Modelling the executive components involved in processing false belief and mechanical/intentional sequences

Hiromi Tsuji1* and Peter Mitchell2

1Osaka Shoin Women’s University, Higashiosaka, Japan
2University of Nottingham, UK

To understand the executive demands of the false-belief (FB) task relative to an alternative theory-of-mind (or mechanical causality) task, picture sequencing, the present study used path analyses. One hundred and sixty-six children between 3 and 6 years old completed the FB and picture-sequencing tasks, three executive function tasks (updating, inhibition, and shifting), and the receptive language test. The model with the best fit indicated that FB performance had a direct contribution from shifting of attention and inhibitory control, which was independent of the significant contribution made by picture sequencing. This model indicates that FB inference requires more executive processing than picture sequencing, which is used as an alternative task to measure theory of mind.

Statement of contribution

What is already known on this subject?
- The majority of researchers use the false-belief task to assess mentalizing ability in young children.
- Sources of information used in various different mentalizing tasks require different levels of cognitive demand.
- Many executive functions (EFs) are involved in children’s judgements of false belief.

What does this study add?
- A statistical model was created to compare processing requirements of false-belief and picture-sequencing tasks.
- The model supported the claim that the false-belief task involves considerably more than just mentalizing.
- Shifting the focus of attention was an EF that was found to be a key component of performance in the false-belief task.

Theory of mind (ToM) is a framework used to illuminate children’s developing ability to impute inner states, principally mental states, to other people. The most common method used to investigate development in this arena employs a test of false belief (Wellman, Cross, & Watson, 2001). A long-standing debate has surrounded how best to explain children’s developing ability to pass a test of false belief (FB), such as the unexpected transfer test in which child participants are asked to predict where a protagonist will
search for a coveted object (Wimmer & Perner, 1983). If children predict that the protagonist, who is unaware of the relocation of the object, will look in the place she or he last saw that object, they are credited with an understanding of FB in that they apparently individuate between what they know is true (the object is in Location A) and what they think the protagonist believes is true (she or he falsely believes the object is in Location B). Some researchers argued that development is driven by children acquiring a concept of belief at about 4 years of age (Gopnik & Wellman, 1992; Perner, 1991; Wellman et al., 2001). In contrast, some researchers argued that children already possess a concept of belief by 4 years of age but still give the ‘wrong answer’ in an unexpected transfer test due to various task demands apart from those relevant to having a concept of belief (Lewis & Osborne, 1990; Mitchell & Lacohée, 1991; Russell, Mauthner, Sharpe, & Tidswell, 1991; Siegal & Beattie, 1991). Specifically, Russell et al. were the first to suggest that young children might err not because they have difficulty with the concept of belief but primarily because of the executive demands of the unexpected transfer test. In agreement, Leslie and Thaiss (1992) suggested that developing an ability to pass an unexpected transfer test depends on maturation of the executive processes that the task draws upon.

Bloom and German (2000) went as far as suggesting that the unexpected transfer test is not fit for the purpose of investigating development of a ‘theory of mind’; at the very least, they suggested a priority is to understand what kind of demands are imposed by the unexpected transfer test. The purpose of the current research was to begin to address this question by testing children on a variety of tests of executive processing in addition to an unexpected transfer test. The aim was to generate a statistical model that would optimally explain variance in performance data associated with the unexpected transfer test and an alternative measure that also putatively taps into the child’s understanding of others’ inner states.

**A framework for analysing mentalizing tasks**

To understand the cognitive demands of the FB task, defining sources of information that a participant processes during the task is important. Achim, Guitton, Jackson, Boutin, and Monetta (2013) outlined the eight sources of information framework (8-SIF) to characterize tasks that are currently available for a mentalizing judgement. According to the framework, mentalizing tasks can involve information for agents and for contexts. Both types of information can be drawn immediately from a given task or from a person’s memory, or from both. The immediate information for the agent and for the context can be presented by perceptual and linguistic means. The information drawn from a person’s memory of the agent and of the context includes specific or general information. Specific information includes the details that the participants already know about the protagonist’s characteristics as well as the surrounding environment related to the task, whereas general information refers to general knowledge about people and contexts, which the participants can recall and use spontaneously. Achim et al. suggest that these sources of information could contribute to, and influence, the accuracy of mentalizing judgements.

**False-belief task and picture-sequencing task**

A classic FB task involves a verbal presentation of a scenario with a puppet enactment in which a protagonist comes to hold a FB about the location of an object due to the unexpected transfer of that object. In this task, participants are required to process the perceptual and linguistic information in the story (immediate information), and then, they
need to access social knowledge about what people would normally do in a case of the unexpected transfer of an object (stored information) to make a correct judgement.

