This version of the article has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/ 10.1007/s11098-023-01986-8.

Title: A new rationalist account of the development of false-belief understanding

Author: Francesco Antilici, francesco.antilici@gmail.com

Abstract. Rationalist accounts of the development of folk-psychology maintain that the acquisition of this capacity is aided by special-purpose mechanisms rich in innate structure. Rationalists have typically maintained that false-belief understanding (FBU) emerges very early on, before the age of two. To explain why young children nonetheless fail the false-belief task, rationalists have suggested that they may have troubles expressing their FBU. Here I do two things. First, I argue that extant proposals about what might prevent children from expressing their FBU cannot explain some of the relevant data. Second, I put forward a new rationalist proposal, the processing-time account, according to which young children fail because they cannot carry out all the required processing in the time available. I argue that the processing-time account overcomes the challenges extant rationalist accounts face while being compatible with the evidence in their support, thereby providing a compelling explanation of the development of FBU.

1. Introduction

Folk-psychology is the capacity to infer what someone thinks, wants, or feels. To a creature lacking this capacity, the world must look rather different. Such a creature might perceive physical objects bumping against each other, some of them moving unpredictably and emitting strange sounds; but there would be no people, no actions, no reasons. It takes some effort to picture this. Yet, if knowledge of the mind is something we acquire through experience, at some point in our lives we must all have perceived the world that way. So, is that the case? Or do we come to the world already equipped to interpret behaviours as expressions of an inner mental realm? We can now tackle such questions experimentally, reigniting the classic debate between empiricism and rationalism (Chomsky, 1965; Margolis & Laurence, 2012).

Several psychological tasks have been used to investigate the development of folk-psychology, but none are as well-known and as widely used as the false-belief task (Wimmer & Perner, 1983). Its purpose is to establish whether children (or some other population of interest) are aware that people have beliefs, that beliefs guide action, and that beliefs are either true or false; in brief, whether they have False-Belief Understanding (FBU henceforth). In a classic version of the false-belief task (Baron-Cohen et al., 1985), children are told about Sally, who hides a marble in a basket; while Sally is away, a second girl, Anne, moves the marble to a box. Children are then asked, "Where will Sally look for her marble?" Since Sally still thinks, falsely, that the marble is in the basket, the right answer is that she will look for it there. The traditional finding, replicated hundreds of times, is that most 3-year-olds point to the box containing the marble, seemingly unaware of Sally's false belief, while most 5-year-olds point correctly (see Wellman et al., 2001, for a meta-analysis).

Many have taken this classic finding to settle the question: The average three-year-old, while perhaps not completely oblivious to the mental realm, has not yet learned that the mind represents, and sometimes *misrepresents*, the world (Gopnik & Wellman, 1992; Perner, 1991; Rakoczy, 2017). The relatively late emergence of this insight has been taken as evidence that FBU is acquired through learning thanks to general psychological mechanisms relatively unburned by innate structure. This leaves experience free rein in furnishing the mind, be it through cultural transmission (Heyes & Frith, 2014) or individual construction (Gopnik & Wellman, 2012). Following Chomsky (1965), I will refer to this as an empiricist account of the development of FBU.

Despite its prevalence and intuitive appeal, empiricism is not forced upon us. Several theorists have argued that two and three-year-olds already have FBU but struggle to express it in the false-belief task (Baillargeon et al., 2010; Bloom & German, 2000; Scholl & Leslie, 1999). As we will see in §2, there is evidence supporting this hypothesis. Strikingly, some studies suggest that infants only a few months old can already ascribe false beliefs (Hyde et al., 2018; Southgate & Vernetti, 2014)! If FBU emerges early on, this suggests that its acquisition is aided (and constrained) by special-purpose mechanisms rich in innate structure (Scholl & Leslie, 1999). This view has been called "nativist" (Margolis & Laurence, 2012) or "rationalist" (Chomsky, 1965); I will be using the latter term here.

All findings are open to alternative interpretations, and the evidence for rationalism is no exception. The debate is unlikely to be settled by any "critical" experiment. Both sides must explain away some of the data, and the view that does this best is likely to prevail. Rationalists must provide an account of what prevents young children from expressing their FBU in the original false-belief task. This has proven difficult. As I will argue in §3, the available proposals do not fit the data well. The worry is that this may tilt the scales in favour of empiricism.

My aim here is to help the rationalist meet this challenge. Drawing from the literature on cognitive-ability, processing-speed, and executive-functioning, I put forward a new proposal, the *processing-time account*: roughly, young children fail to express their FBU because they cannot carry out all the required processing in the time available. In §4 I introduce the account and discuss the evidence supporting it; in §5, I argue that it can explain the findings that challenge extant rationalist proposals as well as the evidence in their favour.

2. Evidence for rationalism

In this section, I summarise the evidence that FBU emerges earlier than traditionally assumed. First, let me introduce a distinction between two types of measures. The original false-belief task uses an "explicit" measure: children are simply asked where Sally will look, or what Sally thinks. Call this type of task *explicit-FB task*, and children's accuracy on such tasks *explicit-FB accuracy*. Other studies have used implicit or "nontraditional" (Scott & Baillargeon, 2017) measures, e.g., looking times or patterns of neural activation. Call these *implicit-FB tasks*.

Two main pieces of evidence support rationalism. First, minor procedural changes significantly, and in some cases dramatically, improve young children's performance on explicit-FB tasks (§2.1). Second, even infants can succeed at implicit-FB tasks (§2.2). As we will see, this evidence is not quite conclusive, but it is strong enough to warrant taking rationalism very seriously.

2.1 Explicit-FB tasks

The finding that three-year-olds fail explicit-FB tasks has been widely replicated and is robust in the face of many procedural variations (Wellman et al., 2001). Nonetheless, *some* procedural changes have been found to significantly improve children's explicit-FB accuracy. Such findings were often inspired by rationalist hypotheses about what could prevent children from expressing their FBU. There are two main rationalist proposals in the literature: processing-load accounts and pragmatic accounts. I will discuss the evidence supporting each in turn.

2.1.1 Processing-load accounts

According to processing-load theorists, explicit-FB tasks impose substantial demands on processing-resources scarce in young children (Carruthers, 2013; Leslie, 1994; Setoh et al., 2016). I will focus here on the influential account by Setoh et al. (2016). Setoh and colleagues argue that explicit-FB tasks impose demands on both response-inhibition and response-generation processes. First, the test question (e.g., "Where will Sally look for the marble?") triggers a prepotent response to answer based on one's own knowledge of where the marble is, which must be inhibited to succeed (response-inhibition). Second, children must interpret the test question and select the correct response (response-generation). Since young children have immature response-inhibition and response-generation processes, these processing-demands jointly overwhelm them, causing them to fail. Supporting this, procedural changes aimed at simultaneously reducing both types of demands have been found to dramatically improve performance.

First, young children respond more accurately in "unknown-location" false-belief tasks, where the marble is transferred to an undisclosed location or destroyed (Setoh et al., 2016; Wellman et al., 2001, p.

667; Wimmer & Perner, 1983). According to Setoh et al. (2016), this manipulation helps because, when children do not know where the marble is, they cannot point to it, i.e., there is no prepotent response to inhibit.

Second, Setoh et al. (2016) show that reducing response-generation demands further improves performance. Setoh and colleagues used an unknown-location task with two practice-trials. In each practice trial, children were shown pictures of two objects side-by-side and asked to point to one of them (e.g., “Where is the Frisbee?”). The rationale was that allowing children to rehearse answering questions in the practice phase would make that easier in the test phase, reducing response-generation demands. This proved enough to raise two-and-a-half-year-olds above-chance, an unprecedented result at the time. Importantly, this finding has since been replicated (Grosso et al., 2019; Kaltefleiter et al., 2021) and extended (Scott et al., 2020).

2.1.2 Pragmatic accounts

Pragmatic theorists argue that young children, due to their poor pragmatic-understanding, misunderstand the test question typically used in explicit-FB tasks (Helming et al., 2016; Westra, 2017; Westra & Carruthers, 2017). Pragmatic theorists have highlighted several ways this may happen.

