Cognition 119 (2011) 197–215

Contents lists available at ScienceDirect

Cognition

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

The accidental transgressor: Morally-relevant theory of mind

Melanie Killen [a,*], Kelly Lynn Mulvey [a], Cameron Richardson [a], Noah Jampol [a], Amanda Woodward [b]

[a] Department of Human Development, University of Maryland, College Park, MD, USA
[b] Department of Psychology, University of Chicago, USA

ARTICLE INFO

Article history:
Received 18 May 2010
Revised 5 January 2011
Accepted 20 January 2011
Available online 4 March 2011

Keywords:
Theory of mind
Moral judgments
Developmental social cognition
Social knowledge

ABSTRACT

To test young children's false belief theory of mind in a morally relevant context, two experiments were conducted. In Experiment 1, children ($N = 162$) at 3.5, 5.5, and 7.5 years of age were administered three tasks: prototypic moral transgression task, false belief theory of mind task (ToM), and an "accidental transgressor" task, which measured a morally-relevant false belief theory of mind (MoToM). Children who did not pass false belief ToM were more likely to attribute negative intentions to an accidental transgressor than children who passed false belief ToM, and to use moral reasons when blaming the accidental transgressor. In Experiment 2, children ($N = 46$) who did not pass false belief ToM viewed it as more acceptable to punish the accidental transgressor than did participants who passed false belief ToM. Findings are discussed in light of research on the emergence of moral judgment and theory of mind.

1. Introduction

Understanding the intentions of another person reflects a core aspect of moral judgment (Killen & Smetana, 2008; Turiel, 2006; Zelazo, Helwig, & Lau, 1996) and theory of mind (Perner, Frith, Leslie, & Leekam, 1989; Wellman, 1990; Wellman & Liu, 2004; Woodward, Sommerville, & Guajardo, 2001). For several decades, researchers in the field of moral development have demonstrated how young children, as early as 3 and 4 years of age, evaluate moral transgressions on the basis of the negative intrinsic consequences to others rather than on external consequences such as teacher mandates or punishment (Smetana, 2006; Turiel, 2006). In addition, several decades of research on children's theory of mind has documented the emergence of increasingly robust reasoning about others' mental states during the same time period (Carpendale & Lewis, 2006; Perner et al., 1989; Wellman, Cross, & Watson, 2001).

1.1. Moral judgment and theory of mind

Recently, there has been interest in whether theory of mind competence is related to understanding another's intentions regarding *morally relevant* actions (Astington, 2004; Chandler, Sokol, & Wainryb, 2000; Knobe, 2005; Lagattuta, 2005; Leslie, Knobe, & Cohen, 2006; Wellman & Miller, 2008; Zelazo et al., 1996). The foci of the studies differ but converge on the overall expectation that theory of mind and moral judgment are interrelated. What is apparent is that the way that theory of mind is assessed is fairly consistent across studies, with measures including false belief competence in childhood (most often) and measures assessing reasoning about the desires of others. The moral judgment tasks, however, reflect a wide range of measures, from punishment acceptability for transgressions, to ratings of severity of a transgression as well as the

* Corresponding author. Address: Department of Human Development, 3304 Benjamin Building, University of Maryland, College Park, MD 20742-1131, USA. Tel.: +1 301 405 3176 (office); fax: +1 301 405 2891.
E-mail address: mkillen@umd.edu (M. Killen).

0010-0277/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2011.01.006

desire to acquire something that has been prohibited by adults. Moreover, the designs often reflect the administration of two discrepant tasks, one for moral judgment and one for theory of mind.

A small handful of studies have measured children's evaluations of tasks that involve both theory of mind competence as well as moral judgment evaluations. Early research that demonstrated children's differentiation of accidental and intentional acts, contrary to Piaget's theory (1932) about the lack of differentiation, revealed, for example, that 3–7 year olds judged that recipients identified a bad outcome (harm) to be the result of an intentional rather than accidental action, and that children would assign more blame when an act was intentional rather than accidental (Yuill & Perner, 1988). Chandler and colleagues (Carpendale & Chandler, 1996; Chandler, Sokol, & Hallett, 2001) also demonstrated that children 5–7 years old rated intentional acts as "more bad" than accidental acts. Further, Leslie, Knobe, and colleagues (2006) found that contrary to the traditional expectation that theory of mind is necessary for moral judgment, children's moral decisions regarding attributions of blame influence their interpretations of other's intentions (theory of mind). Thus, these studies provided support for the theory that young children think about the motives of others, assign blame when acts are intentional, and, at times, interpret intentionality from a moral perspective.