The picture-sequencing task (Baron-Cohen, Leslie, & Frith, 1986) is another type of task intended to assess children’s mentalizing ability (Achim et al., 2013). In this task, participants are asked to rearrange four snapshots of an event into sequential order such that it is coherent in relation to an agent’s mental states, including their false beliefs. This task is a good point of comparison for an unexpected transfer task because it imposes different processing demands while still tapping into the child’s understanding that others can hold false beliefs. According to the 8-SIF (Achim et al., 2013), these two tasks impose similar processing demands with one major difference. Both tasks include sources of information drawn from memory for general assumptions about the agent and for the context but do not include specific information for the agent or the context. However, these tasks differ in that uniquely for the FB task, the immediate information for the agent and for the context is presented using both perceptual and linguistic modalities, whereas no immediate linguistic information is available for the picture-sequencing task. In summary, a FB task requires participants to engage in the dual tasks of processing two different sources of information, which would place extra cognitive load on participants when processing linguistic information; this extra load is absent from the picture-sequencing task.

When characterizing mentalizing based on the 8-SIF, it becomes clear that tackling a FB task requires processing two sources of information simultaneously. Thus to make a correct mentalizing judgement in a FB task, children would need to have sufficient executive processing capacity. Even adults are reported to find the cognitive load challenging in a FB task. In their study of implicit FB processing, Schneider, Lam, Bayliss, and Dux (2012) found that adult participants who were subject to the dual-task condition failed to track the false location based on a protagonist’s FB, whereas those who were under the no-load condition correctly tracked the false location. This result suggests that mentalizing at the implicit level is also disrupted when extra cognitive load associated with an unexpected transfer test is applied.

Relationship between executive functions and false belief

To cope with the cognitive load in the context of unexpected transfer, certain levels of executive processes may be required. Development of executive functions (EFs) has become one of the foci in the discussion of how children acquire their capacity to exercise mentalizing judgements in the FB task. EFs encompass a wide range of abilities such as monitoring and updating of working memory (updating), inhibitory control of prepotent responses (inhibition), and cognitive flexibility in shifting mental sets (shifting) (Miyake et al., 2000), and provide the foundation for reflective thinking, reasoning, and self-control of behaviours (Blair, 2016; Zelazo et al., 2013). These three components are commonly recognized as distinct functions that are assumed to contribute differently to more complex cognitive processing tasks (Miyake et al., 2000), perhaps depending on the degree to which higher order rules are involved. Indeed, some researchers argue that certain EF and ToM tasks are united by drawing upon such higher order rules (Riggs, Peterson, Robinson, & Mitchell, 1998), as articulated in Zelazo et al.’s (2013) ‘Cognitive Complexity and Control’ theory.

The dissociation of EF components remains inconclusive in developmental studies, especially for preschool children (Best & Miller, 2010; van der Ven, Kroesbergen, Boom, & Leseman, 2013). Most studies that have examined the role of executive processes in
mentalizing used some but not all of the components of EFs to measure this capacity. This may have led to lack of clarity over the form of any causal relationships between EFs and ToM in that EFs were handled sometimes as a single component and sometimes as separate elements in the statistical analyses. The present study assumes all three components are important factors in executive processing in mentalizing tasks. If each component had a unique association with mentalizing, it would provide support for the three-component structure of EFs in preschool children and bring new insights to the understanding of developmental trajectories for these components (Best & Miller, 2010).

The findings from a meta-analysis suggest a weak to moderate association between executive processing capacity and FB judgement across different cultural contexts (Devine & Hughes, 2014). Devine and Hughes (2014) support a hybrid emergence–expression hypothesis, which is derived from both the emergence account that EFs play a functional role in the conceptual development of FB judgement (Russell, 1996) and the expression account that FB tasks place incidental demands on EFs (Russell et al., 1991). Devine and Hughes concluded that early developmental variations predict later variations in FB judgement. The expression account assumes that if mentalizing tasks that place less executive demands on EFs were used, then the correlations between such mentalizing tasks and EFs would be reduced in comparison with the original FB task. In an effort to remove pure incidental task demands in FB judgement, several studies (Carlson, Claxton, & Moses, 2015; Perner, Lang, & Kloo, 2002) examined the associations between tasks with fewer executive demands and EFs, and these studies still found associations between mentalizing tasks and EFs, especially for the component of inhibition, thereby supporting the emergence account. So far, it has become clear that EFs play a role in the development of mentalizing in general. However, the exact relationships between these three components of EFs and different forms of mentalizing tasks have yet to be explored.

When the sources of information are considered in these tasks with fewer executive demands, irrespective of the degree of conceptual difficulty in tasks, all the tasks require processing of immediate information conveyed through both a linguistic and a perceptual channel on the 8-SIF (Achim et al., 2013). In this respect, even conceptually easier tasks may still require a certain level of executive processing, which led to the identification of significant relationships between EFs and these modified tasks. Thus, it is unclear whether these observed relationships are due to common demand derived from conceptual overlap or cognitive load.

The approach in this study
As argued by Bloom and German (2000), if a FB task is more cognitively demanding than another mentalizing task such as the picture-sequencing task, then associations between EFs and mentalizing tasks for the FB and picture-sequencing tasks will show a difference. It is possible that the picture-sequencing task may have weak associations with EFs in comparison with FB tasks. Alternatively, different components of EFs may be associated with respective mentalizing tasks. To examine these relationships, the present study conducted path analyses. A recursive model was employed in which a path is only drawn from picture sequencing to FB according to the assumption that FB is measuring more than mentalizing (Bloom & German, 2000). The magnitude of the direct effects from three components of EFs (working memory, shifting of attention, and inhibitory control) was examined while controlling variance associated with receptive language and age, as well as indirect effects from these variables via the picture-sequencing task. We predict that a
FB task, which requires more cognitive load in processing sources of information, is likely to place more demands on the EFs of inhibitory control than a picture-sequencing task.

**Method**

**Participants**

One-hundred and sixty-six children (94 boys) between 3 and 6 years old ($M_{age} = 5.0$, $SD = 0.86$, range 3.5–6.5) participated in this study. They were from a kindergarten located in middle to lower-middle class communities located on the outskirts of Osaka city. Ethical approval was granted by the author’s academic institution, and written informed consents were obtained from the parents of the participants.

**Design and materials**

Mentalizing abilities were assessed using picture-sequencing and FB tasks. Executive functions were assessed using the pointing–Stroop task (Berger, Jones, Rothbart, & Posner, 2000), the Dimensional Change Card Sort (DCCS) task (Zelazo, 2006), and the digit span task. Receptive language was assessed using a Japanese version of the PVT (Ueno, Nakoshi, & Onuki, 2008). The language measure was used to partial out variance shared between the focal tasks that were explained by language ability.

**Mentalizing tasks**

**FB tasks**

Two unexpected transfer tasks (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983) in which a puppet protagonist holds a FB about the location of an object were used to measure theory-of-mind ability. Each task included a memory question for the initial location of the item, a reality question for the present location of the item and a FB question for the protagonist’s belief about the location of the item. All three questions had to be answered correctly for children to be coded as passing the task.

**Picture-sequencing task**

Four pictures, depicting scenes that represented an event involving causality or mental states, were presented to the children. The children were asked to put them in order to make a story. Four sets of task stimuli were chosen from the picture-sequencing test (Baron-Cohen et al., 1986). These included two sets for causality events, a person causally interacting with an object and an object causally interacting with an object (mechanical condition), and two sets of stimuli involving the attribution of people’s mental states for a missing teddy bear and a missing bar of chocolate (intentional condition). The children’s ability to represent events was assessed based on the correct sequencing of pictures to make the story. The scoring system used by the original study was followed.

**Executive function tasks**

For the three components of EFs, we chose the pointing–Stroop task (Berger et al., 2000) for inhibitory control, the DCCS task (Zelazo, 2006) for shifting of attention and
cognitive flexibility, and the digit span task to measure verbal working memory (WM). These tasks were chosen because they are well established, are reliable, and are suitable for children as young as 3 years old (Diamond, 2013; van der Ven et al., 2013; Zelazo, 2006). These tasks only require a simple response of touching one of the stimuli or repeating the digits. The first two tasks were presented to the participants on a computer with a touch panel screen. The digit span task was administered manually.

**Pointing–Stroop task**

This task was based on Berger et al. (2000) and comprises three blocks, which were administered in the following order: practice (two trials), compatibility (eight trials), and incompatibility (eight trials). The trial was started by touching a ‘focal point’ on the screen, which initiated the display of cat and dog pictures with simultaneous audio presentations of the animal’s sound ‘meow’/’woof’. The child needed to touch the picture that was compatible/incompatible with the animal sound she or he heard.

---

**Figure 1.** Trial sequence of stimuli presentation for the audio Stroop task. For the compatibility trials (a), when the audio stimulus of ‘meow’ is presented, the child needed to touch the cat picture, whereas when the audio stimulus of ‘woof woof’ is presented, the child needed to touch the dog picture. For the incompatibility trials (b), when the audio stimulus of ‘meow’ is presented, the child needed to touch the dog picture, and vice versa.
(Figure 1). For scoring, the number of correct responses in incompatibility trials was used.