First, children may take the experimenter to be asking where Sally will have to look for the marble to find it, as opposed to where Sally will look first. Pointing to the marble would then be the correct answer. To prevent this misunderstanding, Siegal and Beattie (1991) added a temporal marker to the question (e.g., “where will Sally look *first* for her marble?”) and found that this significantly improved three- and four-year-olds’ performance, a result later replicated by Surian and Leslie (1999) (see also Wellman et al., 2001, p. 668).

Second, children may take the question to concern the location of the marble. In this case, as well, it would be correct to point to the marble. Hansen (2010) rephrased the question to discourage this interpretation (e.g., “You and I know that the marble is in the box, where does Sally think it is?”), and found that this significantly improved three-year-olds’ performance. Furthermore, three-year-olds perform better when neither the test question (Rubio-Fernández & Geurts, 2013) nor the control question immediately preceding it (Rubio-Fernández & Geurts, 2016) mention the marble. As Helming et al. (2016) note, mentioning the marble may draw children’s attention to its current location, causing them to misunderstand the question.

(Note that attempts to replicate Rubio-Fernandez & Geurts’(2013) finding have had mixed results (Białecka-Pikul et al., 2019; Dörrenberg et al., 2019; Kammermeier & Paulus, 2018). For discussion, see (Paulus & Kammermeier, 2018) and (Rubio-Fernández, 2018a, 2018b)).

Third, children may take the question as a request to help Sally find the marble; this is plausible given that children are typically eager to help those in need. As Helming et al. (2016) note, this is supported by the finding that three-year-olds perform significantly better when the experimenter involves them in the change of location, with the explicit motive of deceiving Sally (Sullivan & Winner, 1993) (see also Wellman et al., 2001, p. 666).

2.2. Implicit-FB tasks

Converging evidence that FBU emerges early on has come from implicit-FB tasks. In a seminal study, Onishi & Baillargeon (2005) adapted the Sally-Anne story to the violation-of-expectation paradigm. An experimenter, playing the role of Sally, puts a toy watermelon slice inside a green box on the right, then leaves. While Sally is away, the toy spontaneously crawls out of the green box and into a yellow box on the left. In the test phase, half the infants are shown Sally reaching for the green box, while the other half are shown Sally reaching for the yellow box. Onishi & Baillargeon (2005) found that fifteen-month-olds looked significantly longer at the yellow-box event compared to the green-box event, suggesting they expected Sally to reach for the green box, where she thought her toy was. In a true-belief condition where Sally saw the toy crawling into the yellow box, however, infants now looked longer at the green-box event, suggesting that their expectations tracked Sally's beliefs.

Since the publication of Onishi & Baillargeon's (2005) ground-breaking finding, evidence that infants have FBU has continued to pile up, with more than thirty positive reports overall (Scott & Baillargeon, 2017). Positive findings have been obtained using a variety of implicit paradigms, including not only the violation-of-expectation method but anticipatory-looking, active-helping and neuroimaging paradigms as well. For example, a recent functional near-infrared spectroscopy study by Hyde, Simon, Ting & Nikolaeva (2018) found that the temporo-parietal junction, a key cortical area involved in FBU, was significantly more active as 7-month-old infants watched videos depicting false-belief scenarios compared to videos depicting true-belief scenarios.

Rationalists argue that implicit-FB tasks have lower performance-demands than explicit-FB tasks, explaining how infants can succeed at the former while failing the latter (Baillargeon et al., 2010). Implicit-FB results have proven particularly controversial, however. First, empiricists have argued that FBU is not required for succeeding at implicit-FB tasks, since infants may rely on other, less sophisticated abilities (Apperly & Butterfill, 2009; Heyes, 2014; Perner & Ruffman, 2005). Rationalists have criticised these alternative explanations, leading to a prolific debate that cannot be summarised here (Baillargeon et al., 2010; Carruthers, 2013). Second, against the backdrop of the replication crisis in psychology (Shrout & Rodgers, 2018), several failed replications of implicit-FB results have recently been published, further exacerbating the controversy (see Sabbagh & Paulus, 2018 and other articles in the same issue). An indepth discussion of failed replications is beyond the scope of this article, but a couple points are worth making.

In general, there are many reasons why a finding may fail to replicate, e.g., contextual factors (Van Bavel et al., 2016) or lack of statistical power (Maxwell et al., 2015). Concerning implicit-FB findings, many of the replication studies used different stimuli and procedures compared to the original studies; this, together with the fragility of the phenomenon in question, may account for some of the replication failures (see Baillargeon et al., 2018; Poulin-Dubois et al., 2018 for discussion).

Second, attempts to replicate implicit-FB findings have not had wholly negative results but mixed, with some paradigms (e.g., anticipatory-looking; see Barone et al., 2019 and Southgate's commentary in Baillargeon's et al., 2018) fairing worse than others. Those mixed results must then be weighed against more than 30 reports of positive findings (Baillargeon et al., 2018). As Barone and colleagues note in their meta-analysis of implicit-FB results, overall infants succeed more often than they fail, "suggesting that the tasks are tapping a real phenomenon" (Barone et al., 2019, p. 16). The failed replications certainly raise concerns and must be investigated further, but at this stage it seems an overreaction to dismiss the evidence that infants have FBU.

2.3 Summary

Taking a step back and looking at all the data discussed so far, there are clear indications that FBU emerges earlier than traditionally assumed. The evidence that subtle procedural changes can improve young children's explicit-FB accuracy is quite compelling, with some of the key findings (e.g., Setoh et al., 2016) already replicated (§2.1). The finding that infants can pass implicit-FB tasks is less robust, but nonetheless encouraging (§2.2). Overall, the case for rationalism may not be conclusive, but it is strong enough to warrant optimism about its prospects.

Unfortunately, the data discussed so far are only a subset of the evidence. There are many more findings, to be discussed presently, which challenge extant rationalist accounts. If the evidence for is not conclusive, and the evidence against cannot be explained away, rationalism is in dire straits. I will later suggest how rationalists can address this challenge.

3. Challenges to extant rationalist proposals

In this section, I explain some of the main challenges extant rationalist accounts face. I will discuss processing-load accounts first (§3.1), before turning to pragmatic accounts (§3.2).

3.1 Processing-load accounts

3.1.1 Executive-functioning

Processing-load accounts maintain that explicit-FB tasks impose demands on a processing-resource scarce in young children. What could that processing-resource be, however? Executive functions, such as inhibitory-control and working-memory, seem perfect candidates. Executive-functioning is notoriously poor in young children, and it is plausibly recruited in explicit-FB tasks. For example, Setoh et al. (2016) argue that, to succeed, children must inhibit a prepotent response, and that is precisely what inhibitory-control is for.

Supporting the claim that children's executive-functioning can affect their explicit-FB accuracy, dozens of studies have found a correlation between the two (see the meta-analysis by Devine & Hughes, 2014). This may seem to support processing-load accounts, but there is a problem. As Westra (2017) notes, the correlation, while robust, is relatively weak, suggesting that executive-functioning only plays a modest role. Indeed, in studies controlling for age and language-ability, the correlation appears to be only weak-to-moderate (r=.22) (Devine & Hughes, 2014, p. 1784).

Some processing-load theorists argue that executive-functioning is not the only performance-limiting resource; other processing-resources may thus account for the missing variance. But what could those resources be? Besides executive-functioning, Setoh et al. (2016, p. 5) mention language-ability and processing-speed as plausible candidates. Language-ability raises its own problems, to be discussed below (§3.1.2-3). The relationship between processing-speed and explicit-FB accuracy has not been investigated in children and, from the point of view of Setoh and colleagues' processing-load account, it is not clear *how* slow processing could cause children to fail explicit-FB tasks. With no evidence that processing-speed affects explicit-FB accuracy, and no explanation of how it could do so, the case does not look promising.

Notably, empiricists argue that variance in explicit-FB accuracy largely reflects variance in FBU. The fact that executive-functioning only accounts for a relatively small proportion of the variance thus plays in their favour. Some of the findings discussed below (§3.1.3) support this explanation. Furthermore, several empiricists endorse an "emergence" explanation of the correlation between executive-functioning and FBU, where executive-functioning assists children in learning that beliefs can be false, and there is evidence supporting this hypothesis (Benson et al., 2013; Sabbagh et al., 2006). The training study by Benson et al. (2013), for example, found that children with better inhibitory-control benefitted significantly more from training on false-belief scenarios.

