Applying the Causal Theory of Reference to Intentional Concepts

John Michael and Miles MacLeod*†

We argue that many recent philosophical discussions about the reference of everyday concepts of intentional states have implicitly been predicated on descriptive theories of reference. To rectify this, we attempt to demonstrate how a causal theory can be applied to intentional concepts. Specifically, we argue that some phenomena in early social development (e.g., mimicry, gaze following, and emotional contagion) can serve as reference fixers that enable children to track others' intentional states and, thus, to refer to those states. This allows intentional concepts to be anchored to their referents, even if folk psychological descriptions turn out to be false.

1. Introduction. It is a curious fact that many philosophical discussions about the reference of everyday concepts of intentional states—from folk psychology in the 1980s to theory of mind and mind reading more recently—have, in recent decades, been predicated on descriptive theories of reference. A particularly salient illustration of this can be gleaned from the debate about eliminative materialism (Churchland 1981; Stich 1983; Dennett 1987; Fodor 1987; Lycan 1988; Bermudez 2003), as we demonstrate.[1] But, given that casual theories have become increasingly widespread in philosophy of lan-

Received June 2012; revised January 2013.

*To contact the authors, please write to: John Michael, Center for Subjectivity Research, Copenhagen University; e-mail: johnmichaelaarhus@gmail.com. Miles MacLeod, Georgia Tech; e-mail: milesfromanywhere@gmail.com.

†We would like to thank Ruth Millikan, Brian Epstein, Dan Dennett, Edouard Machery, Joe McCaffrey, two anonymous reviewers, and also audiences at the University of Pittsburgh and the University of Vienna for very helpful comments and suggestions on earlier versions of this article. Miles MacLeod's participation was supported by the Konrad Lorenz Institute for Evolution and Cognition Research, Austria, and the US National Science Foundation (DRL097394084).

1. We believe that descriptivism has also shaped research on concepts of intentional states in psychology, and we will demonstrate (secs. 6 and 7) how the application of a

Philosophy of Science, 80 (April 2013) pp. 212–230. 0031-8248/2013/8002-0005$10.00
Copyright 2013 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

guage and philosophy of science (i.e., since Kripke 1972; Putnam 1975) as well as in philosophy of mind (i.e., Fodor 1990; Laurence and Margolis 2002; Prinz 2002) over the past 30 years, this must be regarded as anachronous. We therefore regard the neglect of causal theories of reference as a significant gap in research on intentional concepts.

In this article, we aim to fill in this gap by demonstrating how a causal theory can be applied to intentional concepts. Specifically, we will (*a*) formulate conditions under which bodily or cognitive processes can function to fix reference to others' intentional states, (*b*) consider some candidate processes that may fulfill those conditions, and (*c*) articulate the roles that these reference-fixing relations might play in the development of intentional concepts.

With respect to *a* and *b*, our main claim is that some processes that are engaged in social interactions—such as gaze following, imitation, and emotion contagion—may stand in reliable causal relations with others' intentional states, such as their attention, their intentions, and their emotions, and thereby fix reference to those states. These reference-fixing processes are akin to devices used by scientists to interact with and thereby to maintain a causal link to phenomena that are being investigated. Importantly, such devices can be used and can serve their reference-fixing function in a relatively theory-free manner.

With respect to *c*, we propose that intentional concepts are likely to have been shaped evolutionarily and to be shaped ontogenetically by the need to integrate, modulate, and build on the bodily responses that are engaged during social interaction. As a result, the development and refinement of intentional concepts enables children to become increasingly proficient at using the information that their own bodily responses provide about others' intentional states in order to track ever more fine-grained or sophisticated intentional states. Thus, a causal theory of reference enables us to capture a sense in which reference to others' intentional states is fixed very early in childhood and structures the subsequent development and progressive refinement of the descriptive contents of intentional concepts. Hence, we claim that the function of the descriptive contents of intentional concepts is not to establish reference but to elaborate and make use of it.

In a first step (sec. 2), we demonstrate how discussions of intentional concepts in philosophy and psychology have been shaped by a descriptive theory of reference. This is most striking in the case of the eliminativism debate, which—we argue—presupposes a descriptive theory. Second (sec. 3), we draw a contrast with causal theories, which allow for the reference of

causal theory can provide a framework for modeling the development of such concepts. See Millikan (1998) for an argument to the effect that psychological research on concepts in general has been shaped by descriptivism.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

intentional concepts independently of the particular theories or descriptions associated with these concepts. The rest of the article is then devoted to considering how to apply the causal theory to intentional concepts, that is, to addressing issues *a*–*c*.

2. Eliminativism and Descriptive Theories of Reference. Eliminativists have argued that intentional terms like 'belief' and 'desire' should be and are likely to be abandoned as scientific psychology reaches a mature stage (Churchland 1981; Stich 1983; see also Dennett 1987; Fodor 1987; Lycan 1988; Cummins 1991; Bermudez 2003). The argument is that intentional concepts are defined by a folk theory of mind that is likely to be false, and they are therefore likely to be nonreferring. But, as Mallon et al. (2009) have noted (see also Lycan 1988; Cummins 1991; Schouten and De Jong 1998), this argument only works in conjunction with the implicit claim that reference is fixed by descriptions.[2] In other words, eliminativists assume a descriptive theory of reference (also referred to as a description theory of reference or descriptivism) for intentional concepts like 'belief' and 'desire'. While descriptive theories may take a variety of forms, they typically agree on the following points:

D1. Competent speakers associate a description with a term *t*.[3] This description, also called the 'meaning of *t*', specifies a set of properties and relations.

D2. An object is the referent of *t* if and only if it uniquely or best satisfies the description associated with it.

Descriptive theories, the origin of which is often imputed to Frege (1892) or Russell (1905), have played a significant role in the development of philosophy of science over the past 50 years. They are central to the incommensurability claims of Feyerabend (1962) and Kuhn (1970). Indeed, Kuhn and Feyerabend assume a strong and, above all, holistic reading of descriptive theories since they maintain that the meaning of a term is dependent on the entire theoretical structure in which it occurs. As a result, theory change

2. Whereas Lycan (1988) replaces descriptivism with a causal theory and argues on that basis that intentional concepts are in fact referring, Schouten and De Jong (1998) maintain that neither a causal nor a descriptive theory yields an adequate account of the reference of intentional concepts, and Mallon et al. (2009) present evidence that intuitions about reference vary cross-culturally and argue on that basis that theories of reference are a poor basis for philosophical theorizing.