**DCCS task**
This task was developed from the original manual presentation of DCCS (Zelazo, 2006). The children’s correct responses after introducing a new sorting rule were measured. A block of two practice trials, which was repeated until the child understood the task, preceded the pre-switch block (six trials) and the post-switch block (six trials). When the child touched one of the pictures (hat or bag), a sound was made to acknowledge the action regardless of accuracy. The child initiated the next trial when she or he was ready by touching the focal point on the screen (Figure 2). For scoring, the number of correct responses in the post-switch condition was used.

**Digit span task**
Following the instructions of KABC (Kaufman & Kaufman, 2004), a series of digits were presented to the child orally. The child needed to repeat back the digits to the experimenter in a forward order. The child was awarded one point for each trial in which the series of digits were repeated back correctly.

**Receptive language**
A Japanese version of PVT (Ueno et al., 2008) assessed the children’s general receptive language. In the PVT, the child was presented with four pictures from which they selected the picture that matched with the language they heard. Adjusted scores, taking into account the probability of guessing the correct picture, were used.

**Procedures**
Children were tested individually in a quiet room in one session, which lasted 25–30 min. The order of the tasks was counterbalanced with the exception of the PVT, which was always administered last because the duration of this task varies dramatically depending on the individual’s language ability and this may have had an effect on the child’s motivation for any subsequent tasks.

![Figure 2](image_url). Trial sequence for the stimuli presentation of DCCS. In the pre-switch block (e.g., colour rule), when a blue bag is presented, the child needs to touch the picture of a ‘blue hat’. In the post-switch block (e.g., shape rule), when a blue bag is presented, the child needs to touch a picture of a ‘red bag’. 
Analytic strategy
Path analysis was used to test the model and the relationships between the measured variables. Analyses were performed using the statistical package AMOS 21.0. (IBM SPSS, Chicago, IL, USA)

Specification of the model
A theoretical model specified in the literature (e.g., Bloom & German, 2000) is a recursive model in which there is a unidirectional relationship from the picture-sequencing task to FB task, where the FB task requires additional and independent input from EFs. For the picture-sequencing task, fewer EFs are required; specifically, the ‘shifting’ component is not needed, as the picture-sequencing task does not require processing of linguistic input for competing representations of reality and belief.

Identification of the model
To test the specified model using the observed data, both the parameter estimation and the adequacy of model fit were examined. To test the model for goodness-of-fit, the following criteria were computed: Chi-squared test ($p > .05$), comparative fit index (CFI $\geq .90$) and root mean square error of approximation (RMSEA $\leq .06$), and goodness-of-fit index (GFI $> .90$) (McDonald & Ho, 2002; Hu & Bentler, 1999). For the parameter estimates, age and PVT were included in the model regardless of statistical significance because they were considered as compound variables to be controlled for in explaining the relationships between EFs and mentalizing.

Results
Descriptive statistics for all the measures are summarized in Table 1. These are tabulated by age to show developmental trends. There were no gender differences for the tasks, with the exception of DCCS in which girls scored higher than boys: $t(164) = 3.23$, two-tailed $p < .001$. Cohen’s $d = .51$. Thus, gender was not considered in subsequent analyses. For picture-sequencing scores, the correlations between FB and the two types of picture sequencing (mechanical and intentional) were similar ($r = .55$ and $r = .58$, $ps < .001$) and there was a significant correlation between the intentionality and mechanical subdomains ($r = .61$, $p < .001$), and intertask item analysis (Table S1) found no justification for separating these variables of picture sequencing into their component parts. Thus, these scores were combined to make a single measurement for further analyses. The skewness of all the measures was below 1.5, and the kurtosis was below 2.0.

Pearson’s correlation coefficients between the measurements are summarized in Table 2. All the variables interrelated positively with a moderate to strong degree of association.

To examine how performance measured by the FB and the picture-sequencing tasks was related to EFs, path analyses using AMOS with maximum-likelihood estimation were conducted following Analytic strategy described in Method section. A total score for the FB and the picture-sequencing tasks was, respectively, set as endogenous variables in a

---

1 Univariate distributions were within the recommended ranges for skewness < 2 and kurtosis < 7 (West, Finch, & Curran, 1995).
recursive path model with a path from the picture-sequencing task to the FB task while WM, Stroop, DCCS, receptive language, and age were set as exogenous variables. From the theoretical model specified in Analytic strategy section, the initial model which excluded a path from DCCS (shifting component of EFs) to the picture-sequencing task was tested. In this model, the parameters in the paths from WM to both the picture-sequencing task ($b = .066, p = .33$) and FB task ($b = .007, p = .92$) were found to be non-significant. The model fit index was deemed to be sufficient: $\chi^2(1) = .72, p = .40$, CFI = 1.00, AGFI = .97, RMSEA < .0001, 90% CIs [0.00, 0.19], AIC = 54.72.