I argue below (§4.2) that Setoh et al. (2016) were in fact right that processing-speed is an important performance variable; to this extent, their suggestion will be vindicated. Yet, the mechanism I will propose is not compatible with the processing-load framework. The take-home message, for now, is that processing-load theorists have yet to provide a satisfying explanation for the data on executive-functioning.

3.1.2 Language-ability

Given that explicit-FB tasks are verbal, it may seem plausible that young children's difficulties are due to their poor language-ability. Supporting this, dozens of studies have found that language-ability and explicit-FB accuracy also correlate (see the meta-analysis by Milligan et al., 2007). Even in this case, however, there is a problem. When administering explicit-FB tasks, experimenters typically include control questions to check that children understood, and remember, the key events in the story (Wimmer & Perner, 1983). Furthermore, two-and-a-half-year-old children have been shown to pass implicit-FB tasks with linguistic stimuli comparable to those in explicit-FB tasks (Scott et al., 2012). This suggests that most three-year-olds already have the level of language-ability required to interpret the relevant linguistic stimuli. And yet, language-ability continues to correlate with explicit-FB accuracy well after that age (Milligan et al., 2007). As above, empiricists have their own explanation. Children with higher language-ability can benefit more from social stimuli, such as exposure to mental-state talk, and thus have an easier time learning that beliefs can be false (Milligan et al., 2007). Indeed, there is evidence, to be discussed presently, that social stimuli help children acquire FBU.

3.1.3 Learning and maturation

Many studies have found that young children's explicit-FB accuracy correlates with social factors that could help them acquire FBU, such as exposure to mental-state talk and having a sibling (McAlister & Peterson, 2006; Ruffman et al., 2002; Slaughter, 2015). Furthermore, children who have undergone training on false-belief scenarios are more successful than those who have not (see the meta-analysis by Hofmann et al., 2016). Finally, in young children, explicit-FB accuracy correlates with maturational changes in the neural areas associated with FBU (Richardson et al., 2018; Wiesmann et al., 2017). Together with the data discussed above (§3.1.1-2), these findings appear to show that the age-related increase in explicit-FB accuracy tracks the emergence of FBU. Children who have been exposed to more social stimuli will have progressed more in their learning, resulting in a more mature FBU network and in better explicit-FB accuracy.

A processing-load theorist may try to explain away these data in terms of the mediating effect of language-ability and executive-functioning. Children with a sibling, for example, may have better language-ability and executive-functioning compared to only children; and we already know (§3.1.1-2) that children with better language-ability and executive-functioning have higher explicit-FB accuracy.

Unfortunately, several of the studies mentioned above have tested and ruled out this type of explanation. For example, the study by McAlister & Peterson (2006) found that, after controlling for both executive-functioning and language-ability, having a sibling accounted for *additional* variance in explicit-FB accuracy. Similarly, the correlation between explicit-FB accuracy and brain maturation persists after controlling for executive-functioning and language-ability (Wiesmann et al., 2017).

3.2 Pragmatic accounts

Having looked at some of the challenges processing-load accounts face, it is time to look at pragmatic accounts. Pragmatic accounts have important advantages. For example, a pragmatic theorist can argue that exposure to more social stimuli leads to higher pragmatic-understanding, and that children with higher pragmatic-understanding are more successful because they are less likely to misunderstand the test question (Westra, 2017). Several of the findings discussed above (§3.1.2-3) could be explained along similar lines (Westra & Carruthers, 2017). Nonetheless, pragmatic accounts also have important limitations.

3.2.1 Executive-functioning

If poor pragmatic-understanding is the sole factor preventing children from expressing their FBU, how do we explain why children with better executive-functioning tend to have higher explicit-FB accuracy? The challenge here is opposite to that faced by processing-load theorists. While processing-load theorists struggle to explain why the two variables do not correlate more strongly, pragmatic theorists struggle to explain why they correlate at all.

A potential solution to this problem is to go hybrid, i.e., to concede that poor pragmatic-understanding is not the sole factor preventing the expression of FBU, and that poor executive-functioning contributes as well. Several pragmatic theorists have pursued this strategy (Helming et al., 2016; Westra & Carruthers, 2017). This seems the right approach to take, but it needs to be fleshed out further. To make testable predictions, one must specify the role executive-functioning plays and the way it interacts with pragmatic-understanding in affecting accuracy.

Pragmatic theorists have made suggestions in this direction, but not without problems. To illustrate, Westra & Carruthers (2017) propose that succeeding at explicit-FB tasks requires not only interpreting the question as intended, but also inhibiting responses suggested by other interpretations. This suggests that different children can fail for different reasons: those with insufficient pragmatic-understanding will fail because they misinterpreted the question, while those with insufficient executive-functioning will fail because they could not inhibit incorrect responses. Although Westra & Carruthers (2017) do not acknowledge this explicitly, it seems that children would have to satisfy one requirement before they can satisfy the other. To inhibit incorrect responses, children must know what the incorrect responses are, which presupposes they interpreted the question as intended. E.g., a child who interprets the question as a request to help Sally would see the box containing the marble as the correct response; they would thus not attempt to inhibit it, even if they could. It follows that executive-functioning should make more of a difference to (and thus correlate more strongly with) explicit-FB accuracy in children who can interpret the question as intended compared to children who cannot; for the former would succeed or fail depending on their executive-functioning, whereas the latter would fail regardless. But on pragmatic-accounts, the proportion of children who can interpret the question as intended must increase with age; otherwise, how could pragmatic theorists explain how explicit-FB accuracy improves with age? Thus, we should expect the correlation between executive-functioning and explicit-FB accuracy to become stronger with age. Unfortunately, this prediction is not borne out by the evidence. The meta-analysis by Devine and Hughes, for example, found no change in the strength of the correlation between the ages of three and six (Devine & Hughes, 2014, p. 1785).

3.2.2 Further challenges

At least two further problems must be mentioned. First, there is evidence that young children already possess the level of pragmatic-understanding required to interpret the question as intended. Scott et al. (2012, exp. 2) carried out a violation-of-expectation false-belief task where two experimenters enacted the Sally-Anne story while a third experimenter (call her Jackie) watched in the back. After moving the marble in Sally's absence, Anne asked Jackie the usual test question (e.g., "Where is Sally going to look for her marble?"). The two-and-a-half-year-olds in this study correctly expected Jackie to point to where Sally thought the marble was. The question seems as open to alternative interpretations in this implicit task as in the original false-belief task. Participants could have taken Anne to be asking Jackie where the marble was, or to be encouraging her to help Sally, yet they did not.

Discussing this finding, Westra & Carruthers (2017, p. 171) note, "one might expect that the helpfulness-interpretation would be less salient because the infant herself is not involved in the task, and has no opportunity to help". Yet, what matters is not whether children can help, but whether Jackie could, since the question was directed at Jackie. It is possible, of course, that children interpret the question differently when targeted at another person; but the burden is on pragmatic theorists to argue that this is indeed the case. Perhaps, a pragmatic theorist could suggest that toddlers do not ascribe to Jackie a desire to help Sally, which would make the cooperative interpretation less salient for her[1]. However, there is evidence that infants expect agents to help those in need (Jin & Baillargeon, 2017; Köster et al., 2016; Lee et

[1] I thank one of the anonymous reviewers for this suggestion.

al., 2020). Strikingly, in these studies, infants were not given any *evidence* that the agent had a desire to help but made such an attribution by default. It seems plausible, then, that they would do the same with Jackie. It thus remains unclear whether pragmatic theorists can explain the finding at hand.