3. To avoid confusion, it is important to note that theories of reference in philosophy of language and philosophy of science are typically concerned with the reference of terms. In contrast to this, our aim in secs. 4–8 will be to apply a causal theory to intentional concepts.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

leads to meaning change and, thus, also to the failure of the terms of the old theory to refer. In a similar manner, eliminativists have used the descriptive theory of meaning and theory dependence of meaning to argue that belief/desire psychology is likely to turn out to be incommensurable with neuroscience. The putative truth of the latter and the eventual replacement of belief/desire psychology by a neuroscientific theory signals the referential failure of folk psychology.

Descriptive theories of reference, however, are not the only game in town. Indeed, over the past 30 years, reference has become a hotly contested issue in the philosophy of language and philosophy of science. This discussion, sparked somewhat by the negative conclusions that follow from incommensurability, has generated a series of alternative theories of reference, beginning with the causal theories of Kripke (1972) and Putnam (1975), to which we now turn.

3. Causal Theories of Reference. Causal theories aim to account for the fact that scientists (and individuals in everyday life) often make referential connections despite differences in the descriptions they and others attach to terms. In other words, adapting a causal theory to intentional concepts would support the independence of the reference of intentional concepts from particular theories of mind. Causal theories applied to scientific contexts (i.e., for natural kinds and entities) take the following form (see Putnam 1975):

C1. The reference of a term *t* is fixed by applying the term ostensively to refer to the cause of a set of observed events.
C2. Users of the term refer successfully to the same referent of *t* by being linked to the original use through a historical causal connection.

'Ostension' here refers to some event (or process) by which a language user acquires a causal relation to a putative entity or kind in the world. This can be as simple as pointing to a phenomenon and assigning it a natural kind term (where the putative essence of the kind causally explains the properties of the phenomenon) or naming a putative unobservable causal agent that produces the phenomenon. Either way the dubber is causally connected with the causal agent or natural kind since that agent or its essence is causally involved in the dubbing event through its connection with the observed phenomena (the samples, etc.). These phenomena are themselves part of the causal chain (or causal conditions) that leads to the dubbing and, thus, ultimately so is the causal agent or kind essence.

In this respect, the stereotype or description associated with the term does not play a role in its reference but only in the causal relations between the essence or agent, the phenomenon, and the dubbing event. Further uses of

https://doi.org/10.1086/670328 Published online by Cambridge University Press

the term are thus historically connected to this dubbing event through interactions between language users who pass on the term. New users thus acquire the term causally through reading and conversation. In consequence, it is presumed that the reference of a term is taken to be given for a speaker (and not under revision) in normal use. Thus, when speakers make assertions using a term, they are generally presumed not to be modifying its reference but to be relying on these causal-historical relations. If this were not the case, causal theorists argue, conversation would be much more difficult.

Causal theories are thought particularly strong in the context of proper names (on which Kripke especially focused), for which it seems that speakers and listeners do tend to outsource the reference of their terms to historical usage and thus are perfectly capable of making false statements about specific things and people without it being presumed that they refer to something else. For example, when people talk about Napoleon, we presume they are referring to the deceased French emperor and not to someone else, even if they say false or ridiculous things about him.

However, in the case of theoretical entities and natural kind terms, the theory seems much more problematic as a general account of reference. For one thing, it seems to impute too much reference or 'too easy' reference. For example, on a causal theory of reference, 'phlogiston' and 'oxygen' could be considered co-referential terms, given that phlogiston and oxygen were dubbed as causal agents with respect to similar causal contexts (i.e., combustion phenomena). Since they stand in the same causal relations to their users, they appear to have the same referents—even though scientists would not generally agree that 'phlogiston' really referred to oxygen all along (Kitcher 1993). Plainly, it must be possible to speak of reference failure in cases such as this one, and yet causal theories seem not to offer the resources for doing so.

Moreover, there are also cases in which reference is ambiguous. According to modern chemical theories, for example, the concept of an acid picks out three different sets of substances: Arrhenius acids, Brønsted acids, and Lewis acids (e.g., Stanford and Kitcher 2000). To which, if any, of these types does the term 'acid' refer (as it was used in the past or today)? It does not seem apt to say that it always referred to one or the other alone on the basis of a particular causal relationship.

A more serious philosophical concern is the so-called qua problem (Devitt 1981). There seems to be no theory-free way to ostend a natural kind or particular causal agent, given that any element instantiates numerous natural kinds or any phenomenon is the result of many causal elements acting conjointly. A sample tiger, for instance, instantiates a species but also a genus, and so on, up the hierarchy of mammals, and so forth.

One way of addressing such concerns is to acknowledge that some level of description is unavoidable for fixing reference. This is the strategy en-

https://doi.org/10.1086/670328 Published online by Cambridge University Press

dorsed by *causal-descriptive theories* (see Enç 1976; Nola 1980; Kroon 1985, 2011; Psillos 1999). The core insight of such approaches is that even in the absence of a theory of some phenomenon, scientists structure their investigation of the phenomenon by postulating a causal agent and generating descriptions of the mechanism(s) by which the causal agent could bring about the phenomenon. On such a conception, continuity in the reference of a term across theory change does not come as easily as it does on straightforward causal accounts. Rather, it requires continuity at least in those descriptive components that scientists consider to be most crucial or most productive in making particular phenomena or relations among phenomena appear salient, in shaping the construction of theories and explanations, and in guiding the formulation of hypotheses.

Before moving on to consider how such a conception may be applied to intentional concepts, it is important to point out that (very prominent) causal theories have been developed in philosophy of mind and applied to concepts (although not specifically to intentional concepts). Broadly speaking, there are two different types, namely, nomological and teleosemantic theories, both of which abandon the historical construal of the causal relation that underpins reference (with a dubbing event fixing the reference, followed by a chain of events that transmit the reference). According to the most influential nomological theory, namely, Fodor's asymmetric dependence theory (1990), a concept C refers to R's by virtue of the fact that there is a nomic connection between C and the R's. Thus, for example, the concept of a tiger is reliably activated by tigers.[4] According to Millikan's (1984) teleosemantic theory, representations (such as concepts) refer to whatever they have the biological function of tracking.[5] This, too, is a causal theory insofar as the tiger concept is taken to exist because it has enabled the reidentification of tigers; that is, the fact that it refers to tigers is the cause of its current existence.