Another model excluding the non-significant paths from WM to picture sequencing and FB was considered next. This model indicated good fit with goodness-of-fit statistics: $\chi^2(3) = 1.95, p = .58$, CFI = 1.00, AGFI = .97, RMSEA < .0001, 90% CIs [0.00, 0.11], AIC = 54.72.

recursive path model with a path from the picture-sequencing task to the FB task while WM, Stroop, DCCS, receptive language, and age were set as exogenous variables.

From the theoretical model specified in Analytic strategy section, the initial model which excluded a path from DCCS (shifting component of EFs) to the picture-sequencing task was tested. In this model, the parameters in the paths from WM to both the picture-sequencing task ($b = .066, p = .33$) and FB task ($b = .007, p = .92$) were found to be non-significant. The model fit index was deemed to be sufficient: $\chi^2(1) = .72, p = .40$, CFI = 1.00, AGFI = .97, RMSEA < .0001, 90% CIs [0.00, 0.19], AIC = 54.72.

Another model excluding the non-significant paths from WM to picture sequencing and FB was considered next. This model indicated good fit with goodness-of-fit statistics: $\chi^2(3) = 1.95, p = .58$, CFI = 1.00, AGFI = .97, RMSEA < .0001, 90% CIs [0.00, 0.11], AIC = 54.72.

Table 1. Descriptive statistics tabulated by age

| Age groups            | 3-year | 4-year | 5-year | 6-year | Total | Reliability |
|-----------------------|--------|--------|--------|--------|-------|-------------|
|                      | $n = 30$ | $n = 47$ | $n = 54$ | $n = 35$ | $n = 166$ |
| Age (42–78 months)    | $M$    | $M$    | $M$    | $M$    | $M$    |             |
|                       | 45.60  | 54.70  | 64.89  | 74.63  | 60.57  |             |
|                       | $SD$   | $SD$   | $SD$   | $SD$   | $SD$   |             |
|                       | 1.54   | 3.65   | 3.29   | 2.02   | 10.35  |             |
| Picture sequencing (0–8) | $M$    | $M$    | $M$    | $M$    | $M$    | .79$^a$    |
|                       | 2.33   | 3.85   | 6.00   | 7.20   | 4.98   |             |
|                       | $SD$   | $SD$   | $SD$   | $SD$   | $SD$   |             |
|                       | 1.73   | 2.74   | 2.34   | 1.57   | 2.81   |             |
| Executive functions   |        |        |        |        |        |             |
| Stroop (0–8)          | $M$    | $M$    | $M$    | $M$    | $M$    | .90$^b$    |
|                       | 4.17   | 5.83   | 7.07   | 7.54   | 6.30   |             |
|                       | $SD$   | $SD$   | $SD$   | $SD$   | $SD$   |             |
|                       | 3.30   | 2.88   | 2.04   | 1.29   | 2.71   |             |
| DCCS (0–6)            | $M$    | $M$    | $M$    | $M$    | $M$    | .89$^b$    |
|                       | 2.67   | 3.60   | 4.91   | 5.23   | 4.20   |             |
|                       | $SD$   | $SD$   | $SD$   | $SD$   | $SD$   |             |
|                       | 2.50   | 2.40   | 1.78   | 1.61   | 2.28   |             |
| WM (0–15)             | $M$    | $M$    | $M$    | $M$    | $M$    | .77$^c$    |
|                       | 7.83   | 7.85   | 9.37   | 10.06  | 8.81   |             |
|                       | $SD$   | $SD$   | $SD$   | $SD$   | $SD$   |             |
|                       | 2.39   | 2.21   | 2.31   | 1.53   | 2.33   |             |
| FB (0–2)              | $M$    | $M$    | $M$    | $M$    | $M$    | .86$^a$    |
|                       | 0.17   | 0.94   | 1.44   | 1.66   | 1.11   |             |
|                       | $SD$   | $SD$   | $SD$   | $SD$   | $SD$   |             |
|                       | 0.53   | 0.89   | 0.82   | 0.73   | 0.93   |             |
| PVT (3–60)            | $M$    | $M$    | $M$    | $M$    | $M$    |             |
|                       | 12.40  | 16.15  | 23.81  | 31.09  | 21.11  |             |
|                       | $SD$   | $SD$   | $SD$   | $SD$   | $SD$   |             |
|                       | 7.87   | 7.21   | 10.38  | 10.71  | 11.31  |             |

Notes. Measurement ranges are indicated in parenthesis.

$^a$Reliability was calculated using Cronbach’s alpha.

$^b$The split-half reliability was calculated using the Spearman–Brown prophecy coefficient.

$^c$Reliability was calculated using Cronbach’s alpha, which was adjusted by excluding the sets of two- and seven-digit trials, whose variances were zero.

