

# Causing Human Actions

New Perspectives on the Causal Theory of Action

edited by Jesús H. Aguilar and Andrei A. Buckareff

A Bradford Book  
The MIT Press  
Cambridge, Massachusetts  
London, England

© 2010 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.

MIT Press books may be purchased at special quantity discounts for business or sales promotional use. For information, please email [special\\_sales@mitpress.mit.edu](mailto:special_sales@mitpress.mit.edu) or write to Special Sales Department, The MIT Press, 55 Hayward Street, Cambridge, MA 02142.

This book was set in Stone Sans and Stone Serif by Toppan Best-set Premidia Limited. Printed and bound in the United States of America.

"Omissions and Causalism" by Carolina Sartorio. Copyright 2009 Wiley-Blackwell. Reproduced with Permission of Blackwell Publishing, Ltd.

"Intentional Omissions" by Randolphe Clarke. Copyright 2010 Wiley-Blackwell. Reproduced with Permission of Blackwell Publishing, Ltd.

Library of Congress Cataloging-in-Publication Data

Causing human actions : new perspectives on the causal theory of action / edited by Jesús H. Aguilar and Andrei A. Buckareff.

p. cm.

"A Bradford Book."

Includes bibliographical references and index.

ISBN 978-0-262-01456-4 (hardcover : alk. paper)—ISBN 978-0-262-51476-7 (pbk. : alk. paper)

1. Act (Philosophy). 2. Action theory. 3. Intentionality (Philosophy).

I. Aguilar, Jesús H. (Jesús Humberto), 1962—. II. Buckareff, Andrei A., 1971—.

B105.A35C38 2010

128'.4—dc22

2010003187

10 9 8 7 6 5 4 3 2 1

## 14 Teleology and Causal Understanding in Children's Theory of Mind

Josef Perner and Johannes Roessler

In "The Emergence of Thought," Donald Davidson argues that while we have no difficulty in describing, on the one hand, mindless nature and, on the other, mature adult psychology, "what we lack is a way of describing what is in between." He claims that there is a deep and "perhaps insuperable problem in giving a full description of the emergence of thought"; and he expresses relief at not working "in the field of developmental psychology" (Davidson 2001, 128). In this chapter we argue that Davidson was right about the depth and difficulty of the problems involved in describing the emergence of thought. But we think Davidson was unduly pessimistic about the prospect of making progress, empirical and philosophical, with these problems. Indeed we hope to show that describing the emergence of thought may help to shed light on the *nature* of thought.

We should make it clear immediately that our concern here will not be with Davidson's completely general problem, the "conceptual difficulty" he sees with the very idea of attributing beliefs and other propositional attitudes to immature thinkers. For the purposes of this chapter, we simply assume that such attributions can be literally correct. We will be concerned with a much more specific issue: the question of how young children understand intentional action. The problem is akin to Davidson's, though, in that it involves a "conceptual difficulty." According to current philosophical orthodoxy, to understand what it is to act intentionally one has to be able to find actions intelligible in terms of the agent's reason *for* acting, the agent's *purpose* in doing what he or she does. Call this the Reason claim, or R. Current orthodoxy also holds that reasons for action are provided, or constituted, by suitable pairs of beliefs and desires. Call this the Belief-Desire claim, or BD. (The classical source for both R and BD is Davidson's [1963] paper "Actions, Reasons, and Causes.") Put together, R and BD suggest that to see people as acting intentionally you have to think of them as acting on the basis of what they believe: understanding

intentional action requires a grasp of the explanatory role of beliefs. There is a large body of evidence to suggest that 2- and 3-year-olds do not satisfy this condition. On the other hand, there is also convincing evidence that 2- and 3-year-olds do have some grasp of what it is to act intentionally. Hence the "conceptual difficulty": the hypothesis that children have some understanding of intentional action enjoys empirical support, yet it is hard to see how it can be sustained, given that children apparently lack a basic conceptual prerequisite for such understanding.

In what follows we develop and defend the following diagnosis. What is to blame for our "conceptual difficulty" is a dogma of contemporary philosophy of mind, that reason-giving explanations of actions are explanations in terms of beliefs and desires. The suggestion we will pursue is that young children find actions intelligible in terms of reasons that are not conceived as mental states at all—but that can nevertheless intelligibly be taken to provide causal explanations. We'll call this the teleological account of children's conception of intentional action. We argue that the teleological account is of more than developmental interest. Its genealogy of commonsense psychology has an important bearing on the nature of the adult conception of intentional action. For, as Wittgenstein remarked: "the thing about progress is, it tends to look more momentous than it really is."<sup>1</sup>

## 1 The Puzzle

What evidence is there to suggest that young children have some understanding of intentional action? Some psychologists maintain that even toward the end of the first year, as infants begin to engage in joint attention interactions with others, they perceive and understand others' actions as goal-directed (Tomasello 1999). Others have argued that such understanding manifests itself in 18-month-olds' more sophisticated capacities for imitation of intended actions (Meltzoff 1995). Here we will focus on evidence provided by (slightly older) children's performance on classical false-belief tasks. The question put to children in such tasks is what the protagonist in some story will *do* next. For example: Suppose Maxi's mother transferred the chocolate Maxi put into the green kitchen cupboard to the blue cupboard while he is out playing. Maxi, feeling peckish, returns to the house. *Where will he look for the chocolate?* Three-year-olds' performance on this task is poor, but far from random. They reliably predict that Maxi will look in the blue cupboard. They don't suggest that he will look under the kitchen table or in the playground or in the loft. What explains this

general pattern in 3-year-olds' performance? The natural, and standard, interpretation is that they tend to predict what Maxi will do on the basis of *what it makes sense for Maxi to do*, given the actual location of the chocolate. If you want to find the chocolate, looking in the blue cupboard is precisely what you should be doing—what you have a good reason to do. Thus, on the natural interpretation, children do think of Maxi as an intentional agent. They assume that Maxi acts in a purposive manner; that he will do what, given his purpose, he has good reason to do.

The puzzle, then, is this. Young children's performance on false-belief tasks simultaneously provides evidence that they think of others as intentional agents and that they have a systematic deficit in understanding the explanatory role of beliefs.<sup>2</sup> Yet, if recent philosophical orthodoxy is right, failure to grasp the explanatory role of beliefs should make it impossible to understand what it is to act intentionally. For (to repeat) recent orthodoxy holds that

R: Understanding what it is to act intentionally requires finding actions intelligible in terms of agents' reasons for acting

and

BD: Reasons for action are provided, or constituted, by suitable pairs of beliefs and desires.

## 2 Desire Psychology

There is, in the developmental literature, an influential view that may seem to offer a simple solution to our puzzle. The basic idea is that while thinking of someone as acting intentionally certainly requires understanding something about the mental states causally responsible for intentional actions, such understanding may be more or less comprehensive. A fully developed conception of intentional action will require a large and complex set of psychological notions, including, of course, the notion of belief. But, the claim is, children may have a rudimentary grasp of intentional action in virtue of understanding something about the explanatory role of *desires*, without yet appreciating how desires tend to interact with other states, such as beliefs (Bartsch and Wellman 1995).

Straight off, though, it is not obvious how this suggestion speaks to our puzzle. To understand Maxi's behavior as intentional, you have to *put two things together*: the purpose of the action and the means by which Maxi seeks to accomplish his purpose. A simple "desire psychology" may enable you to identify Maxi's purpose (his purpose is to get hold of his chocolate),

but it is silent on what he might be expected to do to accomplish his purpose. So, on the face of it, the proposal has nothing to say on how children understand the intentional, purposive nature of the instrumental action they expect Maxi to perform (his opening the blue cupboard). Of course, you might say that the desire to locate his chocolate is at least part of what gives Maxi a reason to act. But understanding part of what gives someone a reason for acting intentionally is not the same as a partial understanding of intentional action.