Thus, while existing research has demonstrated that young children distinguish between intentions and outcomes, the connections documented, to date, are rather global, and more detailed investigations are necessary to specify how these connections are made in early childhood. To bolster this point, recent neuroscience research has examined the neural underpinnings between mental state attribution and moral evaluations in accidental transgression scenarios and has revealed a well-specified region of the brain that is integral to intent attribution during act evaluation (Young & Saxe, 2009). Moreover, accumulated research with adults has revealed a strong relationship between theory of mind competence and moral judgment (Knobe, 2005; Pettit & Knobe, 2009), but, as mentioned, questions about origins and development remain to be investigated.

Most developmental studies, to date, have used only one assessment for a single scenario, with some studies focused on children's affective responses towards the target and other studies on children's assignment of blame or punishment. What is lacking is a multi-measure approach in which prototypic moral judgment and false belief tasks are administered along with an embedded morally-relevant false belief theory of mind task in a single study. This type of design would directly address developmental questions about these early social cognitive competencies.

1.2. Moral judgment research

In her review of the literature, Smetana (2006) identified a robust measure for assessing moral judgment, in which children are asked to provide social reasons for what makes an act wrong from a moral viewpoint, such as focusing on physical or psychological harm, the unfairness of the act, or the lack of equal and just treatment of others in contrast to a conventional viewpoint, such as focusing on authority mandates, punishment, or rule violations. This measure, which has served as a prototypic moral judgment assessment given its extensive empirical validation (Helwig, Tisak, & Turiel, 1990; Killen, 2007; Nucci, 2001; Smetana, 2006; Turiel, 1998, 2008), derives from social domain theory (Turiel, 2006), and has been replicated in many countries, with cross-cultural generalizability, as well as with urban and rural samples of high and low socioeconomic status (for a review see Wainryb, 2006). In this methodology, assessments are made of children's evaluations of moral transgressions, such as an act of harm, in which few other competing considerations are involved (making it "prototypic") and children are asked to judge the act as well as to provide justifications for their judgments.

1.3. False belief theory of mind competency

A "prototypic" verbal measure of theory of mind that has been used extensively in past research is the false belief task (Wimmer & Perner, 1983). This task assesses children's ability to use a person's belief state to predict his or subsequent actions when those beliefs differ from reality and from the child's own knowledge. Typically, children younger than 4–5 years of age fail this task, predicting the person's actions based on reality rather than the person's false belief (Wellman et al., 2001). This task was originally assumed to measure the onset of ability to represent others' false beliefs (Wellman, 1990). Recent findings with infants (Onishi & Baillargeon, 2005) and young children (Friedman & Leslie, 2005), however, have cast doubt on this strong conclusion about onset. Even so, it is clear that this task remains relevant to the measurement of the child's ability to recruit false belief information to reason about explicit scenarios involving intentionality (Wellman & Liu, 2004).

Prototypic false belief "theory of mind" tasks measure one's access to knowledge about the physical world, such as whether another person who did not witness a location change of an object will know where to look for it ("I know it's been moved, but X does not know it's been moved and therefore will look in the place that he/she last saw it"); by design, the task itself has limited social content. The social aspect of the competence is the realization of how other people's minds work (in contrast to how other non-social objects work) but the non-social aspect of the false belief task is the lack of a specified social relationship between the "mover" and the "owner" of the object.

Further, in the case of the location change false belief task, marbles are moved from one box to another box when another child is out of the room (the participant is asked where the returning child will look for the marbles) and no social information is provided regarding who owns the marbles, the intentions of the "mover" of the marbles, or the relationship between the "mover" and the "observer" (e.g., friends, strangers). Yet the relationship between individuals in social situations has been shown to be significantly related to young children's evaluations of acts. For example, when preschoolers are told that "child X