she picked up two-dimensional toys on the apparatus floor and placed them in a box near her, thus giving infants the opportunity to observe her outfit. In the next familiarization trial, E2 was absent and E1 (on the right) and E3 (on the left) read identical books; one E wore the same outfit as E2, and the other E wore a different outfit. In the test trial, E1 and E3 were joined by E2, who approached either the E from the same group (*approach-same* event) or the E from the other group (*approach-different* event; figure 1.2C) to read along. Infants looked significantly longer at the approach-different than at the approach-same event, suggesting that they expected E1 to approach her in-group member, in accordance with in-group loyalty, and detected a violation when she did not. This effect was eliminated when the first familiarization trial was modified to reveal that E2's outfit served an instrumental role: she now placed the toys she picked up in the large kangaroo pocket on her shirt, instead of in the box near her. Infants looked equally at the approach-different and approach-same events, suggesting that they no longer viewed the Es' outfits as providing information about their group memberships (in the same way, adults would not view pedestrians holding black umbrellas in the rain on a busy street, or travelers pulling black suitcases in a busy airport, as belonging to the same groups).

Similar results have been obtained in tasks using other cues to group memberships. After watching nonhuman adult characters soothe baby characters, 16-month-olds detected a violation if one baby preferred a baby who had been soothed by a different adult (and hence presumably belonged to a different group) over a baby who had been soothed by the same adult (and hence presumably belonged to the same group) (Spokes & Spelke, 2017). After watching two groups of nonhuman characters (identified by both physical and behavioral cues) perform distinct novel conventional actions, infants 7–12 months old detected a violation if a member of one group chose to imitate the other group's conventional action (Powell & Spelke, 2013). Finally, when faced with a native speaker of their language and a foreign speaker,

infants 10–14 months old were more likely to prefer the native speaker (Kinzler, Dupoux, & Spelke, 2007), to select snacks endorsed by the native speaker (Shutts, Kinzler, McKee, & Spelke, 2009), and to imitate novel conventional actions modeled by the native speaker (Buttelmann, Zmyj, Daum, & Carpenter, 2013). One interpretation of these last results is that in this minimal setting contrasting two unfamiliar individuals, language served as a natural group marker, leading infants to prefer and align with the native speaker, in accordance with in-group loyalty.

Favoring the in-group when resources are limited If infants' sense of in-group loyalty modulates their expectations about fairness, they might expect a distributor to favor in-group over out-group recipients, particularly when resources are scarce or otherwise valuable. To examine this prediction, 19-month-olds saw resource-allocation events involving two groups of animated puppets, monkeys and giraffes (Bian et al., 2018). A puppet distributor (e.g., a monkey) brought in either three (*three-item* condition) or two (*two-item* condition) items and faced two potential recipients, an in-group puppet (another monkey) and an out-group puppet (a giraffe). In each condition, the distributor allocated two items: she gave one item each to the in-group and out-group puppets (*equal* event; figure 1.2D), she gave both items to the in-group puppet (*favors-in-group* event), or she gave both items to the out-group puppet (*favors-out-group* event). In the three-item condition, the third item was not distributed and was simply taken away by the distributor when she left. Infants in the three-item condition looked significantly longer if shown the favors-in-group or favors-out-group event than if shown the equal event, suggesting that when there were as many items as puppets, infants expected fairness to prevail: they detected a violation if the distributor chose to give two items to one recipient and none to the other, regardless of which recipient was advantaged. In contrast, infants in the two-item condition looked significantly longer if shown the equal or favors-out-group event rather than the favors-in-group event, suggesting that when the distributor had only enough items for the group to which she belonged (e.g., two items and two monkeys), infants expected in-group loyalty to prevail: they detected a violation if the distributor gave any of the items to the out-group puppet.

Together, these results suggest two conclusions. First, the "first draft" of moral cognition includes not only principles of fairness and in-group support but also a context-sensitive ordering of these principles that befits their contents: one is expected to adhere to fairness except in contexts where doing so would be detrimental

to one's group. Second, a shortage of resources is one such context: when there is not enough to go around, the group must come first.

Authority

According to the principle of authority, when a social group accepts an individual in the group as a legitimate leader, rich expectations come into play that reflect this power asymmetry (Baillargeon et al., 2015; Graham et al., 2013; Rai & Fiske, 2011; Tyler & Lind, 1992; Van Vugt, 2006). On the one hand, the leader is expected to maintain order, provide protection, and facilitate cooperation toward group goals. On the other hand, the subordinates are expected to obey, respect, and defer to the leader. Do infants already possess authority-based expectations about the behaviors of leaders toward their subordinates or about the behaviors of subordinates toward their leaders?