Perhaps one reason the problem tends to go unnoticed is this. It is common practice in developmental psychology to equate goals with desires (Gopnik and Meltzoff 1997;<sup>3</sup> Tomasello et al. 2005). On this way of thinking, the ability to explain bodily movements in terms of desires may be redescribed as the ability to explain actions in terms of goals. The latter ability, in turn, is naturally equated with an understanding of *goal-directed action*. This can make it look as if a grasp of the explanatory role of desire just is a grasp of goal-directed action. But that appearance is deceptive. To ascribe to children a conception of goal-directed action is to say that they have mastered a distinctive way of explaining actions. They understand what it means for someone to do one thing *in order to* achieve another. This mode of explanation essentially involves putting together someone's purpose in acting with a particular way of pursuing the purpose. Put differently, it involves an understanding of instrumental reasons for action. In contrast, explaining a bodily movement in terms of some mental cause, even if that cause is a desire, does not necessarily involve any such "putting together" of end and means.<sup>4</sup> A "desire psychologist" might correctly explain someone's salivating in terms of a desire for chocolate. This does not mean that she thinks of the activity as an intentional action. The ability to explain bodily movements as the effects of desires is not sufficient for even a rudimentary understanding of what it is to act intentionally.

### 3 Objective Reasons

The conception of practical reasons articulated by BD (practical reasons are constituted or provided by suitable pairs of beliefs and desires) is arguably not the only conception familiar to commonsense psychology. We often think of practical reasons not as mental states but as worldly facts. When deliberating about what to do—say, where to look for some prized object—we tend to be more interested in the facts than in our beliefs about the facts. We think of the *fact* that looking in a certain location will enable us

to retrieve the object as giving us a reason to look there. From the perspective of deliberation, only true propositions—facts—can provide genuine reasons. This point is not inconsistent with BD. The claim is not that BD is false. It is only that BD does not exhaust the commonsense psychology of practical reasons. In this observation, we suggest, lies the solution to our "conceptual difficulty." Young children find intentional actions intelligible in terms of "objective" practical reasons.

Let's clarify the basic idea with the help of a relatively uncontroversial example from Bernard Williams. Suppose you believe of the content of a certain bottle that "this stuff is gin," when in fact the bottle contains petrol. You feel like a gin and tonic. Should we say that you have a reason to mix the stuff with tonic and drink it? Williams suggests that you do not have such a reason, although you think you do, and although it would certainly be rational for you to drink the stuff (Williams 1981b, 102). One might object that to say that it would be rational for you to drink the stuff just is to say that you have a reason to drink it—a reason that might be appealed to, for example, in offering a "reason-giving" explanation of your action. But Williams is surely right in drawing our attention to the fact that you are *mistaken* in thinking you have a reason to drink the stuff. Your putative reason can be set out as follows: "I need a gin and tonic. This stuff is gin. So I should mix it with tonic and drink it." Certainly the inference reveals your action to be rational from your point of view. Correlatively, appeal to the inference may figure in a reason-giving explanation of your action. But the fact remains that the inference is unsound. Given that the second premise is incorrect, the inference fails to establish the truth of its conclusion: you are mistaken in taking the premises to establish that you should drink the stuff. In this sense, you are mistaken in thinking you have a reason for drinking it. This is perfectly consistent with acknowledging that there is a sense in which you do have such a reason. The point is sometimes put by saying that you lack a *justifying* or *guiding* reason, but have an *explanatory* reason to drink the stuff (Raz 1978). But this can be misleading, given that an explanatory reason too may be said to justify, at least in the "anaemic" sense (Davidson 1963/1980) of revealing the action to be justified or rational *from your perspective*. Marking the distinction as one between justifying and explanatory reasons would, in the current context, be awkward in another way. For the suggestion we want to pursue would now (confusingly) have to be put by saying that young children *explain* intentional actions in terms of *justifying* (rather than explanatory) reasons. So we'll simply distinguish between *subjective* and *objective* reasons: you have a subjective reason to drink the stuff, but you lack an objective

reason to do so. Note that by itself the distinction does not imply that we are talking about different sorts of *things* here. Subjective reasons need not be taken to *be* mental states. Instead they might be taken to be propositions forming the contents of mental states. The distinction turns on whether or not a reason statement is to be understood as relativized to the agent's current perspective. To say that you have a subjective reason to drink the stuff is to say that, from your perspective, it looks as if you have an (objective) reason to do so.

It might be said that even objective reasons have to be relativized in one respect: they must be relativized to the agent's set of desires or objectives or projects. But on reflection it is not clear that this is so. As is often pointed out by critics of the Humean theory of motivation<sup>5</sup> (and as is acknowledged by some of its defenders<sup>6</sup>), practical deliberation does not always or even typically start from reflection about one's current desires. Practical inferences are often premised on evaluative propositions, to the effect that some action or some state of affairs is important or desirable; or, more specifically, on propositions involving "thick" evaluative concepts (e.g., promise, treachery, brutality, courage), apparently embodying "a union of fact and value" (Williams 1985, 129); or again, as in the example above, on propositions to the effect that someone needs, or needs to do, a certain thing, where claims of need are also best understood as a species of evaluative propositions, not to be confused with, or reduced to, ascriptions of desire (Wiggins 1987).<sup>7</sup> This suggests that there may after all be such a thing as a fully objective reason, relativized neither to the subject's instrumental beliefs nor to her set of desires and projects. A possible illustration might be the suggestion that the subject in Williams's example not only lacks an (objective) reason to drink the stuff, but actually has an (objective) reason *not* to drink it—a reason provided by the fact that drinking petrol is bad for your health.<sup>8</sup>

#### 4 Teleology and the Hybrid Account

We are now in a position to formulate two possible solutions to our puzzle. One proposal might be that young children find actions intelligible in terms of partially objective reasons: reasons that are relativized to the agent's desires but not to her instrumental beliefs. For example, children might take it that given Maxi's desire to eat some chocolate, he has a good reason to look in the blue cupboard (for doing so will as a matter of fact enable him to satisfy his desire). Put differently, they conceive of Maxi's reason as a combination of a desire and an objective instrumental fact.

Children predict and explain intentional actions by "putting together" a mental state, defining the objective of the action, and an objective fact, determining the means to achieve the objective. We will call this the *hybrid account* of children's conception of intentional action. (The hybrid account is one way of developing the idea that young children are "desire psychologists."<sup>9</sup>) The second solution agrees with the hybrid account that understanding the way young children think about intentional actions requires going beyond BD. But it gives a more radical twist to that strategy. It maintains that children find actions intelligible in terms of fully objective reasons, relativized neither to the agent's instrumental beliefs nor to her pro-attitudes.<sup>10</sup> They conceive of Maxi's reason in terms such as these: "Maxi needs his chocolate. (Or: It is important, or desirable, that Maxi obtain his chocolate.) The way to get it is to look in the blue cupboard. So he should look in the blue cupboard." We call this the *teleological account*.<sup>11</sup>

How might we tell which of the two explanatory schemes young children go in for? We can spell out the key difference between them as follows. Finding actions intelligible in terms of objective reasons implies a certain lack of detachment. It involves commitments we do not normally associate with reason-giving explanations (accustomed as we are to taking such explanations to be a matter of citing subjective reasons). The two explanatory schemes differ in the kinds of commitments they entail. An explanation following the hybrid scheme carries with it a commitment concerning the effectiveness of some means for achieving the agent's purpose. The teleological scheme in addition commits the interpreter to some evaluative proposition. Both schemes share a crucial limitation: they break down when it comes to actions informed by false instrumental beliefs. But the teleological scheme is more limited still. It also breaks down in relation to actions informed by values the interpreter does not share. (Thus the teleological account might be said to involve an unusually strong version of the principle of charity.) There are two obvious sorts of cases where this extra limitation makes itself felt. One is the case of interpreting actions informed by objectionable values (e.g., hurting someone because it's fun to do so). The other is the case of interpreting the activities of two agents pursuing conflicting goals. Consider a competitive game. If you understand A's activities in terms of the fact that it's desirable that A win the game you will be hard pressed simultaneously to understand B's reasons for his contributions to the game, on pain of finding yourself committed to inconsistent evaluative propositions: It's desirable that A should win (by beating B), and it's desirable that B should win (by beating A).<sup>12</sup> Thus the

teleological account would predict that young children have difficulties in understanding competitive games; more precisely, that they have difficulties in understanding the reasons informing the activities of participants in such games. By contrast, if young children subscribe to the hybrid scheme, they should find it easy to understand each of the competing activities, viz. in terms of the desires of each of the protagonists. (That A wants to win is plainly consistent with B's wanting to win also.) In general, for a teleologist, cooperation is utterly natural; practical reasons are subject to a kind of preestablished harmony. As conceived by the teleologist, reasons are essentially intersubjective. If we explain Maxi's action in terms of the fact that he needs his chocolate, reference to that very fact may also help to understand the actions of other agents. For example, the fact may explain why someone else is doing something that will assist Maxi in retrieving his chocolate.