Table 2. Pearson’s correlation coefficients between the measurements and age

|               | 1  | 2  | 3  | 4  | 5  | 6  |
|---------------|---|----|----|----|----|----|
| 1. Age        |   |    |    |    |    |    |
| 2. PVT        | .62$^{**}$ |    |    |    |    |    |
| 3. WM         | .43$^{**}$ | .56$^{**}$ |    |    |    |    |
| 4. Stroop     | .45$^{**}$ | .49$^{**}$ | .37$^{**}$ |    |    |    |
| 5. DCCS       | .45$^{**}$ | .46$^{**}$ | .40$^{**}$ | .53$^{**}$ |    |    |
| 6. Picture sequencing | .66$^{**}$ | .60$^{**}$ | .44$^{**}$ | .48$^{**}$ | .44$^{**}$ |    |
| 7. FB         | .56$^{**}$ | .60$^{**}$ | .41$^{**}$ | .56$^{**}$ | .52$^{**}$ | .62$^{**}$ |

Note. **p < .01 (two-tailed).
AIC = 51.95. Adding the path from DCCS to the picture-sequencing task \((b = .07, \ p = .32)\) did not improve the model: \(\chi^2(2) = .97, \ p = .62, \ CFI = 1.00, \ AGFI = .98, \ RMSEA < .0001, \ 90\% \ CIs \ [0.00, 0.13] \ AIC = 52.97.\) Therefore, we decided to use the second model as the final model. The coefficients of parameter estimates are shown in Figure 3.

For picture-sequencing performance, in addition to age \((b = .43)\) and receptive language \((b = .25)\), Stroop \((b = .17)\) made a significant contribution, explaining 51% of the variance. This picture-sequencing task \((b = .28)\) made a significant contribution to FB, with additional unique and significant contributions from receptive language \((b = .19)\), Stroop \((b = .21)\), and DCCS \((b = .15)\), explaining a total of 54% of variance in FB. When the picture-sequencing scores for only the intentional trials were used, then similar parameter estimates were obtained, though the estimate of FB from age was approaching significance \((b = .15, \ p = .053)\), indicating good fit: \(\chi^2(3) = 1.94, \ p = .59, \ AGFI = .97, \ CFI = 1.00, \ RMSEA < .0001, \ AIC = 51.94.\)

**Discussion**

The present study tested a model in which the FB task is more cognitively demanding than the picture-sequencing task, allowing us to examine their associations with the three EF components. In this model, age, receptive language, and Stroop (inhibition) were significant contributors to the picture-sequencing task, explaining 51% of variance. However, DCCS (shifting) and WM (updating) did not make significant contributions to the picture-sequencing task. The picture-sequencing task subsequently contributed to the FB task, with independent contributions from EFs (Stroop and DCCS), age, and receptive language, accounting for 54% of the variance. These results suggest that FB performance could be explained by additional unique variance in DCCS and Stroop in addition to the

---

*Figure 3. The final path model with FB and picture sequencing as endogenous variables and EFs, language, and age as exogenous variables.*
picture-sequencing task; this supports the claim by Bloom and German (2000) that the FB task measures more than just mentalizing ability. Another way of interpreting the data would be that the FB task more strongly draws upon executive processes than does the picture-sequencing task.

For the EFs, the DCCS task was used to measure the flexible shifting of the focus of attention, whereas the Stroop task was used to measure inhibitory control. Additional variance in both inhibitory control and shifting of the focus of attention was needed to pass the FB task, suggesting that these two functions appeared to be key elements that are required to pass the FB task. Although the flexible shifting of the focus of attention as measured by the DCCS task seems to be critical for passing the FB task, it is not essential for the picture-sequencing task. In contrast, inhibitory control as measured by the pointing–Stroop task appeared to contribute to both tasks. Thus, some forms of inhibitory control are likely to be involved in responding to the general executive demands of both tasks. This interpretation accords with the claim that relationships between theory-of-mind and EF tasks are attributable to more than just a common problem of inhibitory control and that different aspects of EFs are needed to explain their developmental relationships (Perner et al., 2002). In the present study, Stroop and DCCS were administered using a touch-screen PC. Therefore, it is unlikely that differential effects of Stroop and DCCS on the FB and picture-sequencing tasks derived from any procedural differences.

The results also highlight that shifting ability as measured by DCCS is not associated with the picture-sequencing task but is associated with the FB task. It is possible to interpret this finding from the EFs required by each task. The mental set shifting in this task is also considered as cognitive flexibility, which incorporates an ability to change perspectives (Diamond, 2013). The FB tasks involve two competing representations of the child’s and the protagonist’s knowledge. To pass the task, children need to keep track of these representations and then inhibit their own representation to enable an attentional shift to the protagonist’s representation. However, to pass the picture-sequencing task, which has no competing representations, the children just need to arrange the events in order while inhibiting a desire to jump straight to the concluding picture. This interpretation resonates with the argument from the cognitive complexity control theory (David Zelazo, Jacques, Burack, & Frye, 2002), in that the higher order rules which are common to FB and DCCS can explain the association between EFs and FB, and therefore, the picture-sequencing task does not make as high demands as the FB task.