Before addressing this question, developmental researchers first had to determine whether infants could represent power asymmetries. Over the past decade, evidence has steadily accumulated that by the second year of life, infants (a) can detect differences in social power (Pun, Birch, & Baron, 2016; Thomsen, Frankenhuis, Ingold-Smith, & Carey, 2011), (b) expect such differences to both endure over time and extend across situations (Enright, Gweon, & Sommerville, 2017; Mascaro & Csibra, 2012), and (c) distinguish between powerful individuals with respect-based as opposed to fear-based power (Margoni, Baillargeon, & Surian, 2018). Building on these results, recent experiments examined whether infants might also hold expectations about one specific type of respect-based power, the legitimate power of an authority figure (Stavans & Baillargeon, 2019). Specifically, these experiments asked whether infants would expect a powerful individual in a group to rectify a transgression perpetrated by one subordinate against another. The rationale was that positive results would suggest that infants cast the powerful individual in the role of legitimate leader and hence expected this leader to restore order in the group, in accordance with the principle of authority.

In these experiments, 17-month-olds watched live interactions among a group of three bear puppets (Stavans & Baillargeon, 2019). One puppet (at the back of the apparatus) served as the leader, and the other two puppets (on the left and right sides) served as the subordinates; in front of each subordinate was a place mat. In different scenarios, the leader was identified either by its larger size (physical cue) or by the subordinates' compliance with its instructions (behavioral cue); results

Authority Violation

Failing to rectify a transgression

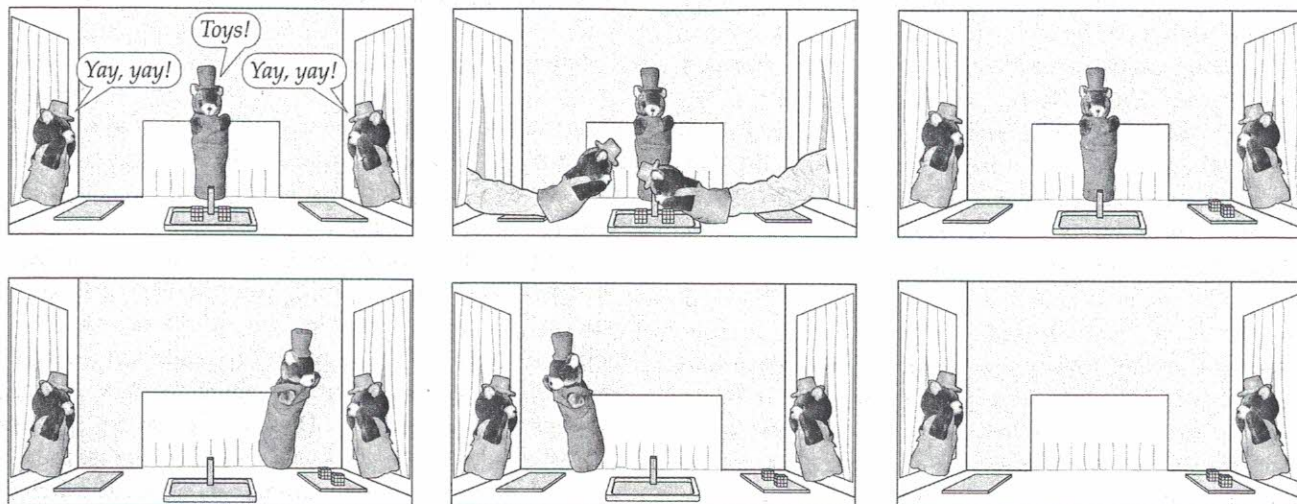


FIGURE 1.3 Infants detect an authority violation when a leader (here marked by its larger size) in a group fails to rectify a transgression between subordinate members of the group.

were identical across scenarios, so the size-based scenario is used here. To start, the leader brought in a tray with two identical toys for the subordinates to share. However, one subordinate (the perpetrator) quickly grabbed both toys and deposited them on its place mat so that the other subordinate (the victim) did not get a toy. In one event, the leader rectified this transgression by taking one of the toys away from the perpetrator and giving it to the victim (*rectify* event). In the other event, the leader again approached each subordinate in turn but did nothing to correct the transgression (*ignore* event; figure 1.3). Infants looked significantly longer if shown the ignore as opposed to the rectify event. This effect was eliminated if the leader was replaced by another member of the group who gave no evidence of being a leader (e.g., another bear of the same size as the two subordinates).

Together, these results suggest that when infants identify an individual as a legitimate leader in a group, they expect this leader to restore order if one subordinate transgresses against another, in accordance with the authority principle.