How should we understand the issue between the hybrid view and the teleological account? And what, if anything, is there of *philosophical* significance in this debate? Well, it certainly looks initially as if the issue were a straightforwardly empirical one. And there is some suggestive initial evidence in favor of the teleological account: young children do seem to have difficulties in understanding both wicked behavior and competitive games. On the other hand, some will be inclined to doubt whether the teleological scheme is even intelligible. If this were right, it would be implausible to attribute the scheme to any rational thinker (even to very young thinkers). Thus the teleological account might be rejected on *a priori* grounds. One philosophical issue, then, is whether an objection of this sort can be sustained. As will become clear later, there is also much of philosophical interest in interpreting the empirical evidence. But there is a further way in which the debate bears on the philosophy of action. Suppose the teleological account is intelligible and enjoys a measure of empirical support—as we will argue it is and does. Suppose, in other words, that at least a rudimentary conception of intentional action is available prior to, and independently of, grasping concepts of propositional attitudes such as belief or various kinds of pro-attitudes. This would raise the question of how to understand the relation between the teleological conception and the mature adult conception of intentional action. We return to this question in section 7.

## 5 Teleology and Causal Understanding

We begin with the issue of intelligibility. A preliminary point to make about the idea of objective practical reasons is that it amounts to nothing

very grand or eccentric, metaphysically speaking. It is not the idea that values belong to the fabric of the world as it is in itself, or that we have, or need, a special faculty for detecting values. It is merely the idea that there are true evaluative propositions. While this commits one to the view that evaluative statements can be true or false, such a commitment need not be seen as a particularly contentious matter. For it may be argued that, as Williams put it, "*truth in itself* isn't much" (Williams 1996, 19). The central point here is that evaluative statements show the sorts of "surface phenomena" (e.g., possible embedding in conditionals) that make it plausible to think of them as bearers of truth-values. There is no commitment to any substantive account of what it is in virtue of which such statements are true (nor even to the possibility of any such account).

If there is a serious worry about the intelligibility of teleology, it is surely over the claim that objective reasons can intelligibly be regarded as *explanatory*. One version of the challenge may be summarized as follows: "The teleological scheme purports to explain actions in terms of facts that are taken to constitute justifying reasons for it. Now, as Davidson has taught us, it is one thing for someone to have a reason to perform a certain action; it is another for that reason to explain her action—to be the reason *for which* the agent acts. (For one thing, the agent may simply fail to do what she has a reason to do. For another, even if she does it, and does it intentionally, she may do it for some other reason.) There are well-known and powerful considerations that suggest the notion of causation holds the key to a proper understanding of the difference: reason-giving explanations have to be conceived as a species of *causal* explanations (Davidson 1963/1980). But of course, causal explanations have to appeal to causes. Yet causes are conspicuous by their absence in the teleological scheme."

One might wonder whether the hybrid view fares any better than the teleological account vis-à-vis the Davidsonian challenge. After all, the hybrid account, too, envisages action explanations in terms of external facts rather than just in terms of mental states. But it might be said, quite plausibly, that there is a key difference. Suppose the upshot of the Davidsonian challenge is that, as one recent commentator put it, intentional actions are to be explained "in terms of mental *items*—such as beliefs, desires, intentions, and associated events . . . or their physical realizers" (Mele 2003, 51, our emphasis). The idea is that for reason explanation to be intelligible it has to make reference to *particulars*—events or, perhaps, states—that stand in a causal relation to someone's doing something intentionally. The teleological account does not provide for that requirement. Values are not events. It would be fanciful to think of Maxi's need for chocolate, or the importance of his getting his chocolate, as an "item" that

stands in a causal relation to the event of his opening the green cupboard. In contrast, the hybrid view may seem to provide for causes, conceived as particulars. True, even on the hybrid view, the explanatory force of early reason explanations cannot be exhaustively explained in terms of causal relations between "mental items" and actions. External facts also play a vital explanatory role. But defenders of the hybrid view might argue that such facts should be seen as "standing conditions," whose explanatory role essentially depends on that of causally efficacious "items." Their role may be not unlike that of the dryness of the ground in an explanation of a forest fire caused by someone's dropping a cigarette. In brief, the Davidsonian challenge might seem to provide materials for an *a priori* argument in favor of the hybrid view.

We want to suggest that this argument rests on an implausible premise. The Davidsonian challenge consists of two central claims. One is that action-explanation must be a species of causal explanation. This is usually motivated by arguing, convincingly, that action-explanations are explanations of the occurrence of events, and that it's hard to see how the occurrence of an event can be made intelligible other than in causal terms. The second claim is that causal explanation has to appeal to causes, conceived as particulars. The correct teleological response to the *a priori* argument for the hybrid view, we suggest, is to accept the first but reject the second claim. The key question here is: What does it mean for an explanation to be causal? Adapting Bernard Williams's remark about truth, perhaps the right thing to say here is that causal explanation *in itself* isn't much. It's not explanation in terms of laws of nature; it's not explanation in terms of event causation; it's not explanation in terms of causal processes or causal mechanisms. The basic idea of a more minimalist account is that causal explanations advert to facts that "make a difference," where this is to be spelled out in terms of patterns of counterfactual dependence.<sup>13</sup> Our suggestion, then, is that teleological explanation is a species of causal explanation. So we agree with the idea underpinning the Davidsonian challenge to teleology, that the development of the commonsense conception of intentional action is inextricably entwined with the development of causal understanding. But we deny that causal understanding in this area takes the form of understanding causal relations between "mental items" and actions.

Specifically, the version of the difference-making approach we propose to draw on is the so-called interventionist approach to causation and causal explanation (Woodward 2003). The central idea of interventionism is this. To say that there is a causal relation between two variables *X* and *Y* is to

say that if there were to be an intervention changing the value of *X*, there would also be a change in the value of *Y*. To say that taking aspirin causes relief from headache is to say that if there were an intervention on the amount of aspirin ingested by someone suffering from headache, there would also be a change in the intensity of her headache. The notion of an intervention is itself a causal notion. Crudely, an intervention on *X* is a causal relation between some variable *I* and *X* such that *I* causes a change in the value of *X* in an "exogenous" way—*I* takes complete control of *X* without affecting *Y* in any other way than through *X*.<sup>14</sup> This sort of account aims to elucidate the meaning of causal claims in terms of interventionist counterfactuals—in terms of how the value of *Y* would change under interventions on *X*—without, of course, undertaking to offer a reductive analysis.

Now on an interventionist approach there is, on the face of it, nothing *a priori* objectionable about the idea that objective reasons may causally explain intentional actions. The idea comes to this: were there to be an intervention on the evaluative or instrumental facts that give someone a practical reason, there would be a corresponding change in her action. The suggestion that young children think about intentional actions in this way is not at all implausible, especially in light of the sorts of social interaction characteristic of early childhood. Consider games of mutual imitation. The simple rule defining such games is that one participant has to imitate the actions of the other. In the context of this game, I can reliably manipulate your actions by intervening on your reasons. For example, my banging my toy gives you a reason to bang yours. (Only by banging yours can you achieve the defining purpose of the game, viz., to imitate my action.) So the fact that I am banging my toy offers a good causal-teleological explanation of your action, a causal explanation in terms of reason-giving facts.<sup>15</sup> To grasp such explanations it is not enough to understand that you have a reason to perform the action. Children also have to understand the reason as explanatory of the action. This they do, in virtue of their developing understanding of interventionist counterfactuals.<sup>16</sup> On this view, the causal understanding that goes into children's conception of intentional action does not require, or consist in, a conception of unobservable entities—"mental items"—conceived as causes of bodily movements. Rather, it is a matter of understanding what will, or would, happen under interventions on someone's practical reasons.