4. A further condition is that, although there may also be a nomic connection between some other concept and tigers, this other nomic connection must be asymmetrically dependent on the connection between the concept of a tiger and tigers. Thus, e.g., I may consistently fail to distinguish between dogs and wolves and, as a result, apply the concept of a dog both to dogs and to wolves, but I would not activate my dog concept when encountering a wolf if it were not the case that encountering a dog also caused me to activate my dog concept, whereas the reverse does not hold. The priority of the dog-dog relation thus explains why the concept of a dog refers to dogs and not to wolves (Fodor 1990).

5. Millikan speaks of "empirical concepts," which she understands as abilities to reidentify substances—substances being "things that retain their properties, hence potentials for use, over numerous encounters with them" (2000, 2). Crucially, for Millikan, the reidentification of substances does not need to involve the kinds of description traditionally associated with concepts. More on this in a moment.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

These theories appear at first blush to render the question of how reference is fixed in ontogeny superfluous: for Fodor, reference is fixed by nomic dependency irrespective of the origin of that dependency, whereas for Millikan evolutionary history may fix reference before individual learning history. This appearance is deceiving, however. Consider Fodor's theory first: Laurence and Margolis (2002) have argued persuasively that the only way to avoid radical concept nativism (and there is a broad consensus in favor of trying to doing so) is to provide an account of how conceptual primitives can be learned. They therefore propose an account of how this is possible, the core component of which is the so-called sustaining mechanism: "a mechanism in virtue of which a concept stands in the mind-world relation that a causal theory of content, like Fodor's, takes to be constitutive of content" (37). Importantly, sustaining mechanisms determine reference not by means of a definitional structure (which would require further concepts and thus not be primitive) but by latching on to syndromes of perceptible properties (e.g., shape, size, typical motions, markings) that are diagnostic of a kind. Such mechanisms, along with "a disposition to treat instances as members of the category only if they have the same essential property as paradigmatic exemplars of the syndrome" (38), could establish just the sort of nomic relation between concept and reference class that Fodor's theory requires.

In a similar vein, Prinz (2002, 2005) maintains that any causal theory must include an account of the perceptual vehicles that enable individuals to identify members of a concept's reference class, otherwise it would be unclear how people could ever use the concepts they have in order to identify the objects those concepts refer to. Indeed, Prinz appropriates Fodor's nomological theory for his own radical concept *empiricism*, arguing that once one has included such an account, it is superfluous to add amodal symbols to play the role of concepts. Instead, the perceptual vehicles (which he calls "proxytypes") themselves can be reused for inferential thinking and thus play the role of concepts.

As for Millikan (1984, 2000), her view is in fact that many or even most concepts (i.e., abilities to reidentify substances) originate in individual learning, with evolutionary history underpinning the more general abilities that enable individual learning (e.g., Millikan 2000, 50, 76–83).[6] Like Prinz as well as Laurence and Margolis, Millikan thinks that concept learning begins with abilities to track substances perceptually. She proposes that this is achieved by applying templates, which can be as minimal as sets of questions for which the children expect to be able to find answers and can enable children to track substances without associating any determinate properties at all with them (82–83). This is illustrated by a study by Xu and Carey (1996), in which 10-month-olds did not dishabituate when an object of one

6. On Millikan's definition of substances, see n. 5.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

kind turned into an object of another kind (e.g., when a yellow duck turned into a white ball) but did dishabituate when it turned into two objects. Such property-blind tracking makes it possible to acquire information about "what sorts of things tend to remain the same and what sorts may change within a short period, yielding clues for keeping conceptual track of substances" (Millikan 2000, 77). Thus, Millikan's teleosemantic theory, too, acknowledges the need for the reference of concepts to be fixed in ontogeny by tracking mechanisms (which, in her case, achieve reference fixation in virtue of the function that tracking mechanisms play in reidentifying and learning about their referents).

In what follows, we will not attempt to adjudicate between these different causal theories. What we will be doing is offering an account of the mechanisms that may serve to fix the reference of intentional concepts, irrespective of which version of the causal theory one endorses. We will also be suggesting that our account of the mechanisms that fix reference for intentional concepts provides a useful framework for modeling conceptual development in the domain of intentional concepts.

4. Conditions for Reference Fixing. In considering how a causal theory can be brought to bear on intentional concepts, our focus will be on the question of how reference is fixed, that is, how the basic mechanisms of reference fixing for intentional concepts can be articulated by analogy to reference fixing for theoretical entities in science. Our main claim is that some processes that are engaged in social interactions may stand in reliable causal relations with others' intentional states, such as their attention, their intentions, and their emotions, and thereby fix reference to those states. Thus, our proposal builds on the accounts of perceptual tracking processes as reference fixers found in Millikan (2000), Laurence and Margolis (2002), and Prinz (2002). However, it also differs from these other approaches insofar as the processes we will consider as candidates for reference fixers, such as gaze following, imitation, and emotion contagion, are not just perceptual processes but in fact bodily responses. In this respect, these processes are akin to devices used by scientists to maintain a causal link with phenomena that are being investigated. In science, reference fixing functions to structure ongoing research, making particular phenomena appear worthy of investigation, suggesting testable hypotheses, and also constraining the possibilities for empirical investigation. In order for candidate processes to fix the reference of intentional concepts in an analogous fashion, they would need to fulfill the following requirements:

(i) Causal interactions between tracking devices and intentional states must be present early in the development of those concepts. It must be emphasized that tracking devices may not initially qualify as rep-

https://doi.org/10.1086/670328 Published online by Cambridge University Press

resentations of intentional states, as they may not be differentially sensitive to intentional states as opposed to behavioral expressions or effects of those intentional states. In other words, a tracking device may initially detect a behavioral pattern caused by an intentional state rather than the intentional state itself. Nevertheless, it may have been evolutionarily selected to detect that behavioral pattern as a proxy of the intentional state and, thus, to gain traction on the intentional state via the proxy. This would enable the cognitive system subsequently to build on the basic causal relation in order to develop a capacity for representing the intentional state.

(ii) They must also remain in place for a sufficiently long period to be used to sustain interaction with intentional phenomena. Thus, a certain degree of continuity in the functional role of tracking processes is essential.

(iii) Children must treat them as sources of information about intentional states; that is, they must support a referential assumption on the part of the developing child.

5. Candidates for Reference Fixing. Let us now look a bit more closely at what kinds of process can fulfill these requirements and thus play the role of tracking devices and thereby fix reference.

5.1. Behavioral Mimicry. Although reports of neonate imitation remain controversial (Meltzoff and Moore 1977; Kugiumutzakis 1999; for a critical perspective, see Anisfeld 1991), it is not controversial that children engage in mimicry in the first year of life (Heyes 2001). Thus, mimicry satisfies i. Moreover, mimicry is also continuously present throughout development and into adulthood (Lang et al. 1993; Dimberg 1997; Doherty 2009) and, thus, satisfies ii.

What about iii? Behavioral mimicry, of course, does not presuppose representations of intentional states. Nevertheless, it does depend on representations of proxies of intentional states, such as the bodily kinematics that instantiate motor intentions (Pacherie 2005). Insofar as such proxies are a crucial source of information about others' intentional states, representing them provides infants and adults with an important foothold on others' intentional states. In fact, one recent study found that different higher-level intentions (e.g., grasping a cup in order to drink from it vs. grasping the cup in order to offer a drink to someone else) are manifested in different kinematics, which are perceptually distinguishable for observers (Becchio et al. 2012). Moreover, a tendency to engage in mimicry also scaffolds development of a nuanced understanding of the intentional states underlying behavior, by making infants attend to features of others' behavior that are relevant to understanding their intentional states.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

And, most important, there is evidence that imitative behavior in young children is sensitive not only to the movements they are observing but to the intentions they take to underlie those movements. Rational imitation, for example, in which infants imitate only those features of an action that are relevant to the goal, is present at around 14 months (Gergely, Bekkering, and Király 2002). Soon after this (18 months), infants become able to imitate uncompleted actions (Meltzoff 1995). Both of these phenomena provide elegant illustrations of the referential assumption demanded by condition iii; that is, what children hone in on is not behavior per se but an intention they assume as an unobservable causal agent underlying behavior.

5.2. Emotional Contagion. There are parallels to this in the domain of emotions. Consider emotional contagion. An early precursor of emotional contagion can be observed in the phenomenon of contagious crying just minutes after birth (Sagi and Hoffman 1976; Geangu et al. 2010). This fulfills i. Moreover, ii is also fulfilled insofar as emotional contagion in adults is a well-documented phenomenon (Hatfield, Cacioppo, and Rapson 1994). More nuanced forms of emotional contagion that occur during the first year of life support conceptual development by enabling children to assume an emotion as a causal agent underlying an observed emotional expression (fulfilling iii). At around 6 months, for example, the phenomenon of affect attunement reveals children in dynamic two-way interactions with caregivers, in which the two participants successively exchange various expressions of the same emotion (Stern 1985). As Stern (1985) has pointed out, these exchanges reflect an incipient abstract understanding of the emotional states underlying emotional expressions, for they require the child to treat distinct expressions (even in distinct modalities) as expressing the same underlying emotional state. Thus, we see here the assumption of an unobservable emotional state enabling children to see a coherent pattern across several expressions, that is, to conceptualize the emotional state behind the expression. In social-referencing situations, which occur by around 9 months, children are sensitive to the objects of caregivers' emotional expressions and adopt those attitudes toward the same objects (Baldwin and Moses 1994), that is, treating as dangerous or disgusting objects toward which the caregiver has expressed fear or disgust. Here we see how the emotional response can be combined with an increasingly sophisticated understanding of the adult's visual attention, in order to track not only the type of emotional state but also its intentional object.

5.3. Gaze Following. The same progression can be observed with respect to gaze following. Gaze following occurs by 6 months at the latest (Senju and Csibra 2008), and Hood, Willen, and Driver (1998) have even found evidence for gaze following in 2.5-month-olds—thereby fulfilling i.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

Condition ii is also fulfilled insofar as gaze following occurs consistently throughout cognitive development. And this is no surprise, given that following an interaction partner's gaze makes it much easier to keep track of what information she has acquired or is seeking, to anticipate her actions, and also to interpret her utterances. This makes it highly useful for children in learning from adults.

Admittedly, it is not possible to rule out a lean interpretation of early gaze following, according to which infants just reflexively look in the direction in which adults are looking without linking the adults to the object on which their gaze settles, that is, understanding that this is the object of the adults' attention. Nevertheless, an innate disposition to detect eyes and to respond to them in a way that reflects sensitivity to their role as sources of perceptual information would direct infants' attention toward a fundamental source of information about others' emotions and intentions (see condition iii). And indeed, infants are born with a predisposition toward face-related stimuli: they are able to discriminate faces from other stimuli (Mondloch et al. 1999), preferentially track moving face stimuli (Johnson et al. 1991), and within days of birth discriminate between the faces of different people (Walton, Bower, and Bower 1992; Bushnell 2001). Eye contact activates specific prefrontal regions, even in 5-month-old infants (Grossman, Parise, and Frederici 2010). Farroni and colleagues (2002) have even found that 2–5-day-old newborns looked longer and more frequently at a photograph of a face whose eyes were facing directly to them than a different image of the same face with the eyes facing away.

Moreover, by 8 months, there is evidence that children will selectively follow an adult's gaze when the adult is performing an action with an ambiguous intention (such as playing with an object and then abruptly cupping her hands over it), apparently attempting to draw on information about the adult's visual attention in order to interpret her intention (Carey 2009, 181–82). By 9 months, children hold up objects for others to look at and actively try to draw others' attention to objects, checking back and forth between the other agent's eyes and the object in order to monitor whether they are attending to it (Tomasello et al. 2005).