Second, remember the finding, discussed above (§2.1.1) that two-and-a-half-year-old children can succeed at an unknown-location task with response-generation practice-trials (Setoh et al., 2016). This is a key finding: no other task manipulation has obtained success with children so young using an explicit paradigm. Setoh et al. (2016) suggest that rehearsing response-generation processes in the practice phase reduces processing-demands during the test phase, but it is hard to make sense of this from a pragmatic perspective. To illustrate, consider Westra's (2017) pragmatic-development account. According to Westra, due to their unfamiliarity with the pragmatics of belief discourse, young children assume that beliefs are rarely a topic of conversation. This causes the intended interpretation of the test question, which concerns Sally's belief, to be less salient for them than alternative interpretations. To explain how Setoh and colleagues' practice-trials improve explicit-FB accuracy, then, one would need to show that the trials can help children realise that the experimenter is interested in Sally's belief. The problem, though, is that they seem rather ill-suited to that task. Children were shown pictures of two objects (e.g., a ball and a Frisbee) and asked a question about the location of one of the two ("Where is the Frisbee?"). It is rather hard to see how a child could gather, from being asked "Where is the Frisbee?", that the experimenter is interested in Sally's belief. Again, pragmatic-theorists need to do more to show they can explain this key finding.

3.2.3 Summary

Pragmatic accounts have important advantages over processing-load accounts, but also disadvantages. Although several pragmatic-theorists have conceded that executive-functioning can impact explicit-FB accuracy, current proposals on its role and interaction with pragmatic-understanding make predictions not supported by the evidence (§3.2.1). Furthermore, it remains unclear whether pragmatic theorists can explain other important findings (§3.2.2). This leaves us with a messy situation. As we saw in §2, much of the evidence for rationalism supports, more specifically, either processing-load or pragmatic accounts. Yet those proposals do not fit well with the rest of the data, putting a question mark on the plausibility of the rationalist position.

4. The processing-time account

In this section, I put forward a new rationalist proposal called the *processing-time account*, which I believe can address the challenges just outlined (§3). This new account fits neither the processing-load nor the pragmatic mould but tries to integrate the insights of both into a new theoretical framework.

4.1 The core hypothesis

To succeed at explicit-FB tasks, not only must children be able to perform all the required processing; they must be able to do so within the limited amount of time they have available. Among rationalists, the prevalent strategy has been to argue that young children, due to either insufficient processing-resources or poor pragmatic-understanding, fail to satisfy the first requirement. What I want to suggest is that they may satisfy the first requirement but not the second. On a first approximation, many young children can perform all the required processing; they just cannot do it *fast enough*. To ease the exposition, it will be useful to introduce some terminology:

- *Time Required for Processing (TRP)*: The time it takes participants to perform all the processing required for expressing their FBU.
- *Time Available for Processing (TAP)*: The time participants have available to perform their processing before responding.

With this terminology in place, the hypothesis can be restated as follows: many young children fail explicit-FB tasks because, for them, TRP is higher than TAP. Crucially, the relationship between TRP and TAP is assumed to vary with age. Thus, in three-year-olds, who tend to perform below-chance, TRP should exceed TAP, while in five-year-olds, who tend to perform above-chance, TRP should be lower than, or at least equal to, TAP.

Below, I will discuss some of the main factors affecting TRP (§4.2) and TAP (§4.3); along the way, I will discuss evidence supporting the account, which includes some of the problematic data discussed above (§3). The remaining challenges will be discussed in §5.

4.2 Factors affecting TRP

4.2.1 Processing-speed

In §3.1.1, while discussing Setoh et al.'s (2016) suggestion that children's processing-speed may affect explicit-FB accuracy, I pointed out that it was not clear how it could do so. The processing-time account provides the required mechanism. If the reason children fail is that they cannot complete the required processing in the time available, those that are faster should clearly be more likely to succeed.

This proposal takes inspiration from the literature on cognitive-ability. It is well-known that processing-speed, as measured on reaction-time or speeded tasks, correlates with cognitive-ability (Vernon, 1983). It is also well-known that, as people age, their cognitive-ability *and* their processing-speed simultaneously decline. To explain this phenomenon, Salthouse (1996) suggested that older adults, having slower processing, may be unable to carry out all the relevant operations in the time available. The hypothesis at hand is that something similar happens with young children on explicit-FB tasks. Crucially, young children are even slower than older adults (Kail, 1991).

Unfortunately, the relationship between processing-speed and explicit-FB accuracy in children has not been systematically investigated. Nonetheless, there is evidence that factors that are likely to affect processing-speed do correlate with explicit-FB accuracy. Ironically, this evidence has so far been taken to speak against rationalism.

4.2.2 Maturation

As mentioned in §3.1.3, children with a more mature FBU network are more likely to succeed at explicit-FB tasks (Richardson et al., 2018; Wiesmann et al., 2017). The authors of these studies have suggested an empiricist-friendly explanation: FBU is acquired through learning over the preschool years, and learning is associated in some way with the observed maturational changes. There is, however, an alternative: the observed maturational changes are associated with higher processing-speed, and children with higher processing-speed are better able to display the FBU they already possess, as suggested by the processing-time account.

Consider, for example, the study by Wiesmann et al., (2017). The key maturational change reported consists in an increase in the integrity of the white-matter underlying the FBU network. Now, processing-speed is thought to depend on neurophysiological properties, such as myelination and neurotransmitter efficiency, that affect nerve-conduction velocity and/or the speed of synaptic transmission, overall resulting in faster spread of neural impulses across the brain (Jensen, 1993). White-matter integrity is thought to affect nerve-conduction velocity and has repeatedly been found to correlate with processing-speed (Chopra et al., 2018; Turken et al., 2008). Notably, white-matter integrity is also associated with cognitive-ability, *and* processing-speed mediates this relationship (Ferrer et al., 2013; Kievit et al., 2016). It is plausible, then, that processing-speed also mediates the relationship between white-matter integrity and explicit-FB accuracy in children.

To see another example, consider the study by Richardson et al. (2018), who found that children who succeeded at an explicit-FB task had higher functional connectivity, i.e., higher correlation between activity in different areas of their FBU network. Higher functional connectivity can plausibly result from more efficient transmission of neural impulses between the relevant areas. Indeed, several studies have found that functional connectivity is positively associated with processing-speed (Krukow et al., 2018; Ruiz-Rizzo et al., 2019) as well as structural connectivity (Hagmann et al., 2010; Hermundstad et al., 2013).

4.2.3 Efficiency-of-processing and memory

It is common to draw a distinction between two factors that can affect reaction-times (e.g., Chi, 1977). Just as the time it takes to get from A to B depends on the speed one travels at but also on the route taken, the time it takes to perform a certain task will depend on one's basic processing-speed (i.e., the speed at which basic cognitive operations can be carried out) but also on the efficiency of the algorithm used (i.e., the number of basic cognitive operations required to perform the task). Both basic processing-speed (simply processing-speed henceforth) and efficiency-of-processing must thus be considered as determinants of TRP.

Memory processes are likely to have a significant impact on efficiency-of-processing. To succeed at explicit-FB tasks, one must hold the information about Sally's belief in mind as one processes the test question, then hold Sally's belief *and* one's interpretation of the question in mind as one plans how to respond. Explicit-FB tasks thus simultaneously engage both storage and processing components of working-memory, suggesting demands comparable to those imposed by complex-span tasks, which are not mastered until adolescence (Gathercole, 1999). Crucially, if any of the required information is lost, children will have to recalculate it, raising TRP substantially. Thus, immature memory processes can lead to more inefficient processing and thus to higher TRP. And indeed, children with better working-memory are more likely to succeed at explicit-FB tasks (Devine & Hughes, 2014). Furthermore, like processing-speed, working-memory improves steeply during childhood (Gathercole, 1999), which may contribute to explaining why explicit-FB accuracy improves with age.

4.2.4 Practice

It is well known that practising a task significantly reduces one's reaction-times on that task (Anderson et al., 1999; Newell & Rosenbloom, 1981). The mechanism underlying this effect is less certain, but there seems to be at least two mechanisms. Practice has been found to induce local increases in white-matter integrity, resulting in increased processing-speed (e.g., Takeuchi et al., 2010; Voelker et al., 2017). In addition, practice may shorten reactions-times through the adoption of more efficient computational strategies, requiring a smaller number of basic cognitive operations and thus increasing efficiency-of-processing (Kail, 1991). Either way, the finding that practice reduces reaction-times is well-established.

On the processing-time account, we would expect children who have had more opportunities to exercise the capacities recruited in explicit-FB tasks (FBU, but also language-ability, pragmatic-understanding, etc.) to have lower TRP, and thus to be more likely to succeed. There is evidence supporting this prediction; once again, the relevant findings have been taken to support empiricism. In §3.1.3, we saw that there are several social factors that can be used to predict children's explicit-FB accuracy, including: whether the child has a sibling (McAlister & Peterson, 2006); how frequent mental-state talk is in the family (Ruffman et al., 2002); and whether the child has received any training on false-belief scenarios (Hofmann et al., 2016). Empiricists argue that social stimuli enable learning, but an alternative possibility is that they enable *practice*, i.e., they allow children to rehearse the knowledge and cognitive capacities they have already acquired; practicing those capacities would then reduce TRP by boosting processing-speed and/or efficiency-of-processing as suggested above.

4.2.5 Summary

In §4.1, I suggested that young children may fail explicit-FB tasks because they cannot complete all the required processing in the time available, i.e., because TRP>TAP (§4.1). In §4.2, drawing from the literature on processing-speed and cognitive-ability, I identified two factors that are likely to affect TRP, processing-speed and efficiency-of-processing, as well as some of their determinants, including brain maturation (§4.2.2), memory (§4.2.3), and practice (§4.2.4). In those sections, I also reviewed evidence that these factors correlate with explicit-FB accuracy, as the account predicts: children with higher brain maturation, children with better working-memory, and children who have had more practice are more likely to succeed at explicit-FB tasks. Importantly, this evidence includes the “learning and maturation” data discussed in §3.1.3, which competing rationalist proposals struggle to explain.

As children grow older, their processing-speed and working-memory improve (Gathercole, 1999; Kail, 1991), and they accumulate more practice. As a result, older children should have lower TRP, which can explain why they have higher explicit-FB accuracy. This, however, is only half the story: it is time to look at the factors that can affect TAP.

4.3 Factors affecting TAP

4.3.1 What constrains TAP?

In explicit tasks we can distinguish between two types of time constraints, external (how much time the experimenter leaves participants to think) and internal (how much time participants are disposed to wait before responding). On the assumption that experimenters leave children enough time to think, performance will be affected mainly by internal constraints. The implications are worth teasing out. On the processing-time account, children fail because TRP>TAP. Based on the discussion so far, one may have assumed that TAP was an externally determined constant, with the accuracy of the response depending entirely on TRP. On the contrary, in explicit-FB tasks, TAP is likely to be determined by how much time children take to think about the answer, and this is just as important a factor as their speed. One can make it the case that TRP≤TAP by lowering TRP, but the same result can be obtained by raising TAP; thus, a child who failed could have succeed by being faster, *but also by taking more time*.

Importantly, we know that young children do not take the needed time on other tasks they struggle with. For example, on the day/night task (which consists in saying "night" when shown a card depicting the Sun, and "day" when shown a card depicting the Moon) Gerstadt, Hong & Diamond (1994) found that although three- and four-year-olds performed accurately when they took a relatively long time to respond, they rarely did so, even though "they were never rushed, did not know they were being timed, and were given as long as they wanted to respond" (as noted retrospectively by Diamond et al., 2002, p. 353). The same pattern has been observed on other tasks. On a computerised inhibitory-control battery, Davidson, Amso, Anderson & Diamond (2006a) found that while adults and older children took more time on difficult trials, four- and five-year-olds kept their reaction-times roughly consistent throughout, even though this negatively affected their accuracy. It has also been shown that forcing young children to wait[2] before responding, e.g., by singing a ditty after turning the card, significantly improves their accuracy on the day/night task (Diamond et al., 2002, p. 356), a finding that has been replicated several times (Montgomery & Fosco, 2012; Simpson & Riggs, 2007).

4.3.2 Inhibitory-control and motivation

What determines how long children wait before responding? Two factors seem relevant: ability to delay the response, and motivation to respond accurately.

First, as Davidson and colleagues note, young children may have "difficulty inhibiting impulsive responding, i.e., difficulty withholding their response long enough to take the time they really needed." (Davidson et al., 2006b, p. 2070). To flesh this out further, since figuring out the correct answer takes time and effort, children will be tempted to cut their processing short by providing premature responses. The slower children are, the stronger that temptation should be. At the same time, children's ability to resist it should depend on their inhibitory-control. Children with better inhibitory-control should be able to delay their response for longer, resulting in higher TAP. This fits well with the evidence that children with better inhibitory-control are more likely to succeed at explicit-FB tasks (Devine & Hughes, 2014), and that children perform worse after their inhibitory-control has been depleted in a previous task (Powell & Carey, 2017). Notably, like processing-speed and working-memory, inhibitory-control increases

[2] It is worth noting that simply telling children to wait is not effective. In the very first false-belief study, Wimmer & Perner (1983) immediately considered the possibility that children's difficulties may be due to premature responding. To test this hypothesis, Wimmer and Perner ran a "stop-and-think" condition where children were encouraged to "think carefully," and found that this did not improve children's performance. The attempt is commendable yet, in hindsight, its failure not surprising: "Asking 4-year-olds to wait before responding is a fruitless exercise" (Diamond et al., 2002, p. 353).

steeply during childhood (Gerstadt et al., 1994). Thus, older children should be not only less tempted to respond prematurely, but also better able to resist that temptation.

Second, to resist the temptation to respond prematurely, children must be not just able but also willing. This suggests a role for motivational factors. Children who lack sufficient motivation to answer correctly may not delay their response even when they could. Thus, TAP has not one determinant but two: inhibitory-control and motivation to answer correctly. That motivation does indeed play a role is supported by evidence that children's explicit-FB accuracy can be improved by incentivising correct answers through rewards (Peterson et al., 2013). Rewards should only be effective for children who *can* delay the response but decide not to; thus, such manipulations may be less effective in younger children, who have less inhibitory-control. Indeed, Peterson et al. (2013) found rewards to be effective in four-year-olds, but not three-year-olds.

4.3.3 Premature v. Prepotent responses

The model just sketched implies that, when children fail, they do so because, being either slow or unable/unwilling to delay the response (or both), they respond prematurely, i.e., before having completed their processing. Premature responses could potentially take many forms: random guessing, professing ignorance, or saying whatever comes to mind. The reason many three-year-olds provide reality-oriented responses (e.g., they point to the box containing the marble) is that such responses tend to have an air of plausibility that the alternatives (e.g., random guessing) lack. E.g., the marble Sally is looking for is in the box, so at first pass it "makes sense" that she would look for it there. Children can figure out that this is not the correct answer, but that might require more time than they are able/willing to wait.

Importantly, premature responses need not be prepotent. Processing-load theorists (e.g., Setoh et al., 2016) typically claim that the incorrect, reality-oriented response is prepotent: it directly competes with the correct response and will be selected unless inhibited. The processing-time account does not rely on this assumption. The reason reality-oriented responses are selected is not that they are prepotent, but that they are quicker and arise before the child has had time to figure out the correct response. Importantly, there are several ways for children to cut their processing short, all leading to reality-oriented responses. For example, knowing that the question contains the words "where" and "marble", children may be tempted to infer, rashly, that it concerns the actual location of the marble. Furthermore, when considering where Sally might look for her marble, children may be tempted to infer, again rashly, that she will look for it at its current location. It is only by resisting such temptations that children can take the time they need to figure out the correct answer. Once they have done that, however, premature responses should lose their appeal.

4.3.4 Summary

In §4.2, I argued that TRP depends on processing-speed and efficiency-of-processing, which are likely to improve with age. However, accuracy depends not just on TRP, but on TAP as well, specifically on whether TRP>TAP (§4.1).

In §4.3 I argued that, in explicit-FB tasks, TAP is likely to be determined internally; it is young children who do not take the time they would need to complete their processing. In §4.3.1, I presented evidence that young children do not take the needed time on other difficult tasks. In §4.3.2, I suggested that the time children take is determined by inhibitory-control and motivation for accuracy. In §4.3.2, I also discussed evidence that TAP is indeed a significant performance variable: children with higher inhibitory-control are more likely to succeed at explicit-FB tasks; children perform worse after their inhibitory-control has been depleted by a previous task; incentives increase accuracy in older children but not in younger ones. In §4.3.3, I clarified the role of premature responses and distinguished them from prepotent responses, thus highlighting an important difference between processing-load and processing-time accounts.