Note that the teleological account is concerned exclusively with the core element of children's understanding of intentional action, their capacity for reason-giving explanation. The account does not imply that the

capacity operates in isolation. Nor does it imply that psychological properties play no role in such explanations. One way in which teleology may be enriched is by adding the idea, which may be in place quite early in development, that people's responsiveness to their practical reasons is subject to certain enabling (and disabling) conditions. For example, the agent has to be awake, not too tired or distracted, she may have to be able to perceive some critical object, and so on. Of particular interest here is the notion of attention: even young children may have some understanding of someone's attention or "engagement,"<sup>17</sup> as something that has to be suitably focused if the person is to respond to the reasons afforded by a situation. Children's understanding of the causal relevance of this condition may manifest itself in their practical interventions on others' focus of attention; for example, their attempts at attracting someone's attention by pointing or shouting. The point is important since it helps to allay the worry that the resources of teleology are simply too limited to offer a credible picture of early social understanding. Pure teleology admittedly leaves many things unexplained. A teleologist who thinks it's time for a game of mutual imitation should be baffled if his chosen partner shows herself to be unresponsive. Of course, being unable to make rational sense of an action need not prevent a teleologist from registering that the action (or inaction) is taking place. (Much of the time, mature commonsense psychologists are in the same position.) Still, the limitation may seem crippling. But there are two ways in which teleologists may overcome this limitation. An initially inscrutable action may become intelligible once the teleologist comes to understand, and share, the value informing it.<sup>18</sup> Alternatively, someone's (in-)action may remain rationally unintelligible, but the agent's failure to respond to her reasons may itself be unsurprising given the satisfaction of some disabling condition. (She may be asleep.) Teleologists need not be at sea in the social world.

What we suggest would be a *non sequitur* is to move from the involvement of psychological material in young children's understanding of intentional agency to the conclusion that they must be finding actions intelligible in terms of subjective reasons—reasons relativized to the agent's cognitive/evaluative perspective. There may be important roles for the psychological material to play other than to delineate what, relative to the agent's subjective perspective, it is rational for her to do. (The same is true in the case of mature commonsense psychology.) It is no trivial task to decide whether a particular example of early mentalism manifests an understanding of subjective reasons—as becomes clear when we turn to the experimental evidence.

## 6 The Evidence

It is sometimes said that children appreciate the subjectivity of desires many years before they grasp the subjectivity of belief (e.g., Repacholi and Gopnik 1997; Rakoczy, Warneken, and Tomasello 2007). There is more than one way to understand this claim. On one reading, it amounts to an endorsement of the hybrid view: children are able to understand the relativity of practical reasons to the agent's desires before they are able to relativize practical reasons to the agent's beliefs. We want to suggest that, on this reading, the claim lacks empirical support. There is a rough distinction we can make between two sorts of evidence that are relevant here. On the one hand, there is evidence that might be appealed to in support of the hybrid view. We discuss two such findings: concerning infants' understanding of subjective preferences, and concerning verbal references to desires in action-explanation. On the other hand, there are findings that appear to support the teleological account. As will become apparent, our empirical defense of the teleological account is somewhat qualified. While there is some suggestive evidence in its favor, and no persuasive evidence against it, we believe that more tightly controlled experimental work is needed to make further empirical progress with these matters.

### 6.1 Subjective Preferences

Even 18-month-olds seem to be aware of the subjectivity of desire in the following sense: they are sensitive to individual preferences. Evidence for this comes from a study where infants were presented with two kinds of food (yummy crackers and yucky broccoli). In a control condition, when another person asked them for something to eat, they almost always gave that person a cracker rather than a piece of broccoli. However, in the discrepant desire condition the other person first demonstrated deviant preferences. Taking the crackers she exclaimed "Yuck, that's awful," underlined by the appropriate facial expression. When she tasted broccoli she looked pleased and also said so: "Mmm, that's good!" Now when this person later asked the infant for something to eat, most children younger than 18 months still handed her the crackers, but most children over 18 months handed her the broccoli (Repacholi and Gopnik 1997). A similar finding is that 2-year-old children freely verbalize subjective preferences: "Daddy likes shaving cream. I hate shaving cream" (Bartsch and Wellman 1995).

One sort of question raised by these findings is how we should characterize children's conception of subjective preferences. Questions under this heading include the following: Do children think of "liking" as a mental



state, something like a desire or an emotion—associated perhaps with a distinctive experience—or rather as something like a character trait (on the lines of “he likes to be in charge”)? Another question raised by the findings is: what sort of explanatory role do subjective preferences occupy in children’s theory of mind? Do children think of subjective preferences as desires in the sense of BD—as part of what makes it rational for the subject of the desire (but for no one else, unless he or she has a desire to cooperate) to use suitable means to satisfy the desire? Or do they think of subjective preferences in some other way, giving them a different sort of explanatory role?

Note that these are quite different issues. Suppose children do think of subjective preferences as involving conscious mental states, say, desires for objects. This should not automatically lead us to assume that they conceive of such desires as states that provide agents with subjective reasons. That would be to ignore the possibility that children think of “likings” as being essentially linked to *objective* purposes. Thus children’s conception of subjective preferences may be part of their conception of objective needs. Subjective preferences may be taken to determine what is good for different sorts of people or creatures. Broccoli is good for A, but not for B. Eating grass is good for cattle but not for humans. This is a relatively sophisticated (perhaps in one sense, nonegocentric) conception of needs. But it does not require any understanding of subjective reasons. On this conception, the import of A’s and B’s differential preferences is that it is objectively desirable that A, but not B, should obtain some broccoli (Perner, Zauner, and Sprung 2005). Correlatively, there is no privileged explanatory relation between A’s preference and A’s action: A’s preference makes it as reasonable, and as intelligible, for B to pass A a piece of broccoli as for A to set off for the greengrocer. The findings concerning subjective preferences are consistent with this interpretation, and hence with the teleological account.

## 6.2 Talk about Desires

We mentioned two potential instances of early psychological understanding: children’s grasp of attention (and perhaps perceptual experience), and their understanding of subjective preferences. Neither of them has any tendency to suggest that children appeal to mental states as providers of subjective reasons; both are naturally incorporated into the teleological scheme of explanation. But there is another sort of connection between teleology and psychology. There is a sense in which teleology *provides* for certain psychological notions. If the teleological account is right, finding

someone’s actions (causally) intelligible in terms of reason-giving worldly facts provides for a primitive notion of intentional action—arguably, a psychological phenomenon. Now setting out a teleological explanation in full would require reconstructing the practical inference that articulates the reason for which the agent performed the action, with an evaluative and instrumental premise and a conclusion to the effect that, say, the agent should open a particular cupboard. A more minimalist way of explaining the action would be merely to identify its goal. Rather than saying Maxi opened the cupboard because he needs his chocolate etc., a teleologist may simply tell us that it was in order to get his chocolate that Maxi opened the cupboard. Or: Maxi opened the cupboard with the intention of taking out his chocolate. In this way, the teleological scheme of explanation can be seen to provide for a simple notion of intention (the intention with which someone acts). This is arguably a psychological notion in good standing, but it is not one that helps to explain actions on the model of BD. The intention with which someone acts, as it is understood by a teleologist, may be classified as a (kind of) desire, but a teleologist does not conceive of it as a state that provides the agent with a subjective reason—a desire that, together with a belief, helps to rationalize intentional actions. Rather, that an agent has this intention *follows* from the fact that his action is open to a reason-giving, teleological explanation.

The point is significant given that it may help to interpret some interesting data concerning early verbalizations. Children talk about desires substantially earlier than about beliefs (see Bartsch and Wellman 1995). Much of this talk is a matter of expressing their own requests and wishes. Some of it is a matter of reflecting on sometimes contrasting subjective preferences (see above). Least frequently, but perhaps most interestingly, some examples have been recorded of early appeals to desires in action-explanations. A striking illustration is the following (Bartsch and Wellman 1995, 118):

Ross (2;10): Look, there’s a car up in the air. [on a hoist]

Adult: Oh. Why is it up there?