6. Reference Fixing and Conceptual Development. In the previous section we reviewed evidence bearing on the suggestion that behavioral mimicry, emotion contagion, and gaze following could fulfill conditions i–iii and, thus, function as reference fixers for intentional concepts. We would now like to look more closely at the role that reference fixing plays in conceptual development. To begin with, consider the analogy to scientific research: scientists do not usually use tracking devices merely to observe a phenomenon. Rather, they actively set up conditions for this observation with the help of experimental equipment and tracking systems, and they continuously

https://doi.org/10.1086/670328 Published online by Cambridge University Press

refine their equipment and observational practices in light of the development of models and theories. This makes it possible to track a target phenomenon with increasing sophistication, that is, to seek it out actively, to predict its appearance in novel circumstances, to recognize seemingly distinct phenomena as instances of it, and so on. Likewise, children actively contribute to sustaining and refining the causal relationships between their bodily tracking devices and others' intentional states. This makes it possible for intentional concepts to further articulate the links to causes of behavior that have been fixed by tracking systems and, thus, to track ever more sophisticated and fine-grained intentional states.

As a result, if a tracking process is being used to sustain reference to a phenomenon in order to enable conceptual development, then we should expect that conceptual development will influence its deployment. Indeed, it is likely that conceptual development will be shaped by the imperative to integrate low-level tracking processes with other processes, to interpret them as sources of information about others' intentional states, and to modulate them in context-sensitive ways. Let us briefly consider some evidence that this is the case for the candidate reference fixers we have been discussing.

Regarding behavioral mimicry, older children and adults not only continue to mimic others but do so with an expanded behavioral repertoire, involving not only the face but also postures, mannerisms, and bodily configurations (Hatfield et al. 1994; Chartrand and Bargh 1999). Additionally—and for our purposes quite importantly—there is some research concerning the situation dependence of mimicry in adults. Several studies have found that mimicry is modulated by various factors, such as group membership and social context (Hess and Bourgeois 2010). Moreover, it has been shown that we are more likely to imitate people with high status—an efficacious strategy, given that imitation tends to increase rapport (Cheng and Chartrand 2003; Wang and Hamilton 2012). There are also reasons to think that mimicry is not only modulated but sometimes actively suppressed by top-down control (Brass, Derrfuss, and von Cramon 2005). Again, mimicry increases rapport. However, there are surely occasions when rapport gets in the way of performing a task, namely, when it is advantageous for individuals to suppress mimicry and thereby prevent the cultivation of rapport. For example, this could be the case when one has the task of detecting liars or cheaters; cultivating rapport may prevent one from suspecting someone of being a liar or a cheater (see Stel, van Dijk, and Olivier 2009). And there is some empirical support for the notion that people indeed suppress mimicry in a way that is sensitive to such considerations. Lanzetta and Englis (1989) found mimicry in a cooperative context (e.g., teammates in a game) but countermimicry in a competitive context (e.g., opponents in a game; see also Hess 1998).

https://doi.org/10.1086/670328 Published online by Cambridge University Press

Similarly, emotional contagion is deployed in an increasingly nuanced manner in social interactions as development proceeds. At around 18 months, when children are able to make a distinction between self and other, as evidenced by their ability to recognize themselves in a mirror, they also begin to react with empathic and sympathetic responses to victims of distress and with appropriate, other-directed comforting and prosocial behavior, thus attesting to an understanding of the other's emotion as the source of one's own response (Bischof-Köhler 1991; Zahn-Waxler et al. 1992; Eisenberg and Fabes 1998). Moreover, there is evidence that emotional sensitivity to others' emotions not only informs an increasing array of other processes but is itself increasingly modulated by other processes. For example, there is evidence that empathetic pain responses, as measured by activation in anterior cingulate cortex, are modulated by numerous contextual factors such as whether one believes the person experiencing the pain deserves it (De Vignemont and Singer 2006).

As in the cases of mimicry and emotional contagion, we can observe a progressive development of a more nuanced deployment of gaze following, which increasingly integrates conceptual knowledge and situation-dependent information. For example, 1-year-olds are able to gaze follow to locations behind barriers (Moll and Tomasello 2004). And 2-year-olds are able to correctly anticipate an agent's gaze direction even in false-belief scenarios (Southgate et al. 2007), which require them to keep track of the agent's preference for a particular object as well as her (nonveridical) belief about its location.

In all three cases, we can observe a type of bodily response that tracks proxies of others' intentional states from an early stage in development and continuously into adulthood and that thereby facilitates conceptual understanding of the intentional states causing the proxies that it detects. Moreover, increasing conceptual development is accompanied by the development of a more nuanced and situation-dependent deployment of the response in question.

7. Consequences for Social Cognition Research. One obvious consequence of the framework we are espousing is that young children may be able to fix reference to some intentional states before being able to understand that they are doing so. And indeed, this appears to be what we observe in the case of recent studies showing that children as young as 11 months old exhibit implicit understanding of false belief (e.g., Onishi and Baillargeon 2005; Southgate et al. 2007; Baillargeon, Scott, and He 2010)—in which looking-time or eye-tracking techniques are used to gauge infants' expectations about how a person with a false belief will behave—although it is not until years later that they succeed at explicit false-belief tests. These findings have stirred theoretical controversy as to whether children really do track false

https://doi.org/10.1086/670328 Published online by Cambridge University Press

beliefs (thick interpretations; Apperly and Butterfill 2009; Baillargeon et al. 2010) or just complex behavioral rules (thin interpretations; Perner and Ruffman 2005; Hutto 2012).

Endorsing a thick interpretation necessitates an explanation of why children do not succeed at explicit verbal false-belief tasks until after age 4, whereas endorsing a thin interpretation is becoming ever more difficult in the face of mounting evidence that young children are sensitive to sources of knowledge (or of false beliefs) that are fairly decoupled from behavior. For example, in a study involving 15-month-olds, Träuble, Marinović, and Pauen (2010) used an apparatus designed such that an agent could cause a ball to be transferred from one bucket to another by manipulating the apparatus without seeing it (with her back turned). The finding was that infants expect an agent not to have a false belief, even though she did not see the object transfer because she was turned the other way. This demonstrates an impressive ability to reason flexibly about the effects that various kinds of evidence (even nonperceptual evidence) will have on agents' beliefs. Song et al. (2008) found that 18-month-old infants' expectations are modulated if the experimenter communicates to the agent that the ball has been moved but not if she says merely that she likes the ball.