With respect to the 8-SIF (Achim et al., 2013), the FB task includes additional linguistic sources of information for the agent and for the context. Extra linguistic information included in the FB task may be helpful in guiding the participants to impute the agent’s mental states. However, this source of information could come at a cost because of the linguistic information about the changes in the context resulting from a translocation of the object. It could inadvertently direct the participants’ attention to the more salient (the true belief) location rather than the agent’s (false) belief, without a volitional flexible shift of attention. Thus, the participants may require an important shift in the focus of attention from the context they hear last. Unique variance in inhibitory control and shift of attention that help explain the FB scores in the model support this interpretation.

These findings suggest that passing the FB task demands both inhibitory control to suppress prepotent information and cognitive flexibility to shift one’s perspective to a different one. For this reason, compared with the picture-sequencing task, passing the FB task places demands on a wider range of EFs. The present findings also highlight the value of considering three components of EFs because the mentalizing tasks are indeed interwoven with different components of executive processing to different degrees. To
address the relationships between mentalizing ability and EFs, an analytical framework is needed to better understand the nature of the task especially to clarify the extent to which certain tasks impose executive demands.

Is it possible that the extra linguistic information given in the FB task could cause the participants to incorrectly impute the mental states of the agent? When the FB task is given without a verbal scenario to young children (Southgate, Senju, & Csibra, 2007), they performed well. In such a non-verbal FB task, the participants do not need to process linguistic information and therefore do not need to make an effortful shift in the focus of attention that may be required in the verbal FB task. Under such circumstances, the younger children were imputing the agent’s mental state without interference from linguistic information. When verbal processing is involved, however, the linguistic representation could make it harder for the participants to shift their focus of attention to the original (falsely believed) location. Previous findings from adult participants whose implicit FB inferences were negatively affected by just listening to unrelated linguistic information (Schneider et al., 2012) raise the possibility that even a low-level cognitive load could degrade a participants’ performance.

The emerging picture for the FB task has important implications for the interpretation of existing studies in regard to the development of mentalizing. EFs seem to play a critical role when mentalizing ability is measured using FB tasks, whereas EFs may be less important in processing other mentalizing tasks such as the picture-sequencing task. This notion becomes crucial when the expression–emergence hypothesis is discussed because many studies are based on the relational findings between mentalizing and EFs. More precisely, the source of information included in mentalizing seems to account for the possibility of finding a relationship between EFs and ToM. Carlson et al. (2015) found that including or excluding the aspect of FB did not change their relationships with EFs. Although cognitive load was manipulated, the tasks Carlson et al. used included both linguistic information and perceptual information, and are similar to each other based on the 8-SIF (Achim et al., 2013). This seems to suggest that analyses of the tasks in terms of the sources of information are important when making a relevant selection of mentalizing tasks.

In the selection of mentalizing tasks, researchers might also like to take into account the relative degree of demands placed by EFs, which emerged from the present comparison between FB and picture-sequencing tasks. Bloom and German (2000) suggest that the FB task can be used to explore the relative difficulty of reasoning representations. The present findings extend this further by enabling researchers to investigate which of the fundamental cognitive processes of EFs are in place. If children pass the FB task, then they are likely to have a wider range of EFs such as inhibitory control and cognitive flexibility than those who only pass the picture-sequencing task. Children who pass the picture-sequencing task, but not the FB task, are likely to have mentalizing ability, but not the full range of EFs needed to reason about mental representations in the more challenging format of the FB task.

This argument resonates with the present finding of a close association between the intentional and mechanical versions of the picture-sequencing task. From the original study (Baron-Cohen et al., 1986), intentional and causality versions of the picture-sequencing task were expected to elicit different results. However, the present study found that such differences in the contents (intentional/mechanical) may be very subtle when compared with the robustness of the executive processing required by the characteristics of the picture-sequencing task. Processing the sources of information in the picture-sequencing task is similar for intentional and mechanical versions, and differences in content had little impact on performance. Modelling executive
components in this present study has elucidated how different mentalizing tasks create differing executive demands. More generally, the results probably lean in favour of the expression account over the emergence account concerning how the development of EF relates to the development of ToM: While children’s ability on the FB task suggests a level of performance that is constrained by limitations in executive abilities, this was not so apparent for the picture sequence task. This finding raises the possibility that the FB task imposes executive demands that prevent children from expressing their ability to mentalize.