Ross: Man put it up there.

Adult: Why did he put it up there?

Ross: He want to fix it.

On one reading, “want to fix it” is to be construed as an ascription of a mental state distinct from, and causally responsible for, the intentional action in question—a desire that both gives the man a subjective reason to put the car up in the air and causally explains his doing so. But there

is an alternative reading. The child may make the man's action intelligible by stating the intention with which it was performed: his intention in putting the car up was to fix it. The example reveals the child's understanding that people act on the basis of good reasons, and provides a partial reconstruction of the reason operative in this case, identifying the purpose informing the action. Note that on this reading, it's possible that Ross takes the reason to be provided by the objective desirability of the man's fixing the car, rather than by the man's desire to fix it. Thus, the example provides no evidence against the teleological account.

### 6.3 Competitive Games

When introducing teleology we mentioned the importance of two sorts of cases: actions informed by (to the interpreter) unacceptable goals (e.g., throwing a ball at someone because one finds it desirable to hurt her); and competitive actions. If children make sense of intentional actions via the teleological scheme, they should find these sorts of cases unintelligible. If they use the hybrid scheme, these cases shouldn't pose any special challenge. As we indicated, there is some evidence to suggest that young children do find these cases challenging. Now the evidence relating to unacceptable goals (Yuill 1984) relies on an experimental paradigm that, at least in the current context, is to be treated with some caution (as we argue in appendix 2). So we will focus here on the second kind of case.

A good test case between teleology and the hybrid view arises when it comes to understanding competitive behavior as intentional actions, done for a reason. In a competitive (zero-sum) game (as are most simple strategic or probabilistic games) one player's win means the other player's defeat. So their respective contributions to the game cannot be understood as attempts to attain some objectively desirable end, valid for both. The teleological account therefore predicts that competition involving reasoned actions can only be appreciated by children who understand differences of perspective, including the case of false beliefs. In contrast, on the hybrid theory, understanding competitive behavior should pose no particular challenge: children should be able to appreciate that relative to their respective objectives (to win the game), both A's and B's behavior makes rational sense. The question is where to find relevant developmental evidence. The only directly relevant evidence that we are aware of is based on the hand-guessing (commonly also known as "penny hiding") game used in a study by Gratch (1964).

The penny-hiding game is a two-person game in which the subject is actively involved either as a guesser or as a hider. Both try to win. The

hider hides the penny in one hand or the other, and then invites a guess. This is repeated over a row of trials, after which the participants change roles. In the role of hider "the child was judged to be *competitive* if he expressed displeasure on any of the trials in which E found the marble, or if any of the following events occurred: 1. when E selected the marble-holding hand, S refused to show the marble or made an attempt to transfer it to his other hand; 2. S extended an empty hand for E to guess from. A child was judged to be *non-competitive* if none of the above events occurred. The non-competitive children frequently told E where the marble would be hidden, after a trial in which E had failed to guess correctly; and many of them extended the marble in an open hand on all trials" (Gratch 1964, 53–54).

The proportion of children who displayed competitive spirit increased steadily from 5 percent below the age of 3 years to 58 percent at 4.5 years to practically 95 percent at around 6 years. This fits very well the typical developmental trend on false-belief tests (see Perner, Zauner, and Sprung 2005, figure 4, for a graphic display). Gratch's study is particularly helpful insofar as he analyzed indicators of competitive spirit separately from indicators of the ability to deceive, fool, or conceal information from the opponent. Since deception requires an understanding of false belief, a developmental link between understanding the deceptive aspects of the game and understanding false belief would not be terribly interesting and no support for the teleological theory developed here. The shortcoming of Gratch's data for present purposes is, of course, that there is no direct comparison of how many of the children in his sample would have passed the false-belief test. Ideally one would also see the use of a control task with similar cognitive demands except for the competitiveness and in which children of all ages can succeed.

The good news is that more recently several studies included the penny-guessing game and the false-belief tasks (Baron-Cohen 1992; Chasiotis et al. 2006; Hughes and Dunn 1997, 1998). The bad news is that, without exception, these studies only analyzed the hand-guessing behavior for indicators of deceptive abilities (or a mix of combative spirit and deception). Hence the reported correlations with false belief understanding provide no convincing evidence against the hybrid theory and, therefore, also no support for teleology.

We should also mention two other pieces of evidence that go well with teleology. Although children seem to have problems understanding the point of competition, they are quite concerned about obeying the rules of a game (Rakoczy, Warneken, and Tomasello 2008) even at the age of 2

years when they just start to use "desire" terms for other people (Bartsch and Wellman 1995; Imbens-Bailey, Prost, and Fabricius 1997). At 3 years (36 months) this concern becomes almost obsessive. Clearly they expect people to act a certain way because it is the right, conventional way, and they seem to have little understanding for idiosyncratic deviation.

Following the pioneering work by Shultz and Shamash (1981), several studies reported that children have difficulty distinguishing intentions from desires (see reviews by Astington 1999, 2001). The latest study by Schult (2002) included children as young as 3 years. They had to toss bean bags into three different colored buckets, some of which contained a ticket for a prize. For each toss they had to indicate which bucket they intended to hit. On some trials they hit the intended bucket, on others they missed it; on some they won a prize, on others they didn't, resulting in four different combinations. The 4- and 5-year-olds were remarkably accurate in answering all types of questions. The 3-year-olds, on the other hand, had serious problems with questions about their intentions, in particular when satisfaction of their intention contrasted with satisfaction of their desire, as shown in table 14.1.

This pattern of results follows from our assumption that children remain basic teleologists until about 4 years, when they can understand differences of perspective. They have no problem knowing what they want, i.e. the desirable goal of the action (winning the prize), and whether they got it or not. They also understand intentions to hit a particular bucket, though only insofar as there are *objective* reasons for such intentions. Consider now a case of fortuitous success, where children accidentally get the prize after hitting a bucket they didn't intend to hit. To understand that they didn't intentionally hit the bucket, children have to understand that they had

**Table 14.1**

Data from Schult (2002): number of children giving correct answers to the satisfaction questions, "Did you do what you were trying to do?" (intention), and "Did you get what you wanted?" (desire).

Condition (correct answer)	3-Year-Olds	4-Year-Olds	5-Year-Olds
<i>Miss target/get prize: fortuitous success</i>			
Intention ("no")	4/15	8/8	8/8
Desire ("yes")	12/15	8/8	8/8
<i>Hit target/no prize: bad luck</i>			
Intention ("yes")	8/15	8/8	8/8
Desire ("no")	13/15	8/8	8/8

no reason for hitting that particular bucket, despite the fact that doing so turned out to be conducive to reaching their goal. Or consider a case of bad luck, where they hit a certain bucket without getting a prize. To understand that they hit the bucket intentionally, children have to understand that they did have a reason for hitting that bucket, despite the fact that doing so turned out *not* to be conducive to reaching their goal. Under the teleological interpretation, it is unsurprising that young children have problems under these kinds of circumstances. Correct judgment of these cases only becomes possible when one understands that one acted on the assumption that the prize might be in the bucket one was aiming for. Since in the critical cases this assumption turns out to be false, the intentionality of the intended action can only be understood if one can understand it in terms of the perspective of that assumption.

In sum, there is some suggestive evidence against the hybrid theory and some support for the teleological account in Gratch's finding that children show little competitive spirit before the age at which they are able to understand false belief as a motivating reason. More importantly (for the developmental psychologists), our account also provides us with a clearer analysis of what young children should find difficult about competition and incompatibility of goals. In appendix 2, we discuss further evidence that is *prima facie* relevant to the debate between teleology and the hybrid view (evidence concerning children's understanding of emotional reactions to the satisfaction or frustration of desires). In appendix 3, we present an outline of a new experimental paradigm to test for children's ability to understand competitive actions.

## 7 Teleology in Perspective

We conclude with a brief suggestion as to how the teleological account bears on our understanding of adult commonsense psychology. Suppose young children find intentional actions intelligible in terms of reason-providing facts, rather than mental states. We can distinguish two different ways in which this initial understanding subsequently gets refined and enriched, corresponding to two distinct elements of the mature adult conception of intentional action. Both involve giving explanatory relevance to an agent's perspective, though in quite different ways.

First, a central element of a more sophisticated conception of practical reasons is an appreciation of the role of knowledge and ignorance. We may think of this as a more subtle application of the possibly earlier developing idea, that agents' responsiveness to reasons is subject to enabling and

disabling conditions. In its subtler (adult) form, the idea is this: if someone doesn't know that action  $x$  causes event  $y$ , it's unsurprising that she won't perform  $x$ , despite having reason to cause  $y$ —for she lacks a *subjective* reason to perform  $x$ . Correlatively, her performing  $x$  can be intelligible in terms of her knowledge that  $x$  causes  $y$ . Note that explanations in terms of knowledge are simultaneously teleological (they accord an explanatory role to reason-giving facts, knowledge being a *factive* state) and psychological (knowledge being a psychological state).

Second, mature interpreters are also able to explain actions in terms of considerations the agent takes to provide her with a reason, without endorsing the consideration or the claim that they constitute reasons. It is this dimension of the "adult theory" that enables us to figure out why someone is adding petrol to her tonic or to understand competition. And it is here that we find the rationale for BD: adopting a detached, relatively noncommittal stance toward others' intentional activities requires finding them intelligible in terms of reasons provided by their propositional attitudes.

The teleological genealogy of the adult theory has important implications for the sorts of psychological properties invoked in the mature conception of intentional action. It should lead us to question the relentless focus on just two kinds of mental states, beliefs and desires, encouraged by the "belief-desire model" of action-explanation. The teleological genealogy highlights the explanatory role given in the adult theory to *knowledge*.<sup>19</sup> It also suggests that the notion of desire in BD should be interpreted not in the narrow Humean sense but along the lines of Davidson's "pro-attitudes,"<sup>20</sup> as subsuming the immense variety of attitudes that constitute agents' perspectives on the purposes for which they act. Reflection on the teleological origin of commonsense psychology also helps to shed light on the sort of explanation required for understanding intentional action. It is not enough to think of certain mental states as the causes of bodily movements. What matters is the ability to see how some of the agent's psychological properties provide her with considerations that from her point of view can be seen to amount to a practical reason. Understanding the subjective reason informing someone's intentional action requires delineating what, from her perspective, presents itself as an objective reason.

#### Appendix 1: Direct and Indirect Tests

The recent investigations of children's understanding of rational action and belief have made it necessary to distinguish two sources of evidence. The

"traditional" evidence is based on what is called in the consciousness literature (Reingold and Merikle 1993; Richardson-Klavehn and Bjork 1988) a *direct test of knowledge*: Children are asked a question (or set a task) that refers (directly) to an actor's action in question. For instance, in the false-belief task children are told a story or made to observe a scene in which a protagonist puts an object in one of two locations, which is unexpectedly transferred to the other location in the protagonist's absence. When he returns children are asked: "Where does the protagonist think the object is?" or "Where will the protagonist go to get the object (he desires)?" On these measures a large amount of data shows that children start to give correct answers between the ages of 3 to 5 years. (See the meta-analysis by Wellman, Cross, and Watson 2001.) The younger children's difficulty stands in stark contrast to their ability to predict with conviction where someone will go for the object when the protagonist has witnessed the transfer and no false belief is involved (e.g., Clements and Perner 1994; Ruffman et al. 2001). Indeed, children not only fail to take belief into account in their predictions of the protagonist's actions but also fail to use it in explanations of the erroneous actions. For instance, when shown that the protagonist in the false-belief story goes to the empty location and they are asked why he went there, many of them answer (not incorrectly but uninformatively) "because he wants to get the object." When asked to specify why he then went to the empty location, they admit their ignorance: "Don't know" (Wimmer and Mayringer 1998; Perner, Lang, and Kloo 2002).

The fact that children feel comfortable talking about desires (want) before they address beliefs is also apparent in Bartsch and Wellman's (1995) study of naturally occurring speech. The authors tried, as best as they could, to identify genuine references to beliefs and desires from the limited context of the transcribed utterances (their figure 5.1). Reference to desire started with the first words as early as 18 months, peaking just before 3 years (33–36 months) to level off at about 3 percent of all utterances. In contrast, references to beliefs started at 3 years, reaching a constant level of about 2 percent of all utterances at 4 years. A similar picture also emerged from induced utterances in a study by Imbens-Bailey, Prost, and Fabricius (1997), conveniently displayed in our figure 14.1. Also, by using a method for eliciting extensive explanations, Bartsch, Campbell, and Troseth (2007) were able to get half the 3- to 4-year-olds to give belief-explanations of erroneous actions, but emotion- and desire-explanations were still much more frequent. However, when relying on nonverbal indicators of understanding belief and desire, recent research has complicated this seemingly clear picture.

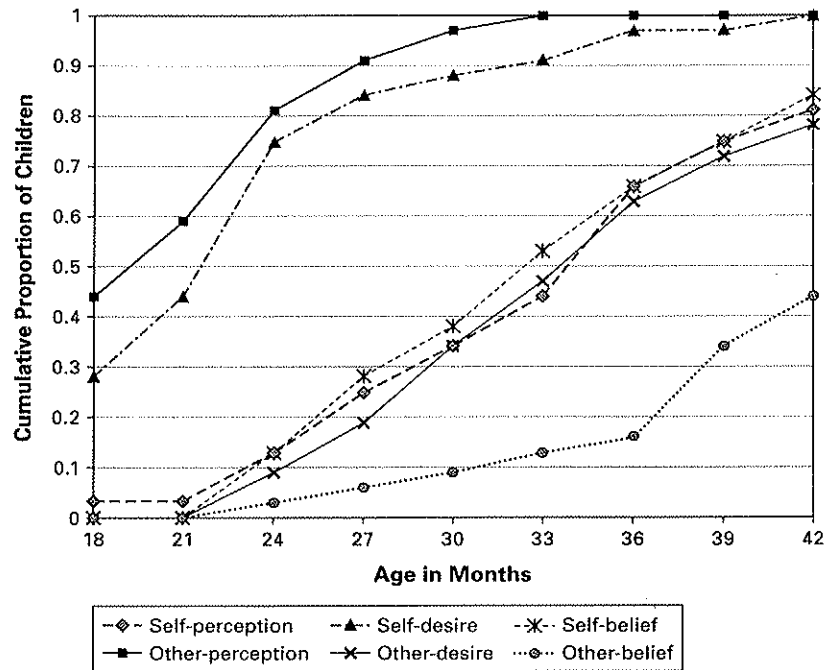


Figure 14.1

Proportion of children who use different categories of mental verbs.

Starting with Clements and Perner (1994), research has recently focused on using *indirect tests of knowledge*: Children are not asked any question about the protagonist's belief or action, but one of three other measures is used. (1) Looking time: The duration of looking at an erroneous action is compared to looking at a successful action. Longer looking at the successful action than at the erroneous action in cases of false belief is interpreted as children being surprised about a successful action when the actor has a false belief. These data indicate sensitivity to the protagonist's belief as early as 14 or 15 months (e.g., Onishi and Baillargeon 2005; Surian, Caldi, and Sperber 2007). One problem with these data is that the looking-time differences are multiply interpretable and not a clear indicator of expectation (Perner and Ruffman 2005; Sirois and Jackson 2007). (2) Looking in expectation: A better measure in this respect is children's direction of eye gaze as an indication of where they expect the protagonist to reappear (Clements and Perner 1994). This method has recently indicated

that children understand belief-based action as early as 2 years (Southgate, Senju, and Csibra 2007) or even earlier (Southgate 2008). (3) Interpretation of referential gestures: Southgate (2008) used a paradigm developed by Carpenter, Call, and Tomasello (2002) and Happé and Loth (2002) on children as young as 17 months. These infants observed a confederate putting object *A* into container *X* and object *B* into container *Y*. In her absence the infants watched the experimenter switch the objects (*A* now in *Y* and *B* in *X*). When the confederate returned she pointed to, say, box *X* and said, "I want this one." Then the covers of the boxes on the child's side were opened so that the child, but not the confederate, could see inside and the confederate asked the child to give the object to her. Children above chance handed her the object in box *Y*, which contained the object that the confederate had put inside box *X* but was no longer in there when the confederate expressed her desire for the object in box *X*.