Although this developmental pattern is difficult for existing approaches to account for, it is precisely what the causal theory—as we are applying it to the reference of intentional concepts in development—would predict. According to the causal theory, children may indeed track false beliefs at 11 months but lack the descriptive/theoretical resources to make use of their own tracking abilities. Thus, their eye gaze direction correctly predicts where the agent with a false belief about the location of a desired object will search for it, and it does so not just by virtue of behavioral rules but by postulating an unobservable intentional state. But explicit false-belief understanding requires interpreting these referential processes and integrating them with conceptual knowledge, background information, language, and so on. Moreover, children must also gradually learn to use increasingly various sources of evidence to detect false beliefs using an increasing variety of tracking processes. It is thus no surprise that explicit false-belief understanding develops piecemeal and over the course of several years.

8. Referential Continuity. Although our aim has not been to resolve the eliminativism debate, our discussion has opened up hitherto neglected theoretical options for approaching that debate. Most important, it has revealed that applying the causal theory to intentional concepts entails that their referential status is independent of the truth of the descriptions associated with folk psychology. Although children's implicit and explicit understanding of beliefs, desires, and other intentional states changes in the course of cognitive development, the reference of the corresponding terms may neverthe-

https://doi.org/10.1086/670328 Published online by Cambridge University Press

less persist when there is a consistent reliance on these causal relations and their connection to particular behavioral phenomena to identify, and investigate, the putative intentional causal agents behind it. Reliance on these causal connections can be used to infer an intention to refer persistently to the same causal agent as the explanation for observed behaviors. The integration of responses to various behavioral proxies reviewed in the previous section reveals that children treat particular behavioral phenomena as manifestations of a common underlying causal architecture.

There is evidence that the relations used to fix reference play a continuous role in the development of a conceptual understanding of the intentional state in question. Indeed, we suggest that conceptual development may depend on the consistency of the causal relationships between other agents' behavior and one's own bodily and cognitive responses to that behavior. As such, in a learning process, having fixed referents is of some considerable importance. They act to stabilize the conceptual system and allow knowledge to be accumulated. Just as with paradigm shifts in the history of science, conceptual overhauls are costly events in terms of the cognitive work required to generate new concepts with new reference classes. Particularly, they tend to imply that prior knowledge claims must be jettisoned to the extent that the original concepts were not referring. There is thus an advantage to fixing reference early in a learning process—particularly when it turns out that referential shifts are not in fact further required (i.e., because the reference is correct or at least adequate).

However, one frequent complaint about causal theories of reference, as we noted earlier, is that they make reference too easy, thereby glossing over discontinuities and overlooking contexts in which the possible causes of a phenomenon are underdetermined. This criticism should indeed be taken seriously, as it alerts us to the need to assess whether the development of intentional concepts is also characterized by discontinuities. A thorough reply would presumably necessitate the formulation of criteria according to which reference should be considered discontinuous. Although such a project would be beyond the scope of this article, we would like briefly to point out an important difference between reference fixing in science and reference fixing in cognitive development: the developing cognitive system is not a novel investigative undertaking exploring unknown phenomena with underdetermined hypotheses but, rather, an evolved system. It therefore seems highly plausible that evolutionary history has built in various early emerging tracking devices that reliably and robustly sustain interaction with others' intentional states (or their proxies) and thereby bias subsequent conceptual development in an instrumental fashion. Tracking devices could, for example, ensure that infants are drawn to information that is relevant to understanding others' intentional states and attract their attention to particular paradigmatic behavioral patterns.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

9. Conclusion. In bringing causal theories of reference to bear on everyday concepts of intentional states, we have been pressing the case that theories of reference can be fruitful resources outside of narrow philosophical circles concerned with traditional issues such as scientific realism. In a sense, this is in tension with Mallon et al.'s (2009) skeptical conclusions about the usefulness of theories of reference as a foundation for philosophical arguments. However, we think it is compatible with their assessment since we have been using the causal theory not as the foundation for an argument but as a resource for modeling conceptual development.[7] Specifically, we have been pressing the case that an understanding of the processes by which infants track and thereby anchor reference to others' intentional states is crucial to modeling the development of intentional concepts. Like scientists who postulate a causal agent underlying some observable phenomenon in order to structure their investigation of that causal agent, children assume that intentional states underlie patterns in others' behavior and that there is a relatively fixed causal connection between those intentional states, the behavioral patterns they cause, and their own cognitive and bodily responses to those behavioral patterns. This set of assumptions enables children to interpret their responses as sources of information about others' intentional states and to treat information from various responses as well as from the context and from background knowledge as pertaining to the same entity. This makes it possible for children to integrate these sources of information and to use them to mutually constrain one another, thus bootstrapping their way to a more sophisticated conceptual understanding of intentional states.

With respect to the broader semantic and epistemological questions animating the eliminativism debate and other philosophical discussions of the nature and status of intentional states, our aim has been modest. We have tried to demonstrate that causal theories proffer an account of reference that supports referential continuity despite descriptive change. One consequence of this is that it is possible to argue that even if a putative folk theory of mind turns out to be false in many respects, it does not follow that the intentional concepts featuring in that theory do not refer or that intentional states do not exist. Thus, we have proposed that further research should aim to articulate criteria for referential continuity that will make it possible to reassess the eliminativism debate from the perspective of a causal theory of reference.