Importantly, like processing-speed and efficiency-of-processing, inhibitory-control improves as children grow older; consequently, older children should have not just lower TRP, but also higher TAP, being more likely to succeed as a result. This, I suggest, is the main mechanism explaining why older children have higher explicit-FB accuracy.

5. Overcoming the challenges

In §4, I put forward the processing-time account and highlighted some of the evidence supporting it, which includes some of the findings extant rationalist accounts struggle with. Specifically, the "learning and maturation" data discussed in §3.1.3, far from challenging the account, turned out support it. So far, then, the processing-time account seems to have the edge over competing explanations. However, two important questions remain. First, processing-load and pragmatic accounts each have evidence in their support, discussed §2.1. Is the processing-time account compatible with those findings? Second, although some of the challenges raised in §3 have been covered, others remain outstanding, particularly those concerning language-ability and executive-functioning. These issues will be discussed in turn.

5.1 Evidence for processing-load accounts

As processing-load accounts predict, young children perform better in unknown-location false-belief tasks, where the marble is destroyed or brought to an undisclosed location (§2.1.1). Response-generation practice trials further improve performance in unknown-location tasks, allowing even two-and-a-half-year-olds to succeed (Setoh et al., 2016), a finding that pragmatic accounts struggle to explain (§3.2.2). Is the processing-time account compatible with this evidence?

As explained above (§4.3.2-3), young children are tempted to answer prematurely. Although premature responses can take many forms, reality-oriented responses are particularly appealing, because they are plausible at first pass. This explains why three-year-olds tend to point to the box containing the marble. In unknown-location tasks, however, children do not know where the marble is, and thus cannot point to it. Children can still respond prematurely, but now their premature responses will likely manifest as guessing. Even if the same percentage of children fails to express FBU, the fact that they are now pointing randomly instead of incorrectly will, on its own, result in improved accuracy. Furthermore, if children care at all about answering correctly, they should be particularly wary of blind guessing; this may motivate them to wait a bit longer and, for some of them, this may be enough to succeed.

What about response-generation practice trials? These are likely to have helped by facilitating, through either priming or automatization, response-generation processes, including both processing the test question and selecting a response (Scott et al., 2020). Notably, although priming and automatization may result in lower processing-demands, they are also known to result in faster processing. Thus, although Setoh et al. (2016) interpret their results through the lenses of their processing-load account, they are just as compatible with the processing-time account.

5.2 Evidence for pragmatic accounts

As pragmatic accounts predict, discouraging alternative interpretations of the test question can improve young children's explicit-FB accuracy (§2.1.2). Yet, there is also evidence that two-and-a-half-year-olds can already correctly interpret the test question (Scott et al., 2012), another finding that pragmatic accounts struggle to explain (§3.2.2). How can we resolve this conflict?

On the processing-time account, most young children can already figure out the intended interpretation of the test question, yet doing so takes time, since alternative interpretations must first be ruled out, and many will struggle to complete the required processing in the time available. Discouraging alternative interpretations should reduce the amount of processing required, thus raising likelihood of success. Thus, pragmatic theorists are right that pragmatic-understanding can affect explicit-FB accuracy, but it does so by affecting TRP.

If this is the correct explanation, however, why was the amount of required processing not a problem in the study by Scott et al. (2012)? This is likely due to a difference in methodology. Scott et al. (2012) used an implicit, violation-of-expectation paradigm where toddlers passively watched the stimuli and were given up to sixty seconds to look at the test event, with no explicit task, while their looking times were being surreptitiously recorded. Since children had a lot more time to complete their processing compared to explicit-FB tasks, it is not surprising that many were able to succeed.

5.3 Language

Children with higher language-ability are more likely to succeed at explicit-FB tasks (§3.1.2). Yet, most young children already seem capable of interpreting the relevant linguistic stimuli (Scott et al., 2012). Again, these findings appear to be at odds with each other, yet the processing-time account resolves the conflict. Even if young children can process the linguistic stimuli in the task, those with higher language-ability should be able to do so more quickly. Language-ability has been found to correlate with speed on language-processing tasks in adults (Lewellen et al., 1993) as well as children (Fernald et al., 2006). Children who score higher on measures of vocabulary and grammar, for example, are typically found to be faster and more accurate at spoken word recognition (Donnelly & Kidd, 2020; Fernald et al., 2006).

In addition, remember that some empiricists (Milligan et al., 2007) have argued that language-ability mediates how much children can benefit from social stimuli, such as exposure to mental-state talk (§3.1.2). Notably, this applies regardless of whether social stimuli help by facilitating learning or through practice effects; thus, the processing-time account is consistent with this explanation.

5.4 Executive-functioning

In children, executive-functioning and explicit-FB accuracy are positively correlated (§3.1.1). This association is in the week-to-moderate range (Devine & Hughes, 2014). For processing-load theorists, the challenge is to explain why the two variables do not correlate more strongly, whereas for pragmatic theorists, it is to explain why they correlate at all. As noted above (§3.2.1), some pragmatic theorists have taken the right approach, claiming that executive-functioning is one of several factors that affect explicit-FB accuracy. Extant proposals of this kind, however, predict that the strength of the correlation should change with age, whereas it appears to remain constant (Devine & Hughes, 2014).

The processing-load account fixes the problem. First, executive-functioning affects explicit-FB accuracy by affecting TRP/TAP. Specifically, working-memory affects TRP (§4.2.4), while inhibitory-control affects TAP (§4.3.2). It is to be expected, then, that executive-functioning and explicit-FB accuracy should correlate. Second, executive-functioning is not the only factor affecting TRP/TAP; TRP can also be affected by processing-speed and practice (§4.2.1-4), while TAP can also be affected by motivation (§4.3.2). Since there are several other factors that can affect the relationship between TRP and TAP, a correlation in the weak-to-moderate range is not surprising. Finally, since all the factors affect explicit-FB accuracy simultaneously, and to an extent that does not change with age, we would expect the correlation coefficient to remain constant.

The processing-time account can also explain away evidence for "emergence" accounts of the correlation, on which executive-functioning helps children learn that beliefs are false (§3.1.1). First, Benson et al. (2013) found that children with better inhibitory-control benefitted significantly more from false-belief training. Benson and colleagues assume that false-belief training helps through learning, whereas I have suggested that it helps through practice (§4.2.4). But why would children with better inhibitory-control benefit more from such practice? The reason is that, being able to delay their response for longer, a smaller reduction in TRP would be required to improve their accuracy. Thus, although practice reduced TRP to the same extent across the high-inhibition and low-inhibition groups, only in the former was this sufficient for bringing TRP below threshold. Second, Sabbagh et al. (2006) found that Chinese children, despite having better inhibitory-control than American children, did not have higher explicit-FB accuracy. Sabbagh and colleagues argue that the Chinese children, not having any siblings, were less likely to be exposed to mental-state talk; this, they suggest, may have balanced out the advantage of having better inhibitory-control. Although Sabbagh and colleagues assume that mental-state talk helps children learn that beliefs can be false, it may instead help by shortening TRP through practice, as suggested above (§4.2.4).

6. Conclusions

My aim in this article has been to show that the data that extant rationalist accounts struggle to explain (i.e., those discussed in §3) can be made sense of from the perspective of a novel rationalist proposal, the processing-time account. Along the way, I also discussed additional data supporting this new proposal. At this point, a review may be helpful.

Children who, having more mature brains (§4.2.2), more efficient memory processes (§4.2.3), more opportunities to practice (§4.2.4), better language-ability (§5.3), can be expected to have lower TRP, tend to respond more accurately. So do children who, having higher inhibitory-control (§4.3.2), can be expected to have higher TAP. Interventions that lower TRP by allowing children to practice, such as false-belief training, improve accuracy (§4.2.4), while those that lower TAP by depleting inhibitory-control negatively affect it (§4.3.2). Task manipulations that lower TRP by reducing the amount of processing (§5.1-2) or increase TAP through incentives (§4.3.2) also improve accuracy. In addition, processing-speed, efficiency-of-processing and inhibitory-control all improve with age, driving TRP up and TAP down, thus providing a natural explanation for why accuracy improves with age. Finally, the account compares favourably with competing proposals, explaining away data that are problematic for other rationalist accounts (§3, §4.2.2-4, §5.3-4) while remaining compatible with the evidence in their favour (§5.1-2).