We concentrate in our discussion on the evidence from direct tests. Answers to questions more clearly express a reasoned commitment, that is, an understanding that the protagonist will act the indicated way because he has good reason to do so. Indirect measures contain no commitment. I can look to a location because of a vague, unreasoned hunch even when to my proper reasoning he will appear somewhere else. Indeed, that is what the looking-in-expectation data seem to indicate: Children look (in expectation) to where the protagonist thinks the object is, but steadfastly and with conviction (Ruffman et al. 2001) they claim that he will appear at the location where the object actually is. Perner, Rendl, and Garnham (2007) proposed that the false-belief data from infancy (see also, e.g., Moll and Tomasello 2007; Tomasello and Haberl 2003; Southgate, Senju, and Csibra 2007) show an unexpectedly strong preoccupation with what other people can and cannot see, and keep an "experiential record" for the other person, using this record to compute likely actions and referential gestures but without clear understanding that these records constitute beliefs that direct action in the real world.

Investigations of children's understanding of knowledge presents a similar picture. In line with the evidence of early sensitivity to beliefs, children as young as 14 or 18 months keep track of what another person has or has not yet encountered (Moll and Tomasello 2007; Tomasello and Haberl 2003; Southgate, Senju, and Csibra 2007) and make their communicative behavior dependent on it by preferring to show the person objects that the person has not yet encountered over encountered objects (regardless of the child's own familiarity with the object). In all these cases, the use of this experiential record (Perner, Rendl, and Garnham 2007;

"registration of encounters," Apperly and Butterfill 2008) is limited to a single purpose, namely, to bring the other person into informational contact with yet unencountered objects. In the only other experimental paradigm used, children are asked to decide whom to ask for information, the one who has seen the hiding of the object or the one who was not able to see it. Here children seem oblivious to the other person's informational access until 4 years (Povinelli and deBlois 1992), or, with improved methodology, not before 3 years (Sodian, Thoermer, and Dietrich 2006).

## Appendix 2: Action versus Emotion

As we mentioned in the text, there is more than one way to interpret the claim that children understand the "subjectivity of desires." One reading is: they understand subjective preferences. A second reading is: they understand that a desire provides an agent with a subjective reason, a reason that makes it rational for the agent (but for no one else—unless they wish to cooperate) to act in a certain way. We argued that evidence regarding subjective preference provides no support for the claim that children understand subjective reasons. Some of the experimental work in this area, however, is concerned with a third claim. This is the claim that young children are able to understand and predict someone's emotional reaction to the satisfaction or frustration of some desire even when they don't *share* the desire. For example, Yuill (1984) addressed the question of whether children can predict that an agent will take pleasure in the satisfaction of a *wicked* desire (such as a desire to hit someone). Similarly, Perner, Zauner, and Sprung (2005) investigated children's ability to attribute emotional reaction to the satisfaction/frustration of desires in the case of two protagonists with mutually *incompatible* desires.

Does evidence from these paradigms help to settle the issue between teleology and the hybrid view? There are two reasons for skepticism. One is simply that the experimental state of play regarding the third claim is currently inconclusive. Crudely: Yuill's (1984) and Perner, Zauner, and Sprung's (2005) findings suggest that until they pass the false-belief task, children have great difficulty attributing emotional reactions to the satisfaction of goals they don't take to be objectively desirable. But more recently, Rakoczy, Warneken, and Tomasello (2007) and Rakoczy, Warneken, and Tomasello (2008) reported some evidence that children can attribute appropriate emotions in competitive situations before they pass the false-belief test.

A second, weightier reason for skepticism is this. It is not clear that explaining emotional reactions to the satisfaction/frustration of desires is a case of reason-giving explanation. It is not clear, for example, that to understand and predict that *A* will feel sad about the frustration of his desire, children have to appreciate that *A* has a *reason for* feeling sad. Emotional reactions *happen* to us: they are caused, for example, by perceived events, not formed for reasons. Correlatively, understanding such reactions is not a matter of reconstructing the agent's reason for undergoing it. To understand *A*'s sadness you merely have to know that the frustration of a desire tends to lead to sadness, not that it makes it *rational* to feel sad. Admittedly, this crude contrast doesn't do justice to the subtleties of adult commonsense psychology in this area. (See Goldie 2000, ch. 2, for some helpful distinctions.) But it is enough to suggest that it would be a non sequitur to move from evidence (if there were such evidence) that children are able to understand emotional reactions to desires they don't share to the conclusion that they grasp subjective reasons—that they think of *A*'s desires as making certain actions (or emotions) rational *from A's perspective*.

## Appendix 3: Sabotage

The empirically pressing question now is how to derive testable predictions from the assumption that young children use teleology before they are able to understand perspective differences. We need a task for which simple teleology is insufficient, i.e. which requires understanding the protagonists' intentional actions in terms of their perspectives on what's desirable. We need another, control task, identical to the experimental task except that teleology can provide the correct answer without need for understanding perspective differences. The predictions then are that the experimental task should be mastered after the control task and not before (within a margin of error) other typical perspective tasks (involving perspective differences, e.g., false-belief task) are mastered. Here is a suggestion involving sabotage.

Children can engage in sabotage before understanding false belief (Sodian 1991)—in seeming contradiction to our claim. In Sodian's tasks the children were asked to think of a way to prevent a robber from getting to the treasure. They correctly locked the treasure box when the robber approached but not when the good prince came to look for treasure. Critically for us, correct responding may not require children to *interpret* the

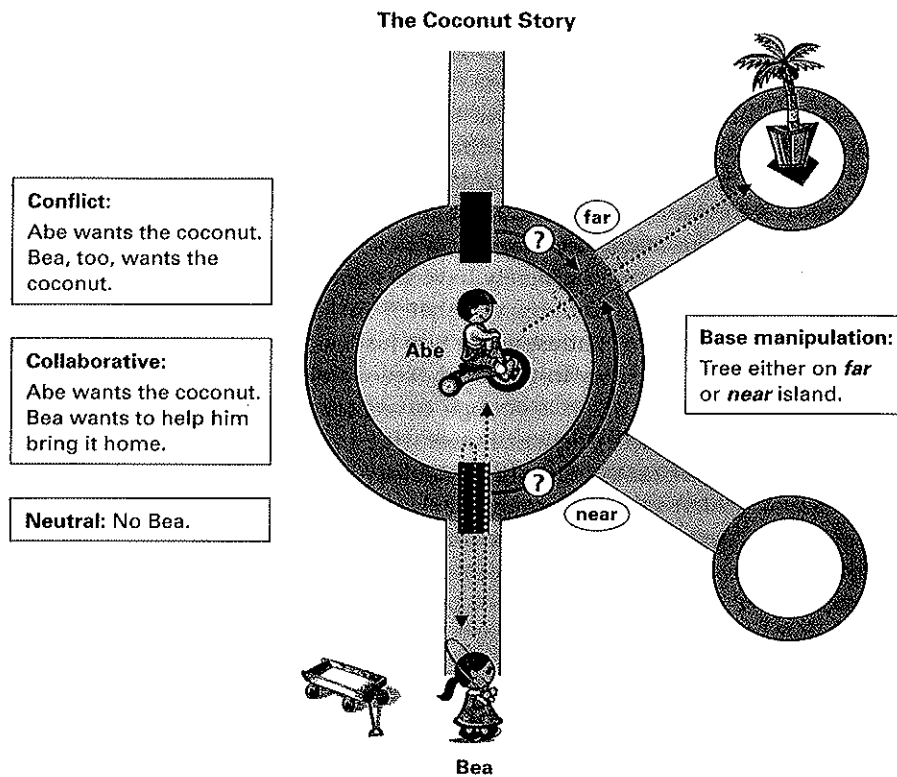


Figure 14.2

Test question: Which board will Abe use to get to the coconut?

robber's behavior—to find it intelligible in terms of his reasons. The child's task is only to find a means to prevent something bad. A new paradigm must ensure that prediction of the saboteur's act must be based on understanding what the sabotee will do for (his) good reasons.