Conclusion
The FB task has been used in a large number of studies that investigate the development of ToM, and has almost become a signature task for the measurement of this ability. However, the path analyses in this study found that passing the FB task requires considerably more than just mentalizing ability, as claimed by Bloom and German (2000). Assessing the development of mentalizing abilities with a variety of mentalizing tasks, giving due consideration to the source of information, could provide a better approach to understanding the development of this ability.

Acknowledgement
This study was supported by the Japanese Society for the Promotion of Science: KAKENHI Grant No. 16K04327.

References
Achim, A. M., Guitton, M., Jackson, P. L., Boutin, A., & Monetta, L. (2013). On what ground do we mentalize? Characteristics of current tasks and sources of information that contribute to mentalizing judgments. Psychological Assessment, 25(1), 117. https://doi.org/10.1037/a0029137
Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind”? Cognition, 21(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8
Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and Intentional understanding of picture stories in autistic children. British Journal of Developmental Psychology, 4(2), 113–125. https://doi.org/10.1111/j.2044-835X.1986.tb01003.x
Berger, A., Jones, L., Rothbart, M., & Posner, M. (2000). Computerized games to study the development of attention in childhood. Behavior Research Methods, Instruments, & Computers, 32(2), 297–303. https://doi.org/10.3758/BF03207798
Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. Child Development, 81(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x
Blair, C. (2016). Developmental science and executive function. Current Directions in Psychological Science, 25(1), 3–7. https://doi.org/10.1177/0963721415622634
Bloom, P., & German, T. P. (2000). Two reasons to abandon the false belief task as a test of theory of mind. Cognition, 77(1), B25–B31. https://doi.org/10.1016/s0010-0277(00)00096-2
Carlson, S. M., Claxton, L. J., & Moses, L. J. (2015). The relation between executive function and theory of mind is more than skin deep. Journal of Cognition and Development, 16(1), 186–197. https://doi.org/10.1080/15248372.2013.824883
David Zelazo, P., Jacques, S., Burack, J. A., & Frye, D. (2002). The relation between theory of mind and rule use: Evidence from persons with autism-spectrum disorders. Infant and Child Development, 11(2), 171–195. https://doi.org/10.1002/(ISSN)1522-7219
Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development, 85*, 1777–1794. https://doi.org/10.1111/cdev.12237

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-145750

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7*(1–2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman assessment battery for children* (2nd ed.). Circle Pines, MN: American Guidance Service.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43*(3), 225–251. https://doi.org/10.1016/0010-0277(92)90013-8

Lewis, C., & Osborne, A. (1990). Three-year-olds' problems with false belief: Conceptual deficit or linguistic artifact? *Child Development, 61*(5), 1514–1519. https://doi.org/10.2307/1130760

McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*(1), 64.

Mitchell, P., & Lacohée, H. (1991). Children's early understanding of false belief. *Cognition, 39*(2), 107–127. https://doi.org/10.1016/0010-0277(91)90040-B

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howarter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Perner, J. (1991). *Understanding the representational mind*. Cambridge MA: MIT Press.

Perner, J., Lang, B., & Kloo, D. (2002). Theory of mind and self-control: More than a common problem of inhibition. *Child Development, 73*, 752–767. https://doi.org/10.1111/1467-8624.00436

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development, 13*(1), 73–90. https://doi.org/10.1016/S0885-2014(98)90021-1

Russell, J. (1996). *Agency: Its role in mental development*. Hove, UK: The Psychology Press.

Russell, J., Mauthner, N., Sharpe, S., & Tidswell, T. (1991). The ‘windows task’ as a measure of strategic deception in preschoolers and autistic subjects. *British Journal of Developmental Psychology, 9*(2), 331–349. https://doi.org/10.1111/j.2044-835X.1991.tb00881.x

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological Science, 23*(8), 842–847. https://doi.org/10.1177/0956797612439070

Siegal, M., & Beattie, K. (1991). Where to look first for children’s knowledge of false beliefs. *Cognition, 38*(1), 1–12. https://doi.org/10.1016/0010-0277(91)90020-5

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science, 17*, 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Ueno, K., Nakoshi, S., & Onuki, S. (2008). *Picture vocabulary test-revised*. Tokyo, Japan: Nihon Bunka Kagakusha.

van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2013). The structure of executive functions in children: A closer examination of inhibition, shifting, and updating. *British Journal of Developmental Psychology, 31*(1), 70–87. https://doi.org/10.1111/j.2044-835X.2012.02079.x

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655–684. https://doi.org/10.1111/1467-8624.00504
Supporting Information
The following supporting information may be found in the online edition of the article:

Table S1. Inter-item correlation matrix