Should we conclude, then, that rationalism is true? Not yet. There remain important open questions I could not address fully here. For example, more work (both theoretical and experimental) is required to understand why implicit-FB findings have proven difficult to replicate. Only time will tell, but the prospects of rationalism should now look more hopeful, with some major obstacles removed.

References

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and Retention: A Unifying Analysis. *Journal of Experimental Psychology*, *25*(5), 1120–1136.

Apperly, I. A., & Butterfill, S. A. (2009). Do Humans Have Two Systems to Track Beliefs and Belief-Like States? *Psychological Review*, *116*(4), 953–970.

Baillargeon, R., Buttelmann, D., & Southgate, V. (2018). Invited Commentary: Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development*, *46*(June), 112–124.

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-Belief Understanding in Infants. *Trends in Cognitive Sciences*, *14*(3), 110–118.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Barone, P., Corradi, G., & Gomila, A. (2019). Infants' performance in spontaneous-response false belief tasks: A review and meta-analysis. *Infant Behavior and Development*, *57*(February), 101350.

Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D. (2013). Individual Differences in Executive Functioning Predict Preschoolers' Improvement from Theory-Of-Mind Training. *Developmental Psychology*, *49*(9), 1615–1627.

Białecka-Pikul, M., Kosno, M., Białek, A., & Szpak, M. (2019). Let's do it together! The role of interaction in false belief understanding. *Journal of Experimental Child Psychology*, *177*, 141–151. https://doi.org/10.1016/j.jecp.2018.07.018

Bloom, P., & German, T. P. (2000). *Two reasons to abandon the false belief task as a test of theory of mind*. *77*, 25–31.

Carruthers, P. (2013). Mindreading in infancy. *Mind and Language*, *28*(2), 141–172.

Chi, M. T. (1977). Age differences in the speed of processing: A critique. *Developmental Psychology*, *13*(5), 543–544. https://doi.org/10.1037//0012-1649.13.5.543

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Chopra, S., Shaw, M., Shaw, T., Sachdev, P. S., Anstey, K. J., & Cherbuin, N. (2018). More Highly Myelinated White Matter Tracts Are Associated with Faster Processing Speed in Healthy Adults. *NeuroImage*, *171*, 332–340.

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006a). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006b). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037–2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Devine, R. T., & Hughes, C. (2014). Relations Between False Belief Understanding and Executive Function in Early Childhood: A Meta-Analysis. *Child Development*, *85*(5), 1777–1994.

Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology*, *38*(3), 352–362.

Donnelly, S., & Kidd, E. (2020). Individual differences in lexical processing efficiency and vocabulary in toddlers: A longitudinal investigation. *Journal of Experimental Child Psychology*, *192*. https://doi.org/10.1016/j.jecp.2019.104781

Dörrenberg, S., Wenzel, L., Proft, M., Rakoczy, H., & Liszkowski, U. (2019). Reliability and generalizability of an acted-out false belief task in 3-year-olds. *Infant Behavior and Development*, *54*(March 2018), 13–21. https://doi.org/10.1016/j.infbeh.2018.11.005

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Ferrer, E., Whitaker, K. J., Steele, J. S., Green, C. T., Wendelken, C., & Bunge, S. A. (2013). White matter maturation supports the development of reasoning ability through its influence on processing speed. *Developmental Science*, *16*(6), 941–951. https://doi.org/10.1111/desc.12088

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410–419.

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: performance of children 33-7 years old on a Stroop-like day-night test. *Cognition*, *53*(1), 129–153.

Gopnik, A., & Wellman, H. M. (1992). Why the Child's Theory of Mind Really Is a Theory. *Mind & Language*, *7*(1–2), 145–171.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing Constructivism: Causal Models, Bayesian Learning Mechanisms, and the Theory Theory. *Psychological Bulletin*, *138*(6), 1085–1108.

Grosso, S. S., Schuwerk, T., Kaltefleiter, L. J., & Sodian, B. (2019). 33-month-old children succeed in a false belief task with reduced processing demands: A replication of Setoh et al. (2016). *Infant Behavior and Development*, *54*(August 2018), 151–155.

Hagmann, P., Sporns, O., Madan, N., Cammoun, L., Pienaar, R., Wedeen, V. J., Meuli, R., Thiran, J. P., & Grant, P. E. (2010). White matter maturation reshapes structural connectivity in the late developing human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(44), 19067–19072.

Hansen, M. B. (2010). If You Know Something, Say Something: Young Children's Problem with False Beliefs. *Frontiers in Psychology*, *1*.

Helming, K. A., Strickland, B., & Jacob, P. (2016). Solving the Puzzle About Early Belief-Ascription. *Mind & Language*, *31*(4), 438–469.

Hermundstad, A. M., Bassett, D. S., Brown, K. S., Aminoff, E. M., Clewett, D., Freeman, S., Frithsen, A., Johnson, A., Tipper, C. M., Miller, M. B., Grafton, S. T., & Carlson, J. M. (2013). Structural foundations of resting-state and task-based functional connectivity in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 6169–6174.

Heyes, C. M. (2014). False Belief in Infancy: a Fresh Look. *Developmental Science*, *17*(5), 647–659.

Heyes, C. M., & Frith, C. D. (2014). The Cultural Evolution of Mind Reading. *Science*, *344*(6190), 1243091.

Hofmann, S. G., Doan, S. N., Sprung, M., Wilson, A., Ebesutani, C., Andrews, L. A., Curtiss, J., & Harris, P. L. (2016). Training Children's Theory-Of-Mind: a Meta-Analysis of Controlled Studies. *Cognition*, *150*, 200–212.

Hyde, D. C., Simon, C. E., Ting, F., & Nikolaeva, J. (2018). Functional organization of the temporal-parietal junction for theory of mind in preverbal infants: A near-infrared spectroscopy study. *The Journal of Neuroscience*, *38*(18), 0264–17.

Jensen, A. R. (1993). Why Is Reaction Time Correlated With Psychometric g? *Current Directions in Psychological Science*, *2*(2), 53–56.

Jin, K., & Baillargeon, R. (2017). *Infants possess an abstract expectation of ingroup support*. *114*(31), 8199–8204. https://doi.org/10.1073/pnas.1706286114

Kail, R. V. (1991). Developmental Change in Speed of Processing During Childhood and Adolescence. *Psychological Bulletin*, *109*(3), 490–501.

Kaltefleiter, L. J., Sodian, B., Kristen-Antonow, S., Grosse Wiesmann, C., & Schuwerk, T. (2021). Does syntax play a role in Theory of Mind development before the age of 3 years? *Infant Behavior and Development*, *64*(May), 101575. https://doi.org/10.1016/j.infbeh.2021.101575

Kammermeier, M., & Paulus, M. (2018). Do action-based tasks evidence false-belief understanding in young children? *Cognitive Development*, *46*(December 2017), 31–39. https://doi.org/10.1016/j.cogdev.2017.11.004

Kievit, R. A., Davis, S. W., Griffiths, J., Correia, M. M., Cam-CAN, & Henson, R. N. (2016). A watershed model of individual differences in fluid intelligence. *Neuropsychologia*, *91*, 186–198.

Köster, M., Ohmer, X., Nguyen, T. D., & Kärtner, J. (2016). Infants Understand Others' Needs. *Source: Psychological Science*, *27*(4), 542–548. https://doi.org/10.1177/0956797615627426

Krukow, P., Jonak, K., Karakuła-Juchnowicza, H., Podkowińskid, A., Jonake, K., Borys, M., & Harciarek, M. (2018). Disturbed functional connectivity within the left prefrontal cortex and sensorimotor areas predicts impaired cognitive speed in patients with fi rst- episode schizophrenia. *Neuroimaging*, *275*(January), 28–35.

Lee, W., Kim, E. Y., & Song, H. joo. (2020). Do infants expect others to be helpful? *British Journal of Developmental Psychology*, *38*(3), 478–490. https://doi.org/10.1111/bjdp.12331

Leslie, A. M. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition*, *50*(1–3), 211–238. https://doi.org/10.1016/0010-0277(94)90029-9

Lewellen, M. J., Goldinger, S. D., Pisoni, D. B., & Greene, B. G. (1993). Lexical Familiarity and Processing Efficiency: Individual Differences in Naming, Lexical Decision, and Semantic Categorization. In *Journal of Experimental Psychology: General* (Vol. 122, Issue 3, pp. 316–330). https://doi.org/10.1037/0096-3445.122.3.316

Margolis, E., & Laurence, S. (2012). In Defense of Nativism. *Philosophical Studies*, *165*(2), 693–718.

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). *Is Psychology Suffering From a Replication Crisis? What Does 'Failure to Replicate' Really Mean?* https://doi.org/10.1037/a0039400

McAlister, A., & Peterson, C. C. (2006). Mental Playmates: Siblings, Executive Functioning and Theory of Mind. *British Journal of Developmental Psychology*, *24*(4), 733–751.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta analysis of the relation between language ability and falsebelief understanding. *Child Development*, *78*(2), 622646.

Montgomery, D. E., & Fosco, W. (2012). The Effect of Delayed Responding on Stroop-Like Task Performance Among Preschoolers The Effect of Delayed Responding on Stroop-Like Task Performance Among Preschoolers. *The Journal of Genetic Psychology*, *173*(2), 142–157.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of Skill Acquisition and the Law of Practice. In J. R. Anderson (Ed.), *Cognitive Skills and Their Acquisition* (pp. 1–55). Lawrence Erlbaum Associates.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255–258.

Paulus, M., & Kammermeier, M. (2018). How to deal with a failed replication of the Duplo task? A response to Rubio-Fernández (2019). *Cognitive Development*, *48*(July 2018), 217–218. https://doi.org/10.1016/j.cogdev.2018.07.005

Perner, J. (1991). *Understanding the Representational Mind*. MIT Press.

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: How deep? *Science*, *308*(5719), 214–216.

Peterson, C. C., Slaughter, V., Peterson, J., & Premack, D. (2013). Children with autism can track others' beliefs in a competitive game. *Developmental Science*, *16*(3), 443–450. https://doi.org/10.1111/desc.12040

Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., Krist, H., Kulke, L., Liszkowski, U., Low, J., Perner, J., Powell, L. J., Priewasser, B., Rafetseder, E., & Ruffman, T. (2018). Do infants understand false beliefs? We don't know yet – A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, *48*(November), 302–315.

Powell, L. J., & Carey, S. (2017). Executive Function Depletion in Children and Its Impact on Theory of Mind. *Cognition*, *164*, 150–162.

Rakoczy, H. (2017). In Defense of a Developmental Dogma: Children Acquire Propositional Attitude Folk Psychology Around Age 4. *Synthese*, *194*(3), 689–707.

Richardson, H., Lisandrelli, G., Riobueno-Naylor, A., & Saxe, R. (2018). Development of the social brain from age three to twelve years. *Nature Communications*, *9*(1).

Rubio-Fernández, P. (2018a). Trying to discredit the Duplo task with a partial replication: Reply to Paulus and Kammermeier (2018). *Cognitive Development*, *48*(July), 286–288. https://doi.org/10.1016/j.cogdev.2018.07.006

Rubio-Fernández, P. (2018b). What do failed (and successful) replications with the Duplo task show? *Cognitive Development*, *48*(January), 316–320. https://doi.org/10.1016/j.cogdev.2018.07.004

Rubio-Fernández, P., & Geurts, B. (2013). How to Pass the False-Belief Task Before Your Fourth Birthday. *Psychological Science*, *24*(1), 27–33.

Rubio-Fernández, P., & Geurts, B. (2016). Don't Mention the Marble! The Role of Attentional Processes in False-Belief Tasks. *Review of Philosophy and Psychology*, *7*(4), 835–850.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relationship between children's and mothers' mental state language and theory-of-mind understanding. *Child Development*, *73*(3), 734–751.

Ruiz-Rizzo, A. L., Sorg, C., Napiórkowski, N., Neitzel, J., Menegaux, A., Müller, H. J., Vangkilde, S., & Finke, K. (2019). Decreased cingulo-opercular network functional connectivity mediates the impact of aging on visual processing speed. *Neurobiology of Aging*, *73*(1), 50–60.

Sabbagh, M. A., & Paulus, M. (2018). Replication studies of implicit false belief with infants and toddlers. *Cognitive Development*, *46*, 1–3. https://doi.org/10.1016/j.cogdev.2018.07.003

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science*, *17*(1), 74–81.

Salthouse, T. A. (1996). The Processing-Speed Theory of Adult Age Differences in Cognition. *Psychological Review*, *103*(3), 403–428.

Scholl, B. J., & Leslie, A. M. (1999). Modularity, Development and 'Theory of Mind'. *Mind & Language*, *14*(1), 131–153.

Scott, R. M., & Baillargeon, R. (2017). Early False-Belief Understanding. *Trends in Cognitive Sciences*, *21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012

Scott, R. M., He, Z., Baillargeon, R., & Cummins, D. (2012). False-belief understanding in 2.5-year-olds: Evidence from two novel verbal spontaneous-response tasks. *Developmental Science*, *15*(2), 181–193.

Scott, R. M., Roby, E., & Setoh, P. (2020). 2.5-Year-Olds Succeed in Identity and Location Elicited-Response False-Belief Tasks With Adequate Response Practice. *Journal of Experimental Child Psychology*, *198*.

Setoh, P., Scott, R. M., & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences of United States of America*, *113*(47), 13360–13365.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology*, *69*, 487–510. https://doi.org/10.1146/annurev-psych-122216

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition*, *38*(1), 1–12.

Simpson, A., & Riggs, K. J. (2007). Under what conditions do young children have difficulty inhibiting manual actions? *Developmental Psychology*, *43*(2), 417–428.

Slaughter, V. (2015). Theory of Mind in Infants and Young Children: A Review. *Australian Psychologist*, *50*(3), 169–172. https://doi.org/10.1111/ap.12080

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130*(1), 1–10.

Sullivan, K., & Winner, E. (1993). Three-Year-Olds' Understanding of Mental States: the Influence of Trickery. *Journal of Experimental Child Psychology*, *56*(2), 135–148.

Surian, L., & Leslie, A. M. (1999). Competence and performance in false belief understanding: A comparison of autistic and normal 3-year-old children. *British Journal of Developmental Psychology*, *17*(1), 141–155.

Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., Yamanouchi, T., Suzuki, S., & Kawashima, R. (2010). Training of working memory impacts structural connectivity. *Journal of Neuroscience*, *30*(9), 3297–3303.

Turken, U., Whitfield-Gabrieli, S., Bammer, R., Baldo, J. V., Dronkers, N. F., & Gabrieli, J. D. E. (2008). Cognitive Processing Speed and the Structure of White Matter Pathways: Convergent Evidence from Normal Variation and Lesion Studies. *NeuroImage*, *42*(2), 1032–1044.

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(23), 6454–6459.

Vernon, P. A. (1983). *Speed of Information Processing and General Intelligence *. 70*, 53–70.

Voelker, P., Piscopo, D., Weible, A. P., Lynch, G., Rothbart, M. K., Posner, M. I., & Niell, C. M. (2017). How changes in white matter might underlie improved reaction time due to practice. In *Cognitive Neuroscience* (Vol. 8, Issue 2, pp. 112–118). Routledge.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-Analysis of Theory-of-Mind Development: The Truth about False Belief. *Child Development*, *72*(3), 655–684.

Westra, E. (2017). Pragmatic Development and the False Belief Task. *Review of Philosophy and Psychology*, *8*(2), 235–257.

Westra, E., & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, *158*(1), 165–176.

Wiesmann, C. G., Schreiber, J., Singer, T., Steinbeis, N., & Friederici, A. D. (2017). White matter maturation is associated with the emergence of Theory of Mind in early childhood. *Nature Communications*, *8*, 1–10.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.