Conclusions

The evidence reviewed in this chapter suggests that from a very young age a skeletal framework of abstract principles guides infants' sociomoral reasoning. These principles include fairness, harm avoidance, in-group support (with its corollaries of in-group care and in-group loyalty), and authority. Although considerable research is needed to fully understand the "first draft" of human

moral cognition and how experience and culture revises it (Graham et al., 2013), available findings indicate that this "first draft" makes possible surprisingly sophisticated moral expectations, evaluations, and attitudes.

Acknowledgments

This chapter was supported by a Graduate Fellowship from the National Science Foundation to Melody Buyukozker Dawkins, a Fulbright Postdoctoral Fellowship to Maayan Stavans, and a grant from the John Templeton Foundation to Renée Baillargeon.

REFERENCES

- Baillargeon, R. (2008). Innate ideas revisited: For a principle of persistence in infants' physical reasoning. *Perspectives on Psychological Science*, 3, 2–13.
- Baillargeon, R., Scott, R. M., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, 67, 159–186.
- Baillargeon, R., Scott, R. M., He, Z., Sloane, S., Setoh, P., Jin, K., & Bian, L. (2015). Psychological and sociomoral reasoning in infancy. In M. Mikulincer & P. R. Shaver (Eds.), E. Borgida & J. A. Bargh (Assoc. Eds.), *APA handbook of personality and social psychology: Vol. 1. Attitudes and social cognition* (pp. 79–150). Washington, DC: American Psychological Association.
- Baumard, N., André, J. B., & Sperber, D. (2013). A mutualistic approach to morality: The evolution of fairness by partner choice. *Behavioral and Brain Sciences*, 36, 59–78.
- Baumard, N., Mascaró, O., & Chevallier, C. (2012). Preschoolers are able to take merit into account when distributing goods. *Developmental Psychology*, 48, 492–498.