In the experimental conflict story of the coconut scenario in figure 14.2 the girl, Bea, wants to get the one coconut left on the tree on the island. She crosses over the black board to the center island. Then she remembers that she needs her cart. She goes back to get the cart. Meanwhile, the boy, Abe, has arrived also with the intention to get the last coconut. To get to the tree he needs another board. He has the choice between the black and the brown board. Test question: Which board will he take to cross the moat, the brown or the black one? Correct answer: the black board.

In a collaborative condition, Bea indicates that she wants to help Abe pick the coconut. The correct answer to the test question in this condition is: the brown board. In a neutral condition, without Bea, we can get an estimate of children's preference for the "near" board (brown) over the "far" board (black). Presumably most children will make him use the nearer one. By counterbalancing the competitive and collaborative conditions with whether the palm tree is on the *near* (to Bea's approach) or the *far* island, we can check whether the understanding of the collaborative and the competitive aspects, respectively, can overrule the natural tendency to predict use of the most convenient nearer board.

We expect young children who fail false-belief tasks (as a measure of their ability to understand conflicting perspectives) to be able to predict correctly what Abe will do in the cooperative condition (use the brown board), but to fail to see that he will act differently in the competitive condition (he will use the black board). Children who pass the false-belief test, we expect, will make correct predictions in both conditions.

### Notes

1. Or rather Johann Nestroy, as quoted in the motto of Wittgenstein (1958).
2. It is tempting to think the solution to our puzzle may be provided by some recent experimental findings, widely taken to show that under certain conditions even very young children have some grasp of the causal role of false beliefs. In Appendix 1 we explain why we think the temptation should be resisted.
3. Compare Gopnik and Meltzoff's claim that 18-month-olds who participated in a study of imitation of failed attempts were able to treat different kinds of movements "as the causal consequence of the same underlying mental state, the same goal" (1997, 151).
4. See Schueler 2009 for illuminating discussion of the importance, and the theoretical implications, of the "putting-together" point.
5. See e.g., Parfit 1997; Scanlon 1998; Schueler 2003, 2009; Hornsby 2008.
6. "[T]ypically, in deliberation what I do pay attention to are the relevant *features* of the external world [rather than one's own desires or impulses]: the cost of the alternatives, the quality of the food, the durability of the cloth, the fact that I made a promise" (Blackburn 1998, 253).
7. Crudely, claims of need have implications to do with matters such as harm and flourishing, but they have no immediate implications regarding the agent's desires. (That you need some vitally important medicine does not imply that you want to take it.)

8. Are fully objective reasons *external* reasons in Williams's (1981a) sense—reasons that cannot be derived by practical deliberation from the agent's "subjective motivational set"? The question raises complex interpretative and substantive issues we cannot pursue here. Two brief comments will have to suffice.

One comment is that objective reasons in our sense are not necessarily external in Williams's. At least in some of his later writings on external reasons, Williams (1995b) allows that the application of "thick ethical concepts" can provide us with practical reasons. Of course, Williams insists that whether a certain application of a "thick" concept provides an agent A with a reason essentially depends on whether A uses the concept ("*chastity* is an example that focuses the mind"; *ibid.*, 37). The idea here is that the reason-giving role of "thick" facts turns on the presence in A's subjective motivational set of the sorts of dispositions tied up with mastery of the relevant thick concept. Still, it is significant that this version of internalism is not committed to the view that only certain kinds of mental states (passions or desires) can constitute the noninstrumental element of practical reasons. Suppose someone is tempted to make a remark that, unbeknownst to him, would be a very tactless remark to make in the circumstances. Williams's "liberal" internalism allows us to say that—provided only that *tactless* is one of the person's concepts—he has a reason not to make the remark, even though from his current perspective the remark seems unobjectionable. In our terminology, this would be an objective reason; but the reason would still count as internal by Williams's lights.

Our second comment is that nevertheless, the developmental suggestion we will be pursuing may be put by saying that we all started life as external reasons theorists. For the suggestion is that young children are familiar with objective reasons before even *understanding* subjective reasons—before they grasp that there are two sorts of perspectives from which to consider what someone has reason to do. It is not clear, though, that there is anything in Williams's account that would be inconsistent with this suggestion. (The "liberal" internalism Williams espouses in this later writings does not imply that external reasons statements are *meaningless*; it merely says they are false.) In fact our developmental account may shed some light on what is going on when people make what appear to be external reasons statements. It is significant that Williams's most convincing example of an external reasons statement (his gloss on James's story of Owen Wingrave) turns on a conflict among members of a family. (Despite Owen's loathing for the military, "[h]is family might have expressed themselves by saying that *there was a reason for Owen to join the army*"; Williams 1981, 106.) Our developmental suggestion would make it unsurprising if the family were a context in which the idea of external reasons could sometimes be seen to linger.

9. Although this is not the way Bartsch and Wellman put things, some of their suggestions fit well with the hybrid account; they stress the importance of desire psychologists' drawing on their *own* knowledge of the world in predicting how someone will go about satisfying his or her desires (see Bartsch and Wellman 1995, 155).

10. We use the term "pro-attitude" in Davidson's sense, as subsuming "desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values" (Davidson 1963/1980, 4).

11. Compare Aristotle's conception of teleological explanation: "the fourth cause is the goal: i.e., the good" (quoted in Charles 1984, 198). More recently, Csibra and Gergely (1998) have used the notion of a "teleological stance" in their theory of infants' expectations concerning rational actions. Their account differs from ours (and Aristotle's) in that it does not appeal to evaluative facts as the source of goals, but takes the notion of a goal as primitive. Another difference is that their account is concerned with children's *expectations* as to the movement of certain kinds of agents under certain circumstances, not with children's ability to *explain* intentional actions in terms of agents' reasons.

12. Of course, there is a (to adults, natural) way of taking these latter propositions that would make them conflicting without being inconsistent, along the lines of "there is something to be said for A's winning, and there is something to be said for B's winning." But construed in this way, the propositions could not do the work a teleologist expects them to do. To provide teleological explanations, the relevant evaluative propositions have to license conclusions as to what the agent should do or has most reason to do; *prima facie* reasons, on their own, do not license such conclusions. That there is something to be said for A's winning, and that A can win by doing x, only gives A a reason to do x if there are no other more important considerations in favor, e.g., of letting B win. In brief, teleology, as we conceive it here, has to appeal to "all-out" practical reasons.

13. See Woodward forthcoming for a suggestive discussion of the contrast between "difference-making" and "causal process" theories of causation. See also Steward 1997 and Hornsby 1993.

14. See Woodward 2003; see also Campbell 2007 for helpful discussion.

15. The case of imitation illustrates that teleology has the resources to conceive of reasons as "agent-specific." Although teleology appeals to worldly facts rather than the individual agent's mental states, this does not mean that two agents cannot have reason to perform different actions. One way in which different agents may be seen to have reasons to do different things is in virtue of their distinctive roles. Having just banged my toy a couple of times, I, currently performing the part of the model, have no reason to keep banging; but you, performing the part of the imitator, have every reason to bang your toy. Again, the very same objective purpose or value can be seen to yield different sorts of practical reasons, depending on the agent's skills and circumstances.

16. Of course, it is not obvious whether imitation is, or when it begins to be, tied up with explanation. Certainly to begin with, infant imitation may well be a more



primitive phenomenon than the ability to make rational sense of others' behavior. All we are saying here is that imitation provides compelling materials for teleological explanations. Acquiring a conception of intentional action may partly be a matter of learning to exploit such materials.

17. See Doherty 2009 for helpful discussion of the notion of engagement.

18. For a "grown-up" version of this point, compare John McDowell: "Finding an action or propositional attitude intelligible, after initial difficulty, may not only involve managing to articulate for oneself some hitherto merely implicit aspect of one's conception of rationality, but actually involve becoming convinced that one's conception of rationality needed correcting, so as to make room for this novel way of being intelligible" (McDowell 1998, 332).

19. See Williamson 2000, and especially Hornsby 2008, on the explanatory role of knowledge.

20. Davidson's gloss on "pro-attitude" bears repeating: "desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values" (1963/1980, 4).