7. Interestingly, Cummins (1991) proposes a similar way of using the causal theory to reformulate the issue of the reference of the concept of a belief into a problem to be resolved by psychological research.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

REFERENCES

Anisfeld, M. 1991. "Neonatal Imitation." *Developmental Review* 11:60–97.
Apperly, I. A., and A. B. Butterfill. 2009. "Do Humans Have Two Systems to Track Beliefs and Belief-Like States?" *Psychological Review* 116:953–70.
Baillargeon, R., R. M. Scott, and Z. He. 2010. "False-Belief Understanding in Infants." *Trends in Cognitive Sciences* 14 (3): 110–18.
Baldwin, D., and L. Moses. 1994. "Early Understanding of Referential Intent and Attentional Focus: Evidence from Language and Emotion." In *Origins of an Understanding of Mind*, ed. C. Lewis and P. Mitchell, 365–73. Hillsdale, NJ: Erlbaum.
Becchio, C., A. Cavallo, C. Begliomini, L. Sartori, G. Feltrin, and U. Castiello. 2012. "Social Grasping: From Mirroring to Mentalizing." *NeuroImage* 61 (1): 240–48.
Bermudez, J. 2003. "The Domain of Folk Psychology." In *Minds and Persons*, ed. A. O'Hear. Cambridge: Cambridge University Press.
Bischof-Köhler, D. 1991. "The Development of Empathy in Infants." In *Infant Development: Perspectives from German Speaking Countries*, ed. M. E. Lamb and H. Keller, 245–73. Hillsdale, NJ: Erlbaum.
Brass, M., J. Derrfuss, and D. Y. von Cramon. 2005. "The Inhibition of Imitative and Overlearned Responses: A Functional Double Dissociation." *Neuropsychologia* 43:89–98.
Bushnell, I. W. R. 2001. "Mother's Face Recognition in Newborn Infants: Learning and Memory." *Infant and Child Development* 10 (March): 67–74. doi:10.1002/icd.248.
Carey, S. 2009. *The Origin of Concepts*. 1st ed. New York: Oxford University Press.
Chartrand, Tanya, and John A. Bargh. 1999. "The Chameleon Effect: The Perception-Behavior Link and Social Interaction." *Journal of Personality and Social Psychology* 76 (6): 893–910.
Cheng, C. M., and T. L. Chartrand. 2003. "Self-Monitoring without Awareness: Using Mimicry as a Nonconscious Affiliation Strategy." *Journal of Personality and Social Psychology* 85: 1170–79.
Churchland, P. M. 1981. "Eliminative Materialism and the Propositional Attitudes." *Journal of Philosophy* 78:67–90.
Cummins, R. 1991. "Methodological Reflections on Belief." In *Mind and Common Sense*, ed. R. Bogdan, 53–70. Cambridge: Cambridge University Press.
Dennett, D. 1987. *The Intentional Stance*. Cambridge, MA: MIT Press.
De Vignemont, F., and T. Singer. 2006. "The Empathic Brain: How, When and Why?" *Trends in Cognitive Sciences* 10 (19): 535–41.
Devitt, M. 1981. *Designation*. New York: Columbia University Press.
Dimberg, Ulf. 1997. "Facial Reactions: Rapidly Evoked Emotional Responses." *Journal of Psycholinguistic Research* 11 (2): 115–23.
Doherty, M. J. 2009. *Theory of Mind: How Children Understand Others' Thoughts and Feelings*. Hove: Psychology.
Eisenberg, N., and R. A. Fabes. 1998. "Prosocial Development." In *Handbook of Child Psychology*, vol. 3, *Social, Emotional, and Personality Development*, 5th ed., ed. N. Eisenberg, 701–78. New York: Wiley.
Enç, B. 1976. "Reference of Theoretical Terms." *Nous* 10:261–28.
Farroni, T., G. Csibra, F. Simion, and M. Johnson. 2002. "Eye Contact Detection at Birth." *Proceedings of the National Academy of Sciences of the USA* 99:9602–5.
Feyerabend, P. 1962. "Explanation, Reduction and Empiricism." In *Scientific Explanation, Space, and Time*, ed. H. Feigl and G. Maxwell, 28–97. Minnesota Studies in the Philosophy of Science 3. Minneapolis: University of Minnesota Press.
Fodor, J. 1987. *Psychosemantics*. Cambridge, MA: MIT Press.
———. 1990. *A Theory of Content*. Cambridge, MA: MIT Press.
Frege, G. 1892. "Über Sinn und Bedeutung." *Zeitschrift für Philosophie und philosophische Kritik* C: 25–50.
Geangu, A., O. Bengu, D. Stahl, and T. Striano. 2010. "Contagious Crying beyond the First Days of Life." *Infant Behavioural Development* 33 (3): 279–88.
Gergely, G., H. Bekkering, and I. Király. 2002. "Rational Imitation in Preverbal Infants." *Nature* 415:755.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

Grossman, T., E. Parise, and A. Frederici. 2010. "The Detection of Communicative Signals Directed at the Self in Infant Prefrontal Cortex." *Frontiers in Human Neuroscience* 4:1–5.
Hatfield, E., J. T. Cacioppo, and R. L. Rapson. 1994. *Emotional Contagion*. Cambridge: Cambridge University Press.
Hess, U. 1998. "Mimicry: Fact and Fiction." In *Proceedings of the Xth Conference of the International Society for Research on Emotion*. Storrs, CT: International Society for Research on Emotions.
Hess, Ursula, and Patrick Bourgeois. 2010. "You Smile—I Smile: Emotion Expression in Social Interaction." *Biological Psychology* 84 (3): 514–20. doi:10.1016/j.biopsycho.2009.11.001.
Heyes, C. 2001. "Causes and Consequences of Imitation." *Trends in Cognitive Science* 5 (6): 253–61.
Hood, B., J. Willen, and J. Driver. 1998. "Adults' Eyes Trigger Shifts of Visual Attention in Human Infants." *Psychological Science* 9 (2): 131–34.
Hutto, D. 2012. "Elementary Mind Minding, Enactivist Style." In *Joint Attention: New Developments in Psychology, Philosophy of Mind and Social Neuroscience*, ed. A. Seemans, 307–42. Cambridge, MA: MIT Press.
Johnson, Mark H., Suzanne Dziurawiec, Hadyn Ellis, and John Morton. 1991. "Newborns' Preferential Tracking of Face-Like Stimuli and Its Subsequent Decline." *Cognition* 40 (August): 1–19. doi:10.1016/0010-0277(91)90045-6.
Kitcher, P. 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.
Kripke, S. 1972. "Naming and Necessity." In *Semantics of Natural Language*, ed. D. Davidson and G. Harman, 253–355. Dordrecht: Reidel.
Kroon, F. W. 1985. "Theoretical Terms and the Causal View of Reference." *Australasian Journal of Philosophy* 63 (2): 143–66.
———. 2011. "Theory-Dependence, Warranted Reference, and the Epistemic Dimensions of Realism." *European Journal for Philosophy of Science* 1 (2): 173–91.
Kugiumutzakis, G. 1999. "Genesis and Development of Early Infant Mimesis to Facial and Vocal Models." In *Imitation in Infancy: Cambridge Studies in Cognitive Perceptual Development*, ed. Jacqueline Nadel and George Butterworth, 36–59. New York: Cambridge University Press.
Kuhn, T. 1970. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press.
Lang, P., M. Greenwald, M. Bradley, and A. Hamm. 1993. "Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reactions." *Psychophysiology* 30 (3): 261–73. doi:10.1111/j.1469-8986.1993.tb03352.x.
Lanzetta, J. T., and B. G. Englis. 1989. "Expectations of Cooperation and Competition and Their Effects on Observers' Vicarious Emotional Responses." *Journal of Personality and Social Psychology* 56 (4): 543.
Laurence, E., and S. Margolis. 2002. "Radical Concept Nativism." *Cognition* 86:25–55.
Lycan, W. 1988. *Judgement and Justification*. Cambridge: Cambridge University Press.
Mallon, R., E. Machery, S. Nichols, and S. Stich. 2009. "Against Arguments from Reference." *Philosophy and Phenomenological Research* 79 (2): 332–56.
Meltzoff, A. N. 1995. "Understanding the Intentions of Others: Re-enactment of Intended Acts by 18-Month-Old Children." *Developmental Psychology* 31:838–50.
Meltzoff, A. N., and K. Moore. 1977. "Imitation of Facial and Manual Gestures by Human Neonates." *Science* 198:75–78.
Millikan, R. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.
———. 1998. "A Common Structure for Concepts of Individuals, Stuffs and Real Kinds: More Mama, More Milk and More Mouse." *Behavioral and Brain Sciences* 21:55–100.
———. 2000. *On Clear and Confused Ideas: An Essay about Substance Concepts*. Cambridge: Cambridge University Press.
Moll, H., and M. Tomasello. 2004. "12- and 18-Month-Old Infants Follow Gaze to Spaces Behind Barriers." *Developmental Science* 7:F1–F9.
Mondloch, Catherine J., Terri L. Lewis, D. Robert Budreau, Daphne Maurer, James L. Dannemiller, Benjamin R. Stephens, and Kathleen A. Kleiner-Gathercoal. 1999. "Face Perception during Early Infancy." *Psychological Science* 10 (5): 419–22. doi:10.1111/1467-9280.00179.

https://doi.org/10.1086/670328 Published online by Cambridge University Press

Nola, R. 1980. "Fixing the Reference of Theoretical Terms." *Philosophy of Science* 47 (4): 505–31.
Onishi, K. H., and R. Baillargeon. 2005. "Do 15-Month-Old Infants Understand False Beliefs?" *Science* 308 (8): 255–58.
Pacherie, E. 2005. "Perceiving Intentions." In *A Explicação da Interpretação Humana*, ed. João Sàágua, 401–14. Lisbon: Edições Colibri.
Perner, J., and T. Ruffman. 2005. "Infants' Insight into the Mind: How Deep?" *Science* 308:212–14.
Prinz, J. 2002. *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: MIT Press.
———. 2005. "The Return of Concept Empiricism." In *Categorization and Cognitive Science*, ed. H. Cohen and C. Leferbvre. Cambridge: Elsevier.
Psillos, S. 1999. *Scientific Realism: How Science Tracks Truth*. London: Routledge.
Putnam, H. 1975. "The Meaning of 'Meaning.'" In *Mind, Language and Reality: Philosophical Papers*, ed. H. Putnam, 215–71. Cambridge: Cambridge University Press.
Russell, B. 1905. "On Denoting." *Mind* 14:479–93.
Sagi, A., and M. L. Hoffman. 1976. "Empathic Distress in Newborns." *Developmental Psychology* 12:175–76.
Schouten, M., and H. De Jong. 1998. "Defusing Eliminative Materialism: Reference and Revision." *Philosophical Psychology* 11 (4): 489–509.
Senju, A., and G. Csibra. 2008. "Gaze-Following in Human Infants Depends on Communicative Signals." *Developmental Science* 18 (9): 678–71.
Song, H., K. Onishi, R. Baillargeon, and C. Fisher. 2008. "Can an Agent's False Belief Be Corrected through an Appropriate Communication? Psychological Reasoning in 18-Month-Old Infants." *Cognition* 109:295–315.
Southgate V., A. Senju, and G. Csibra. 2007. "Action Anticipation through Attribution of False Belief by 2-Year-Olds." *Psychological Science* 18:587–92.
Stanford, P. K., and P. Kitcher. 2000. "Refining the Causal Theory of Reference for Natural Kind Terms." *Philosophical Studies* 97 (1): 97–127.
Stel, M., E. van Dijk, and E. Olivier. 2009. "You Want to Know the Truth? Then Don't Mimic!" *Psychological Science* 20 (6): 693–99.
Stern, D. 1985. *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology.* New York: Basic.
Stich, S. 1983. *From Folk Psychology to Cognitive Science*. Cambridge, MA: MIT Press.
Tomasello, M., M. Carpenter, J. Call, T. Behne, and H. Moll. 2005. "Understanding and Sharing Intentions: The Origins of Cultural Cognition." *Behavioral and Brain Sciences* 28:675–735.
Träuble, B., V. Marinović, and S. Pauen. 2010. "Early Theory of Mind Competencies: Do Infants Understand Others' Beliefs?" *Infancy* 15 (4): 434–44.
Walton, Gail E., N. J. A. Bower, and T. G. R. Bower. 1992. "Recognition of Familiar Faces by Newborns." *Infant Behavior and Development* 15 (2): 265–69. doi:10.1016/0163-6383(92)80027-R.
Wang, B. Y., and A. Hamilton. 2012. "Social Top-Down Social Response Modulation (STORM): A Model of the Control of Mimicry in Social Interaction." *Frontiers in Human Neuroscience* 6:153.
Xu, F., and S. Carey. 1996. "Infant Metaphysics: The Case of Numerical Identity." *Cognitive Psychology* 30:111–53.
Zahn-Waxler, C., M. Radke-Yarrow, E. Wagner, and M. Chapman. 1992. "Development of Concern for Others." *Developmental Psychology* 28 (1): 126–36.

https://doi.org/10.1086/670328 Published online by Cambridge University Press