- Bian, L., & Baillargeon, R. (2016, May). *Toddlers and infants expect individuals from novel social groups to prefer and align with ingroup members*. Poster presented at the International Conference on Infant Studies, New Orleans, LA.
- Bian, L., Sloane, S., & Baillargeon, R. (2018). Infants expect ingroup support to override fairness when resources are limited. *Proceedings of the National Academy of Sciences*, *115*(11), 2705–2710.
- Brewer, M. B. (1999). The psychology of prejudice: Ingroup love or outgroup hate? *Journal of Social Issues*, *55*, 429–444.
- Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, *84*, 422–428.
- Buyukozer Dawkins, M., Sloane, S., & Baillargeon, R. (2017, August). *Evidence for an equity-based sense of fairness in infancy*. Poster presented at the Dartmouth Workshop on Action Understanding, Hanover, NH.
- Buyukozer Dawkins, M., Sloane, S., & Baillargeon, R. (2019). Do infants in the first year of life expect equal resource allocations? In J. Sommerville, K. Lucca, & J. K. Hamlin (Eds.), *Frontiers in Psychology*, *10*, article 116 (special issue on “Early Moral Cognition and Behavior”).
- Cosmides, L., & Tooby, J. (2013). Evolutionary psychology: New perspectives on cognition and motivation. *Annual Review of Psychology*, *64*, 201–229.
- Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, *466*, 794–796.
- Dupoux, E., & Jacob, P. (2007). Universal moral grammar: A critical appraisal. *Trends in Cognitive Sciences*, *9*, 373–378.
- Enright, E. A., Gweon, H., & Sommerville, J. A. (2017). ‘To the victor go the spoils’: Infants expect resources to align with dominance structures. *Cognition*, *164*, 8–21.
- Gelman, R. (1990). First principles organize attention to and learning about relevant data: Number and the animate-inanimate distinction as examples. *Cognitive Science*, *14*, 79–106.
- Gergely, G., Nádasdy, Z., Csibra, G., & Bíró, S. (1995). Taking the intentional stance at 12 months of age. *Cognition*, *56*, 165–193.
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130.
- Hamlin, J. K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants’ social evaluations. *Cognition*, *18*, 451–474.
- Hamlin, J. K. (2014). Context-dependent social evaluation in 4.5-month-old human infants: The role of domain-general versus domain-specific processes in the development of social evaluation. *Frontiers in Psychology*, *5*, 614.
- Hamlin, J. K., & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, *26*, 30–39.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature*, *450*, 557–559.
- Hamlin, J. K., Wynn, K., & Bloom, P. (2010). Three-month-olds show a negativity bias in their social evaluations. *Developmental Science*, *13*, 923–929.
- Hamlin, J. K., Wynn, K., Bloom, P., & Mahajan, N. (2011). How infants and toddlers react to antisocial others. *Proceedings of the National Academy of Sciences*, *108*, 19931–19936.
- Jackendoff, R. (2007). *Language, consciousness, culture: Essays on mental structure*. Cambridge, MA: MIT Press.
- Jin, K., & Baillargeon, R. (2017). Infants possess an abstract expectation of ingroup support. *Proceedings of the National Academy of Sciences*, *114*, 8199–8204.
- Jin, K., Houston, J. L., Baillargeon, R., Groh, A. M., & Roisman, G. I. (2018). Young infants expect an unfamiliar adult to comfort a crying baby: Evidence from a standard violation-of-expectation task and a novel infant-triggered-video task. *Cognitive Psychology*, *102*, 1–20.
- Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences*, *104*, 12577–12580.
- Leslie, A. M. (1995). A theory of agency. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 121–149). Oxford: Clarendon Press.
- Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*, 489–512.
- Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants’ reasoning about events involving inert and self-propelled objects. *Cognitive Psychology*, *58*, 441–486.
- Margoni, F., Baillargeon, R., & Surian, L. (2018). Infants distinguish between leaders and bullies. *Proceedings of the National Academy of Sciences*, *115*(38), E8835–E8843.
- Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences*, *109*, 6862–6867.
- Meristo, M., Strid, K., & Surian, L. (2016). Preverbal infants’ ability to encode the outcome of distributive actions. *Infancy*, *21*(3), 353–372.
- Olson, K. R., & Spelke, E. S. (2008). Foundations of cooperation in young children. *Cognition*, *108*, 222–231.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*, 255–258.
- Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York: Viking.
- Powell, L. J., & Spelke, E. S. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, *110*, 3965–3972.
- Pun, A., Birch, S. A., & Baron, A. S. (2016). Infants use relative numerical group size to infer social dominance. *Proceedings of the National Academy of Sciences*, *113*(9), 2376–2381.
- Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*, 57–75.
- Schmidt, M. F. H., & Sommerville, J. A. (2011). Fairness expectations and altruistic sharing in 15-month-old human infants. *PLoS One*, *6*, e23223.
- Shutts, K., Kinzler, K. D., McKee, C. B., & Spelke, E. S. (2009). Social information guides infants’ selection of foods. *Journal of Cognition and Development*, *10*, 1–17.
- Shweder, R. A., Much, N. C., Mahapatra, M., & Park, L. (1997). The “big three” of morality (autonomy, community and divinity) and the “big three” explanations of suffering. In A. M. Brandt & P. Rozin (Eds.), *Morality and health* (pp. 119–169). New York: Routledge.
- Sloane, S., Baillargeon, R., & Premack, D. (2012). Do infants have a sense of fairness? *Psychological Science*, *23*, 196–204.
- Spelke, E. S. (1994). Initial knowledge: Six suggestions. *Cognition*, *50*, 431–445.
- Spelke, E. S., Phillips, A., & Woodward, A. L. (1995). Infants’ knowledge of object motion and human action. In D. Sperber, D. Premack, & A. J. Premack (Eds.), *Causal cognition: A multidisciplinary debate* (pp. 44–78). Oxford: Clarendon Press.

- Spokes, A. C., & Spelke, E. S. (2017). The cradle of social knowledge: Infants' reasoning about caregiving and affiliation. *Cognition, 159*, 102–116.
- Stavans, M., & Baillargeon, R. (2019). Infants expect leaders to right wrongs. Manuscript under review.
- Thomsen, L., Frankenhuis, W., Ingold-Smith, M., & Carey, S. (2011). Big and mighty: Preverbal infants mentally represent social dominance. *Science, 331*, 477–480.
- Ting, F., He, Z., & Baillargeon, R. (2019a, March). *Group membership modulates early expectations about retaliatory harm*. Paper presented at the Society for Research in Child Development Biennial Meeting, Baltimore, MD.
- Ting, F., He, Z., & Baillargeon, R. (2019b). Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proceedings of the National Academy of Sciences, 116*, 6025–6034.
- Tyler, T. R., & Lind, A. (1992). A relational model of authority in groups. *Advances in Experimental Social Psychology, 25*, 115–191.
- Van Vugt, M. (2006). Evolutionary origins of leadership and followership. *Personality and Social Psychology Review, 10*, 354–371.
- Wang, Y., & Henderson, A. M. (2018). Just rewards: 17-month-old infants expect agents to take resources according to the principles of distributive justice. *Journal of Experimental Child Psychology, 172*, 25–40.
- Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition, 69*, 1–34.
- Ziv, T., & Sommerville, J. A. (2017). Developmental differences in infants' fairness expectations from 6 to 15 months of age. *Child Development, 88*(6), 1930–1951.