Cognitive Development 40 (2016) 9–23

Contents lists available at ScienceDirect

Cognitive Development

# Coding choices affect the analyses of a false belief measure☆

David M. Sobel*, Joseph L. Austerweil

*Brown University, United States*

ARTICLE INFO

*Article history:*
Received 17 June 2015
Received in revised form 15 July 2016
Accepted 8 August 2016

*Keywords:*
Unexpected contents task
Theory of mind
Metaanalysis

ABSTRACT

The unexpected contents task is a ubiquitous measure of false belief. Not only has this measure been used to study children's developing knowledge of belief, it has impacted the study of atypical development, education, and many other facets of cognitive development. Based on a review of articles using this task, we show that there is no consensus regarding how to score this measure. Further, examining both a logit analysis of performance on this measure and performance of a large sample of preschoolers, we show that which coding scheme researchers used to analyze raw data from this measure has a reliable effect on results, particularly when smaller sample sizes are used. Integrating our results, we conclude that the most frequently used coding scheme is flawed. We recommend best practices for scoring the unexpected contents task, and that researchers examine how they analyze data from this measure to ensure the robustness of their effects.



**Contents**



☆ This research was supported by a grant from the National Science Foundation (NSF 1223777). We would like to thank Rose DeRenzio, Isobel Heck, and Antonija Kolobaric who helped with coding, and Nadia Chernyak, Christopher Erb, Tamar Kushnir, Deanna Macris, and Tiffany Tassin for providing fruitful discussion on this topic. Joseph Austerweil is now at the University of Wisconsin-Madison.

* Corresponding author at: Department of Cognitive, Linguistic, and Psychological Sciences, Box 1821, Brown University, Providence, RI 02912, United States.
*E-mail address:* Dave_Sobel@brown.edu (D.M. Sobel).

http://dx.doi.org/10.1016/j.cogdev.2016.08.002
0885-2014/

## 1. Introduction

For over 25 years, researchers have been interested in children's developing *theory of mind* – their developing knowledge of others' mental states (e.g., Flavell, 1999). While these investigations emphasize the nuanced and staggered developmental trajectory of children's mental state knowledge, one aspect of theory of mind development – children's understanding of another's false beliefs – has dominated both the theoretical and empirical landscape (see e.g., Apperly & Butterfill, 2009 Perner, 1991; Wellman, Cross, & Watson, 2001).

Measures of false belief typically fall into one of two categories, *unexpected transfer* and *unexpected contents.* In the unexpected transfer task (Wimmer & Perner, 1983), children are introduced to an agent, who is shown to have a belief about the location of an object, because s/he has perceptual access to that location. The object's location is then changed, and critically, the agent does not have perceptual access to or communication about the new location. Children are then asked about the (false) belief the agent has about the location of the object.

The unexpected contents task is similar in that children reason about another's false belief, but also are often probed about their own belief state. Children are introduced to a familiar container, which is revealed to be deceptive. For instance, in Perner, Leekham, and Wimmer (1987), children are introduced to a *Smarties* box (a type of candy), which is revealed to contain either a pencil or a pencil and Smarties while their friend is outside of the room. In their procedure, children are asked three questions after the contents of the box are placed back inside: A question regarding their memory: "Can you remember what's inside here?" (*control* question), a question asking about their knowledge of their own mental states (*self* question): "But what did you think was in here?", and a question asking about their knowledge of another's beliefs about the contents of the box (*other* question), "What will <name of friend> think is in here?" The results of studies using these procedures are highly consistent. In a metaanalysis of over 50 studies, Wellman et al. (2001) showed that children's ability to answer the test questions develops between the ages of 36 and 60 months in a manner that reflects their own understanding of representational change and that another can hold a false belief.

We have observed, however, that papers reporting this measure score the unexpected contents task differently. For example, Perner et al. (1987) reports the results of the test questions, but excludes the data generated by children who failed the control measure. They write that three of their participants "failed to remember there was a pencil. . .in the box (control question). Their responses to the test questions were therefore meaningless." (p. 133). Dissimilarly, Astington and Jenkins ((1999) see also Wellman and Liu (2004)) used a coding scheme in which "passing" the measure requires a correct response to the control question – that is, if children fail the control measure, they are counted as failing the task, regardless of their response to the test question(s). Many other papers either do not ask this memory question or do not consider the answer to this question when scoring the measure.

Here, we examine the role of coding scheme in research on children's developing false belief. We first examine how the unexpected contents task is scored. We review a subset of the literature that uses this measure, describing the different ways it has been scored. Next, we consider whether differences in the way this task is scored affects results and potential interpretations of findings. We conduct a logit analysis using methods similar to Wellman et al. (2001) to demonstrate that the choice of coding scheme overall influences analyses of performance. We then discuss a published article, which presented their results in sufficient detail such that we can calculate whether the statistical significance of their results depend on the choice of coding scheme. We find that the results in this paper might depend on the choice of coding scheme. We then analyze an in-house data set of ~1200 preschoolers who have been given this measure, using a bootstrap analysis to consider the potential effect of coding scheme on effect sizes. All of our analyses support the same conclusions: (1) There is no agreement regarding how this measure is scored; (2) the choice of coding scheme affects analyses, and (3) many researchers use what we believe is the least optimal coding scheme. Finally, to conclude the paper, we offer a set of best practices based on the results of our analyses.

## 2. Literature review

The goal of the literature review is to examine whether there is agreement in how researchers have scored the unexpected contents measure over the last 25+ years. We conducted a literature review on papers that report children's performance on an unexpected contents task and coded the manner in which data collected from children were scored.

To determine papers to analyze, we first examined the well-known metaanalysis on children's false belief knowledge (Wellman et al., 2001), and included every paper that reported data from the unexpected contents task. We then performed a Google Scholar search for the terms "unexpected content task" to find papers published between 2001 and 2013. Overall, we found 88 papers that reported at least one experiment using the unexpected contents task or a variant of the measure (123 separate experiments on 9365 participants). Table 1 reports these papers.

For each experiment, the first author and a research assistant (who was blind to the goals of this review) read the methods and scoring sections of the experiment to determine whether a control question regarding the child's memory of the contents of the box was asked. If so, the coders examined whether and how the control question was included in the scoring of the task. Experiments were categorized into eight distinct scoring methods: