The Case for a Dual-Process Theory of Transitive Reasoning

Barlow C. Wright

Department of Psychology, Brunel University of London, Uxbridge, UK

All correspondence and requests for reprints should be addressed to:-

B. Wright, Department of Psychology, Brunel University of London, Uxbridge, UB8 3PH,

England. E-mail: Barlow.wright@brunel.ac.uk.

Abstract

Ever since its popularisation by Piaget around 60 years ago, transitive reasoning (deductively-inferring A>C from premises A>B and B>C) has been of psychological interest both as a mental phenomenon and as a tool in areas of psychological discourse. However, the focus of interest in it has shifted periodically first from child development, to learning disability, to non-humans and currently to cognitive and clinical neuroscience. Crucially, such shifts have always been plagued by one core question - the question of which of two competing paradigms (Extensive-Training Paradigm v Non-Training Paradigm) is valid for assessing transitive reasoning as originally conceived in Piagetian research. The continued avoidance of this question potentially undermines several important findings recently reported: Such as about exactly what is involved in deducing transitive inferences, which brain regions are critical for reaching transitive inference, and what links exist between weakened deductive transitivity and mental illnesses like schizophrenia. Here, we offer the view that both of the competing paradigms are indexing transitivity, but each one tends to tap a different aspect of it. Then, we summarise studies from child and adult cognitive psychology, disabilities research, and from cognitive neuroscience. These, together with studies of non-human reasoning, seem to afford a theory of transitive reasoning that has two major components; one deductive but the other associative. It is proposed that only a Dual-Process Theory of transitivity (having analytic v intuitive routes approximate to deductive v associative processing respectively) can account both for the variety of findings and the apparently-disparate paradigms. However, Fuzzy-Trace Theory ("Gist" processes and representations), if not already embodying such a dual-process theory, will need to be incorporated into any complete theory.

Key Words: Children’s Reasoning; Deduction; Dual-Process Theory; Neuroscience; Non-Human Cognition; Transitive Inference

The Case for a Dual-Process Theory of Transitive Reasoning

The past 2 decades have shown us that there may be certain underlying commonalities in the way human thought is structured. One such commonality has become known as "Dual-Process Theory". The basic idea is that many domains of cognition and also social cognition are brought about because of the presence and interplay of two rather distinct functional processing systems (Amsel, Klaczynski et al., 2008; Evans, 2011). The dual-process perspective applies to our "developing abilities" as well as to our "developed abilities" (Ameel, Verschueren & Schaeken, 2007; Barrouillet, 2011; Bouwmeester & Sijtsma, 2007; Markovits & Thompson, 2008; Wright, Robertson & Hadfield, 2011). So far-reaching and persuasive is its scope, that Klaczynski (2009) could confidently state, whether we are considering cognitive or social functioning in children, we have reached the point where we must now accept that "development cannot be explained without recourse to Dual-Process theories." (Klaczynski, 2009, pp.286). This paper aims to show that this statement is as true for the reasoning domain of "Transitivity" as it is of other psychological domains.

On a generic dual-process view, we have evolved two processing-streams - in evolutionary terms an old system and a new system. The newer system is based around symbolic thought (e.g., language - Chapman, 1999; Wright, 2001), Working Memory (WM - De Neys & Dieussaert, 2005; Stanovich & West, 2000) and our highly developed attentional suppression system (e.g., inhibitory control - De Neys & Glumicic, 2008; Kavale, Forness & Lorsbach, 1995). Because of its later evolutionary arrival, this system tends to be called System-2, with the older system labelled System-1 (Stanovich & West, 2000). However, other conceptions of the two systems include Associative v Rule-Based, Heuristic v Systematic, Heuristic v Analytic; and most recently Type-1 v Type-2, or Experiential v Analytic (e.g., see Amsel et al., 2008; Evans, 2011; Klaczynski, 2009; Smith & Collins, 2009; Wright, 2001). Perhaps one distinction best captures all the others - Intuitive versus Analytic (Reyna & Brainerd, 2011). The analytic route plays a major role in planning and problem-solving activities, with the intuitive route playing a major role in guiding actions based on experience or basic associations between entities.

Individual theories tend to agree that the intuitive route is based on experiences built up over extended time, and operates effortlessly, also being contextually cued and relatively independent of one's cognitive ability (Morsanyi & Handley, 2008). It is largely instinctive, automatic and unconscious and makes use of parallel processing (Siemann & Delius, 1996). Resulting from the associative architecture of Long-Term Memory (LTM - Markovits & Thompson, 2008), this system is likely to be species-general (Vasconcelos, 2008; Wynne, 1998). The intuitive route may be inherently slow to build-up representations depending on the efficiencies of LTM, but nevertheless it is fast to respond (De Neys & Dieussaert, 2005).

The analytic route involves manipulation of representations and responses which are inherently more systematic but slow. It involves some degree of serial processing that is typically voluntary, wilful and conscious (Smith & Collins, 2009). Thus, it is largely this route that results in us being able to attend to, plan for, and carry out activities in systematically-ordered ways; abilities which are highly adaptive. Hence, this system is correlated with educational attainment scores, WM capacity and even with general intelligence (Amsel et al., 2008; Stanovich & West, 2000). However, the influence of such factors in system-2 is thought to be moderated by things such as personality dispositions (Stanovich, 2009). The analytic route tends to be relatively context independent but effortful. This allows it to dominate over the Associative System during abstract or hypothetical reasoning (De Neys & Dieussaert, 2005; Evans, 2011), although there is mounting evidence that it may itself often play only a relatively minor role in the actual computations that support the total reasoning process (Reyna & Brainerd, 2011).

Evans (2011) adds that system-1 (our intuitive route) has high capacity whilst system-2 (our analytic route) has limited capacity; and that system-1 processing is similar across individuals whilst system-2 processing exhibits most of our individual differences. Additionally, Evans observes that system-1 may well be unconscious, but we can still become conscious of its operation (e.g., reasoning about risk-taking – Reyna & Brainerd, 2011; Reyna & Farley, 2006). Equally, some advanced processes (e.g., reading) are evolutionary newer and hence should be defined as system-2; and yet they may operate both largely automatically and largely unconsciously just like a system-1 process. For these reasons, Evans has moved to thinking in terms of two categories of processes rather than just two distinct processes. These are Type-1 versus Type-2 processes; each type drawing on its own dynamic mix of system-1 and system-2 depending on the specific activity (e.g., reading, recognition, risk assessment etc.). This nuance serves to show that there are small but sometimes important differences between dual-process theories, although these differences are mostly beyond the scope of the present paper.

What is central is consideration of how the dual-process perspective can capture not only where development has to get to (Evans, 2011) but also the developmental course it takes (Barrouillet, 2011; Reyna & Brainerd, 2011). Regarding our two systems themselves, the above outlines imply that the intuitive route will tend to be efficient early on (Klaczynski, 2009), although it will also improve over extended time as children acquire more and more experience. By contrast, the analytic route is expected to become efficient relatively late, and reach its full potential relatively quickly after its initial arrival (for discussion see Markovits & Thompson, 2008; Morsanyi & Handley, 2008). On the latter issue, it has already been shown that analytic processes increase with age and cognitive development, and are also improved within any single child age group if the individual has higher than average cognitive capacity (e.g., higher I.Q or WM - Morsanyi & Handley, 2008). Thus, when we see performance on a task being effortful and varying with age, cognitive capacity and cognitive attainments, we may have good reason to hold that this task draws on the analytic route. Conversely, when we see performance on a task being effortless and remaining fairly stable or even reversing across large variations in these factors, we may have good reason to hold that this task employs the intuitive route (Bouwmeester & Sijtsma, 2007; Markovits & Thompson, 2008; Reyna & Brainerd, 2011).

As intimated above, the main aim of this paper is to make the case for a dual-process conception of the development to maturity of one particular reasoning domain, which is often taken to be basic to many others - Transitive Reasoning (Brainerd & Kingma, 1984; Piaget, 1954). This research area has recently attracted dual-process accounts (e.g., Ameel et al., 2007; Bouwmeester & Sijtsma, 2006; cf. Piaget, Grize, Szeminska &Vinh Bang, 1977; Wright & Howells, 2008), and yet the dominant view remains that transitive reasoning is one and only one process (Bryant, 1998; Vasconcelos, 2008). We therefore explore what single versus dual conceptions have to offer transitive research. The paper is organized around six main sections. The first presents the standard single-process view regarding the development of transitive reasoning. This presents two mutually-exclusive research traditions, one of which completely dominates the field (the Extensive-Training Paradigm). This is then followed by a long-overdue evaluation of the claim that the dominant paradigm really does fulfil the claims made by its originators and most of us today (i.e., to be indexing deductive transitive reasoning). The status of non-human transitive performance adds to the developing picture regarding this paradigm. An evaluation of the alternative (Non-Training Paradigm) is then presented. In the fifth section, both paradigms are evaluated in the context of brain research both with humans and non-humans. This leads to the premise of the sixth section, which is that these paradigms need to be integrated together if we are finally to arrive at an adequate theory of transitive reasoning spanning various child, adult and non-human groups. This section also considers whether one particularly well-grounded and well-specified theory (Fuzzy-Trace Theory) either can mitigate the need for a fully fledged dual-process theory of transitive reasoning, or already constitutes such a theory. These six sections are then drawn together into a clear set of conclusions, the most pertinent of which is that we really must have some version of a dual-process theory for transitive reasoning.

An Overview of Transitive Reasoning and its Development

Transitivity is an instance of relational reasoning, and is a mathematical property of an ordered set, as long as certain criteria are met (Wright, 2001). Take the set denoted by A, B and C respectively. Next, let B take on some value such that it lays somewhere between A and C in terms of a relation "r". The result is that we can now move from any item in the set to any other item in terms of r (ArBrC). As a concrete illustration, if person A is a faster runner than person B, and person B is a faster runner than person C, then a transitive inference permits us to deduce that person A is a faster runner than person C, without seeing A and C run against each other (Wright et al., 2011).

Many argue that the capacity for transitive inference is logical, partly because the conclusion necessarily follows as long as the relation used is a linear comparative term, and partly because we can deduce that conclusion upon simply being told the two premises, rather than having to directly perceive and verify them as true for ourselves. But for the founder of this view within psychology (Piaget, 1954, 1970), logic represents a description of the adult state of reasoning, in terms of axioms, rules and mathematical symbols. Today, most acknowledge that the actual psychological implementation of logic might conform to logical rules etc., but equally it might not (Barrouillet, 2011; Reyna & Brainerd, 2011; Wright, 2001). Thus, here, we use the term "deduction" to refer to the process of coordinating pieces of information in a rational and systematic way, and leave the question of whether this really does approximate to a logic to others. This said, we will substitute the term "logic" when this is essential to grasp an argument from any theorist.

A well-developed deductive transitive capacity has been shown to assist in the appreciation of things from the understanding of numerosity and measurement to inferences made during text comprehension, from class inclusion to assessing friendships, and from the understanding of Theory of Mind to balancing of risks/gambles (Barrouillet, 1996; Castle & Needham, 2007; Halford, Wilson & Phillips, 1998; Markovits & Dumas, 1999; Reyna & Farley, 2006; Smith & Collins, 2009). However, "although transitivity seems simple to adults, it causes difficulties for children under 5 years" (Andrews & Halford, 1998, pp.479-480).

Transitivity became of interest to psychologists largely because of the pivotal role it was afforded in Piagetian theory (e.g., Piaget, 1954; Piaget, Inhelder & Szeminska, 1960). The main research paradigm for investigating it was based around three-items and did not require any training. Tasks were loosely based on recommendations from theorists working in the area of reasoning or logic (Bara, Bucciarelli & Lombardo, 2001; cf. Clark, 1969; Goel, Makale & Grafman, 2004; Goodwin & Johnson-Laird, 2006). Thus, in line with the definition and examples given above, the minimum two premise pairs were used (A:B and B:C). Each premise contained a pair of items with one more favourable than the other along some dimension, usually "length" (see also Bara et al., 2001; Goodwin & Johnson-Laird, 2006; Markovits & Dumas, 1999). From the relationship between items within one premise pair (say items A and B) plus the relationship within a second pair (say items B and C), the reasoner was then required to make the inference between items A and C (Wright & Dowker, 2002; Wright et al., 2011).

Using such tasks among other tasks, Piaget and colleagues determined that transitivity develops quite gradually until it reaches adult competence (which they took to approximate to logic). Because children did not tend to get the A:C answer reliably correct until between 7 and 8 years, they maintained that our transitive competence is non-deductive before this time, but tends to involve deduction soon after. This said, Piaget and colleagues did concede children can make transitive inferences from as early as 3 or 4 years, although putting this down to learning by rote rather than logical deduction. Furthermore, Piaget (1954) viewed this lesser competence as intuitive reasoning or experiential reasoning, rather than deductive reasoning (one half of his preoperations stage was actually named the "intuitive" sub-stage). As acknowledged by Barrouillet (2011), this looks like an early example of a dual-process theory of transitivity within the domain of psychology, in all but name. This duality was crystallised by Piaget et al. (1977), into a theory distinguishing between "function logic" (the intuitive route) versus "operations logic" (the analytic route - see also Barrouillet, 2011; Chapman & Lindenberger, 1988; Reyna & Brainerd, 2011).

By contrast with Piagetian theory, almost the entire transitive field seems to endorse the view that children "are logical" by at least 4 years (e.g., Bryant, 1998; Chapell & Overton, 2002; Goodwin & Johnson-Laird, 2006; Oakhill, 1984; Pears & Bryant, 1990; Russell, McCormack, Robinson & Lillis, 1996). Indeed, even key theorists working with non-human groups begin by accepting this view (e.g., see McGonigle & Chalmers, 1992). The dominant view begins with the assertion that the "traditional Piagetian method of measuring transitivity in children is plagued by two great weaknesses" (Bryant, 1998, pp.266-267, emphasis added). These weaknesses are the failure to control for children's memory lapses (false negatives) and the failure to control for non-logical cues given by the experimenter (false positives). Such false-positives include provision of salient cues to absolute magnitude (e.g., knowing that item-A is exactly 25cm, item-B is 20cm and item-C is 15cm), and the parroting of labels in verbal tasks (e.g., by omitting B, only item-A had the label "large" when the choice is between only A and C).

The false-positive claim predicts that Piaget was too generous in attributing the beginnings of logic at 7 or 8 years. But this possibility seems yet to receive empirical evaluation. Nevertheless, to prevent false-positives (however unlikely they might be), Bryant and Trabasso (1971) simply boosted the number of items in the transitive series from three up to five items (the transitive series A>B>C>D>E is implied by premises A>B, B>C, C>D and D>E - see Bryant & Trabasso, 1971 experiment 1 for example of such graded feedback). This meant that none of item-B, C or D could be observed to be largest in magnitude, because each of these had been seen as of larger magnitude in one comparison but smaller in another (for discussion see Bryant, 1998; de Boysson-Bardies & O'Regan, 1973; Perner & Aebi, 1985; Perner & Mansbridge, 1983; Reyna & Brainerd, 1990; Wright, 2001). Additionally, the investigator could ignore premises and inferences containing either end-item (item-A and item-E), which they argued could be solved without the need to coordinate thinking around any common item (i.e., solved by using categorical labelling strategies). Hence the premises for deduction were now B:C and C:D, and the critical inference was B:D (Bryant, 1998).

One thing that this procedural change did was increase Bryant and Trabasso's own false-negatives: Remembering four interlinked premise pairs in the Bryant and Trabasso task is very much harder than remembering only two pairs in the classical task (Holcomb, Stromer & Mackay, 1997; cf. Perner & Mansbridge, 1983; Riley & Trabasso, 1974; Wright, 1998a; submitted). To get around this new problem, Bryant and Trabasso were forced to intensively train participants on all four premises (A:B, B:C, C:D and D:E) for as long as it took to achieve near perfect premise retention, before testing transitive relationships. This training regime is arguably the most critical aspect of the task (Markovits & Dumas, 1992; Reyna & Brainerd, 1990; Russell et al., 1996; Siegal, 2003), and so to keep it salient, Verweij, Sijtsma and Koops (1996), Wright and Howells (2008) and Yamazaki (2004) independently advise the term "the Extensive-Training Paradigm".

In two experiments reported in a 1971 paper, Bryant and Trabasso reported that children of 6 years, 5 years and even 4 years passed on the critical B:D inference at greater than 78% (>84% for experiment 2). They additionally reported that the critical B:D inference approximated the joint probability of its two antecedents (B:C and C:D). This confirmed Bryant and Trabasso's false-negative claim had been correct about Piaget's classical three-term task, although it seemed also to simultaneously uphold Piaget's intimation that the individual premises had to be coordinated in order to deduce the inference (Bryant, 1977, 1998; Reyna & Brainerd, 1990). On this latter issue, Reyna and Kiernan (1994, pp.179) also interpret Bryant and Trabasso (1971) as insisting that "memory for inputs (e.g., for presented sentences) must be preserved until reasoning can operate on them" (see also Trabasso, 1977 and see Reyna & Kiernan, 1994 for an empirically-supported opposing view).

Concerning the critical B:D inference, de Boysson-Bardies and O'Regan (1973) pointed out that if the experimenter/child dismisses items A and E because these have an unique label (e.g., "the big one" v "the small one"), then items B and D are now to some extent free to take on those nominal labels themselves. In other words, if Bryant and Trabasso's aim was to remove all possibility that B or D could be given labels which could in turn obviate the need to coordinate inferential responses around item-C, then at least on the theoretical level, that aim was not met. De Boysson-Bardies and O'Regan (1973) then went on to demonstrate that children do use their suggested categorical/nominal labelling strategies in many transitive-like tasks and even in a task closely resembling that of Bryant and Trabasso.

Although Bryant and Trabasso's B?D inference may initially have been the main phenomenon of interest, Trabasso (1977) and other theorists soon noted other interesting effects, such as end-anchoring, lexical marking and symbolic distance (for reviews see Allen, 2006; ; Wright, 2001, 2006; Wright & Howells, 2008; Wynne, 1998). For example, since the year 2000 alone, over 100 research papers have used the finding of a symbolic distance effect as the most direct evidence (i.e., the critical test) of genuine transitive deductions (Bond, Kamil & Balda, 2003; Browne et al., 2010; Frank, Rudy, Levy & O’Reilly, 2005; Titone et al., 2004; Vasconselos, 2008). Briefly, the symbolic distance effect refers to the profile of faster and more accurate transitive performance, the further apart the two items being compared (e.g., B:E better than B:D and A:E best of all - Moyer & Landauer, 1967; Trabasso, 1977). Interestingly, the actual taught premises are closest together and so these tend to be reported least well (Breslow, 1981). However, logicians hold that if an inference is logical then it can never be more certain than the information on which it is based (Clark, 1969). So we should already have reason to doubt the extensive-training paradigm routinely indexes logic when interpreted in terms of effects such as symbolic distance. Additionally, recall that Bryant and Trabasso (1971) had stipulated that only the B:D inference (in a five-term series) is sufficient for ascribing deduction to the participant. Hence, again any over-reliance on the symbolic distance effect, will overshadow the B:D inference, and therefore should already be suspected as not alluding to deduction (see next section on Trabasso, 1977).

The roughly 4 year age discrepancy between tasks using the extensive-training paradigm and those using the three-term paradigm is actually highly pertinent: It must be resolved if we are to reach a view on whether either paradigm used has been assessing its target deductive transitive ability (Bouwmeester & Sijtsma, 2004). Indeed, acknowledging this in their own comprehensive theory on transitive reasoning, Brainerd and Reyna state "the theory's overriding aim is to explain age variations in reasoning and memory on standard childhood cognition paradigms" (Brainerd & Reyna, 1990, pp.7). Making a start at resolving this age issue, Wright (2006a) previously noted that the critical aspect of the classical three-term task is not after all that it uses only three-terms, but actually that it avoids using prolonged training in learning the premises (hence the term "non-training paradigm"). Then as will be seen below, both three-term non-training tasks plus non-training tasks with five or more items, seem actually to support the view that transitive reasoning either develops over around 8 years (Andrews & Halford, 1998; Castle & Needham, 2007; Markovits & Dumas, 1999; Rabinowitz & Howe, 1994) or that it is not possessed by some children even between around 8 and 12 years (Bouwmeester & Sijtsma, 2007; Wright, 1998a).

Before progressing, it is imperative to note that, it appears that only one study (Bryant & Trabasso, 1971) has found 4 year-olds to be logical according to the extensive-training task. However, two non-training studies have also reported 4 year-olds are logical. In the first of these (Bryant & Kopytynska, 1976), 4 year-olds passed on a three-term task about transitivity of depth judgements, which was based directly on Piaget's famous "Two Towers Study" (Piaget et al., 1960). There are two things to note here. 1, Bryant's use of a three-term task is in conflict with his wider view that such tasks are plagued by great weaknesses (e.g., see Bryant, 1977, 1998; Bryant & Trabasso, 1971; Pears & Bryant, 1990). 2, When Bryant and Kopytynska themselves gave the same children a classical three-term task, they found that performance was even lower than did Piaget (i.e., 0% - all children failed). In the second study here (Pears & Bryant, 1990), transitive problems included three to six items. Items were blocks stacked one above the other; and children passed the transitive question, which this time was about which of two non-adjacent blocks would be higher up if they were in the same block pile (see Figure 1 for a three-item depiction of this task).

There is a problem regarding ascribing "logic" from either of these studies. Basically, the tasks allowed what Bryant had previously called "non-logical solution strategies", "non-transitive strategies", "categorical labelling" or "absolute cued solutions" (Bryant, 1977, 1998; Bryant & Trabasso, 1971). This occurred either deliberately (via colour cues - Bryant & Kopytynska, 1976) or inadvertently - (via absolute height cues - Pears & Bryant, 1990). When others have tested 4 year-olds on the block task after correcting for these cues (see Figure 1), in every case 4 year-olds did not perform at above chance (e.g., Markovits, Dumas & Malfait, 1995; Wright et al., 2011). The above reduces to two fears: 1, Although accepted as a robust paradigm, the extensive-training paradigm may actually be founded on only a single result regarding 4 year-olds. 2, That finding may not be robustly supported by any other paradigm. We believe the extensive-training paradigm needs to be properly and systematically evaluated, to uncover whether these fears are the reality, and whether this situation carries ramifications to its status as the gold standard for assessing transitive reasoning in other groups too.

Insert Figure 1 about here.

The following sections present empirical evidence regarding theoretical assertions and intimations about transitive reasoning, as conceived of as a deductive capacity (Bryant, 1998; Piaget et al., 1977). En-route we will outline some interesting findings and phenomena first stemming from the extensive-training paradigm and then from the alternative paradigm which avoids such training. It will be shown that the extensive-training paradigm tends not to readily index the deductive form of transitive reasoning, and hence the well cited conclusion that it disproves Piagetian theory or findings should now be seen as perhaps unsound, at least upon current evidence. We will also show that, by contrast, non-training tasks, although perhaps going some way beyond Piagetian theory in their designs, nevertheless tend to support Piagetian findings, most notably on age of competence in deductive transitive reasoning.

Studies from non-humans given extensive-training tasks and from human children and adults given both types of task, will be shown to offer additional support to the contention that, although clearly of relevance to transitivity, the extensive-training paradigm generally misses its target deductive transitive competence. Furthermore, we will support the view that transitive reasoning generally develops much as espoused by Piagetian theory. Then, building on Trabasso's (1977) theory and Breslow's (1981) theory surrounding extensive-training tasks and also on Piaget et al.'s variant on Piagetian theory, plus recent dual-process theories of transitivity (e.g., Ameel et al., 2007; Bouwmeester & Sijtsma, 2006), we will present a new theory. This is a new and comprehensive dual-process account of transitive reasoning that gives a place both to the deductive form and the non-deductive form of transitivity. The result will be a theory that accounts not only for developmental phenomena but also for all other groups to whom transitive tasks have been applied to up to now (including children and adults with learning difficulties, adults with clinical disorders and even research with animals). To date, we know of no post-Piagetian theory that has attempted all of this (although Halford et al.'s, 1998 theory comes closest to doing so - missing mainly the learning disabilities dimension). Within this dual-process conception, the extensive-training transitive phenomena reveal intuitive (non-deductive) reasoning and the non-training phenomena reveal analytic reasoning; mapping well onto the two processes in generalised dual-process theory as outlined at the start of this paper.

Finally, we will evaluate our own dual-process account of transitive reasoning relative to a very well specified existing account of transitivity (fuzzy-trace theory), which both has explained extensive-training task performance and generated new avenues for research for over 20 years (e.g., Brainerd & Kingma, 1984; Brainerd & Reyna, 2004; Reyna & Brainerd, 1990; Reyna & Ellis, 1994). We will first consider whether fuzzy-trace theory and our new dual-process theory might simply be different views of the same underlying theoretical perspective on transitivity; and then outline some findings yet to be accounted for by each theory and some promising potential future research avenues stemming from them both.

Evaluating a Purely Extensive-Training Account for Transitive Reasoning

Trabasso's lab conducted a large amount of research using the extensive-training paradigm between 1971 and 1977 (e.g., Lutkus & Trabasso, 1974; Riley & Trabasso, 1976; Trabasso, Riley & Wilson, 1975). This culminated in arguably the first highly detailed theory of transitive reasoning not only for children (Trabasso, 1975) but also encompassing adults (Trabasso, 1977). The theory explained transitivity in terms of Information-Processing constructs such as encoding, mentally-scanning and retrieval cuing; chief of which was "memory" (see Reyna & Brainerd, 1990; Wright, 2001 for detailed discussion). However, a second key development beyond the Bryant and Trabasso (1971) paper, was a reliance on response-time data (RT) in addition to response-accuracy data (Lutkus & Trabasso, 1974; Riley, 1976; Trabasso, 1977). Because these new tasks were based on RT, they could now be given via computerised tasks (e.g., Riley, 1976; Trabasso et al., 1975).

Summarising data from such tasks both with children and adults, Trabasso (1977) could explain transitive reasoning without recourse to logic at all. The main claim was that transitive phenomena result largely as a bi-product of premise memory - the better was memory, the better was the corresponding transitive inference. This aspect of Trabasso's theory was in line with assertions made in the Bryant and Trabasso 1971 paper (note, this memory-dependence view is not without challenge - e.g., Reyna & Brainerd, 1990). The basic thesis was that, during the training phase of extensive-training tasks, participants gradually construct an internal representation isomorphic with the entire logically implied series, by integrating the premise information during training (Halford & Andrews, 2004). This is done by first learning about the most positive item (A), followed by the least positive item (E), followed by the filling in of intermediate items in the same way (Trabasso, 1975, 1977).

In their 1971 paper, Bryant and Trabasso had argued that item-B and item-D were well suited for testing deduction precisely because neither item is an end item. However, as intimated above, de Boysson-Bardies and O'Regan (1973) demonstrated across four experiments, that for example, children tend to give item-B the same categorical label as item-A, apart from when it has to be considered against item-A. In other words, items one in from the end-items attract the same label as their respective end-item, but the label is qualified slightly (e.g., A = "positive", B = "quite positive", E = "negative", D = "quite negative" - Trabasso et al., 1975). De Boysson-Bardies and O'Regan had concluded that, since categorical labelling effects deemed by Bryant and Trabasso to invalidate Piagetian tasks, are still likely in Bryant and Trabasso's own tasks, extensive-training task conclusions (e.g., regarding age) are no more reliable than Piagetian conclusions. Notwithstanding de Boysson-Bardies and O'Regan's conclusion here, the process they offer for building-up the series actually provides a mechanism of great use within Trabasso's (1977) later theory (see Bryant, 1977 for independent application specific to the earlier Bryant & Trabasso account). It is worth noting two things at this juncture. First, although it is clear that de Boysson-Bardies and O'Regan's (1973) categorical strategy can clearly benefit humans, it is probably not available to non-humans (Premack, 2007). Second, we will see later that non-humans probably use a graded representation for the transitive array rather than a categorical one; but this raises the issue of why humans would need anything more than this graded strategy (Vasconcelos, 2008; cf. Wynne, 1998).

For now, we acknowledge that studies confirmed that in a five-term series, the end items (A and E) are indeed responded to even faster than any of the four premises which had actually been taught (A:B, B:C, C:D or D:E - Wright, 2001). Trabasso (1977) argued that, on a Piagetian view, the A:E inference should be inferior to even the critical B:D inference: For example, it might require three inferential steps - A:B + B:C gives A:C, then A:C + C:D gives A:D, and finally A:D + D:E gives A:E (of course many other permutations exist). Continuing this line of argument, the critical B:D inference should be easier because it can be made by combining only the B:C and C:D premises. Thus, Piagetian theory (or indeed any theory based on coordination of pairs of premises in a deductive manner - Trabasso, 1977) anticipates that inferences further apart should take longer and be less accurate (known as "the reverse symbolic distance effect" - see Barrouillet, 1996; Moyer & Landauer, 1967; Trabasso, 1975; Wright, 2001).

However, in at least one critical respect, the data from Trabasso and colleagues various studies refuted not only Piaget's account but indeed any account that was couched in terms of logical deduction, including the account that had been given by Bryant and Trabasso in 1971. On Trabasso's (1977) theory, the critical B:D inference should be superior to B:C and C:D. This directly contradicts the originally-reported Bryant and Trabasso finding that B:D was inferior to B:C and B:D (recall B:D approximated the joint probability of its two antecedents - Bryant & Trabasso, 1971). Trabasso (1977) could therefore strengthen the Bryant and Trabasso view that transitive reasoning was the result of memory phenomena rather than deduction; although he was now rejecting the view that the memory in question was memory for the premises logically required to make the inference (i.e., reasoners used memory for the global series rather than using memory for the individual premises - see Brainerd & Reyna, 1990; Reyna & Brainerd, 1990).

For Trabasso, the use of an extensive-training procedure remains more valid than the Piagetian tasks it had replaced in 1971. Extensive-training tasks result in the gradual and intuitive build-up of an integrated memory array that captures the global relations between the items learned. Thus, the array might end up being "A>B>C>D>E". Now, any premise can be reported by simply consulting the array. But crucially, any inference can also be reported from this same array. Take the B:D inference as the critical case: The participant accesses the series from the most positive end (A) and mentally moves along it serially until item-B is located. The participant then continues along the series from that point, until s/he has this time located item-D. Now the relative positions of these two items can be compared and the B:D inferential question answered, by simply comparing the locations of each in the global array (B= position 2; D = position 4; B is closer to position 1 so we report this item). Because this strategy might be applied starting from position 1 or position 5 in the array (i.e., the end-items), Riley and Trabasso (1974) and later on Trabasso (1977) called this an "ends-inward" mental search strategy. Also, it is because item-B and D are on average closer to the two ends than either B:C or C:D (i.e., would tend to be located sooner assuming searches start from the end nearest each respective item), that the B:D inference is reported faster and with less interference (i.e., more accurately) than its two antecedents (B:C and C:D). In the limit, the A:E inference involves no need to move along the array once the end items have been accessed (Trabasso quite reasonably assumes a self-terminating search through mental space); and this is why it is the fastest and most accurate inferential comparison to report.

Clearly Trabasso's theoretical account, although not involving logical deduction as such, can very neatly explain performance on extensive-training transitive tasks. However, Breslow (1981) pushed this theory to its limits in at least one crucial respect. He pointed out that the global array can be set up in mind without virtually any information about the comparative relations between items of each premise. The premises can be linked fully by using an intuitive process - what Breslow called "Sequential Contiguity". Assuming five entities are indexed by colour, if say "red" and "yellow" are shown together and then yellow is shown in a new pairing with say blue, the reasoner might note that yellow featured with red and with blue. But this is not the same as computing some transitive inference about these three items. In much the same way, we might intuitively link blue with say green and green with say brown. If these premises are presented to us enough times, especially when this is done sequentially (Bryant & Trabasso, 1971; Riley & Trabasso, 1974; Trabasso et al., 1975), a contiguity eventually emerges, because of the structure of LTM. Further repetitions of the premise information eventually leads to the common items of each successive premise becoming closer and closer in terms of its memory representation. With enough training, the desired integrated array begins to appear in mental space.

We can actually set up either of two interlinked series without necessarily needing to know which item had been the more positive in any of the individual pairs. Here the series would be either Red o yellow o blue o green o brown or else Brown o green o blue o yellow o red. Now, if we had noticed the positive-negative relation in only one of the premises, say that Brown should be selected over green, then we would move "from" brown "to" green in the array being passively built up in memory. Thus, without necessarily having any prior active intent, we would respond as though we know that, from most positive item to least positive item, the entire array will be

Brown>green>blue>yellow>red rather than the other way around. So, in tasks based on the extensive-training paradigm, mentally constructing the entire array does not even have to entail acknowledging the relationships within all of the premise pairs.

On either Breslow's or Trabasso's theory, clearly it remains the case that the use of highly extensive training is critical to mentally representing the transitive series, regardless of whether the reporting stage is as effortless as Trabasso (1977) maintained or as effortful as Breslow (1981) maintained. But it is important to note that both Trabasso's theory and Breslow's theory asserted that the transitive inferences revealed by the extensive-training paradigm, are not a matter of logical deduction at all. We will review evidence below, that suggests the reasoning process captured by the extensive-training paradigm tends to be one closely resembling intuitive processing as set out at the beginning of this paper.

Verweij et al. (1996) argue that, as revealing as the extensive-training paradigm has been regarding how transitive reasoning can operate, it is questionable as the tool for specifically "deductive" human transitive reasoning as was originally claimed by Bryant and Trabasso in 1971. This is because it is overly dependent on a training procedure ahead of testing for inferential capacity; a training procedure which may significantly change the nature of the task. This is even acknowledged by one of its originators (Trabasso). After twice failing to replicate the Bryant and Trabasso finding with 4 year-olds, Riley and Trabasso's (1974) third experiment had to use 4 days of training with 4 to 6 year-olds, had to train from largest to smallest and back again continually, and even then they had to additionally show all the items lined up together. As well as disqualifying the task as far as the need for logical coordination is concerned, it is important to note that all three manipulations had been absent in Bryant and Trabasso's study. However, only by going through such extreme lengths did Riley and Trabasso get their children to pass at levels even close to (i.e., within 10% of) those reported by Bryant and Trabasso. No surprise then that in Trabasso's own reflection about his unusual success in the 1971 paper, "we suspect that Bryant and Trabasso's (1971) inadvertent use of spatial cues was critical in aiding the children to construct ordered, spatial arrays for use in storing the ordered set information and answering transitive questions" (Riley & Trabasso, 1974, pp.196). Trabasso summarises, “we believe that, indeed, the children we have studied in these tasks do not use operational transitivity to solve the problem if one means by that term coordination of the members of the premises via a middle term” (Riley & Trabasso, 1974, pp.201). We will now outline empirical findings that these claims are correct and ought to be enough to encourage a shift in the dominant (extensive-training) account of deductive transitive reasoning and its development in humans.

As our first illustration, consider a study by Van der Lely (1997). Van der Lely contrasted 8, 9 and 10 year-olds' performance on various tasks (including a five-term extensive-training transitive inference task), as compared to a 10 year-old boy with specific language impairment. We focus here on the groups' performance. On the transitive task, the youngest control group (8 year-olds) performed at 92% across all 10 premise and inferential pairs. However, they also out-performed a second control group of 10 year-olds in terms of overall transitive performance. It is important to note that standard tests confirmed that the older children were 2 years advanced of the younger children on WM tasks, other cognitive tasks and language tasks. Clearly, extensive-training task performance was not related to the children's intellectual abilities. Earlier, we noted this as an indication of intuitive processing. One could argue that the 8 year-olds may simply have tended towards a different strategy to the 10 year-olds - Van der Lely did confirm the 8 year-olds were trained for a longer time. But that only re-emphasises our point here - intuitive processes are the ones that are slow to build-up. If the 8 year-olds' strategy was more induced because of being given more training, it is even more unlikely to have been based on deduction than the 10 year-olds (Trabasso, 1977; Riley & Trabasso, 1974).

Russell et al. (1996) anticipated and tried to circumvent such issues, conceding some extensive-training tasks do tap an associative competence. But they argued the ones used by Bryant and Trabasso exhibit "logical necessity" and therefore test the desired deductive competence (see Markovits & Dumas, 1992 and Von Fersen, Wynne, Delius & Staddon, 1991 for a fuller debate and theoretical treatment of this issue). They were responding to over 20 years of transitive research which claimed that other species were just as logical as human children (e.g., Davis, 1992; McGonigle & Chalmers, 1992; Wynne, 1998); and made it clear in their paper that their main aim was to rule out all tasks with non-humans as invalid.

Russell et al. offered empirical data to support this view. They tested two groups of 6.5 year-olds using separate training and test sessions. One group did a graded task following Bryant and Trabasso's length task (experiment 1, 1971). The other group did a nominal task following McGonigle and Chalmers' weight task for monkeys (1992). The investigators then waited between 7 and 12 minutes after completing extensive-training on the graded length task, so that premise memory would fall to that of the nominal task. Furthermore, they filled this delay with a task designed to occupy WM, before even testing that particular group. This ensured the premises and the implied transitive series were absent from WM (see Halford, Maybery & Bain, 1986) and instead were stored in LTM (Brainerd & Reyna, 2004). It was found that children's transitive performance was better on the graded than on the nominal task. This, according to Russell et al. (1996), begs the conclusion that the nominal task must have been harder because it had to be solved "non-logically" (associatively), whereas the graded task must have been easier because it demanded a logical (deductive) solution (i.e., exhibited logical necessity).

However, it would have been more parsimonious to conclude that it was the harder task that drew on deductive competencies (i.e., the nominal task). There were two additional potential issues: First, both of the Russell et al. tasks, rather than just the graded task, had essentially been used in the Bryant and Trabasso study (the McGonigle & Chalmers task was actually a replication of Bryant & Trabasso's, 1971 nominal task, experiment 2). In that experiment, the children never saw any length gradations at all, and so were presented only with binomial information (here it was verbally), just as in McGonigle and Chalmers (1992). In their own words, "'big' and 'little' were the terms used with the youngest group as they were more easily understood." (Bryant & Trabasso, 1971, pp. 457). Second, Russell et al. did not acknowledge that deductive coordination of premises must rely on WM (see Barrouillet, 1996; Crone et al., 2009; Halford et al., 1986), and this is why they had gone through considerable lengths to ensure that the information for their "graded task" was NOT in WM but rather in LTM. De Neys and Dieussaert (2005) found direct experimental evidence that WM capacity does correlate with deductive (analytical) solutions but it does not correlate with solutions that can be reached without the need for analytical thinking. Finally, it is worth noting that Russell et al.'s 6 and 7 year-olds did not approach the level of performance by even Bryant and Trabasso's 4 year-olds.

Despite Russell et al. not replicating Bryant and Trabasso's findings with 4 year-olds, other extensive-training tasks have resulted in children younger than Russell et al.'s 6.5 year-olds doing even better than suggested so far. In one variant, Holcomb et al. (1997) showed that 5 year-olds can do well on transitive tasks. But here it must be noted that it took one 15 minute session per day for between 24 and 57 days of training, before Holcomb et al. got their 5 year-olds to pass at reliably above chance (see also Kallio, 1982; Riley & Trabasso, 1974; Stromer, Mackay, Cohen & Stoddard, 1993 for confirmation that training is never ever "rapid" and can take hours, days or even weeks with child groups).

In another variant on the extensive-training paradigm, Stromer et al. (1993) successfully trained a 5 year-old child and two adults with mental ages similar to their child participant on not one but two transitive tasks, each involving a five-term series (which we denote series A and series B). Participants gave the usual symbolic distance profile for each series separately (for details see Breslow, 1981; Trabasso, 1977). However, in addition all participants also gave meaningful responses on comparison pairs that were formed by displaying one item from series A as a pair with one item from series B. Stromer et al. (1993) concluded that participants were responding in terms of the unique ordinal position of items within individual series - what Riley and Trabasso (1974) had earlier called "spatial cues" leading to "ordered spatial arrays". Indeed, Stromer et al. found within-series performance was at above 90% for inferential pairs but performance for analogous pairs in the cross-series comparisons was even higher at virtually 100%. Although an explanation in terms of comparisons between different ordinal positions in the different series is almost certainly correct, as far as logical deduction is concerned, there was no basis for reaching any single conclusion with the cross-series comparison pairs (nor did these researchers claim that there was). In Bryant and Trabasso's terminology, the Stromer et al. (1993) study confirms that the extensive-training paradigm is capable of precisely the "false-positives" it was designed to avoid.

Even tasks with typically developing adults suggest the extensive-training paradigm tends not to be about deduction. For example, data in Titone et al. (2004) show that their control participants actually improved their performance from testing immediately following completion of training to a further test session, with no intervening training between these two test sessions. Indeed, establishment of the transitive series in memory is improved even further by allowing the participant time enough to sleep before the transitive test session is run (Ellenbogen, Hu, Payne, Titone & Walker, 2007). This means that by not engaging in reasoning (i.e., sleeping) participants can reason better on extensive-training tasks. The implication is that extensive training calls into play an LTM consolidation mechanism that facilitates representation of the entire transitive series. Ellenbogen et al. also found that a particular area of the brain is important for such transitive responding (the hippocampus - see later).

The nature of the consolidation process on extensive-training tasks is so pervasive that the reasoner does not even need to know s/he is building up a representation of the entire series in memory and simply reading off answers from that representation (Wright, 1998a; 2006b). For example, Siemann and Delius (1996) gave adults a computer-game transitive task which used pairs of polygons to infer the six-term series +A:B, +B:C, +C:D, +D:E and +E:F (the + here denotes the item selected from each taught pair following Bryant & Trabasso's 1971 experiment 2). After extensive training, participants learned the series and gave transitive responses during test. A post-test questionnaire revealed that most participants were unaware of the implied series early on in training but became more aware as training progressed. Importantly, however, Siemann and Delius found that on completion of training, participants who remained unaware of the series performed just as well as those who had become aware (see also Libben & Titone, 2008; Wright, 1998a). Moses, Villate and Ryan (2006) gave adults another six-term transitive task in which participants reported post-test, about the block of training trials where they became aware of the series. Upon comparing this to the block in which transitive performance became reliable, it was found that participants improve performance first and only become aware of the transitive hierarchy later on. Such findings support the conclusion reached by Reyna and Brainerd from their own studies, which was that the extensive-training paradigm "can not only alter reliance on memory, but the nature of reasoning itself" (Reyna & Brainerd, 1990, pp.42). Indeed, on the view that logic at the psychological level is about wilful coordination of premise information (Bara et al., 2001; Bryant, 1977; Clark, 1969; Goodwin & Johnson-Laird, 2006; Piaget, 1954; Piaget et al., 1977), few would expect that participants can be logical but not conscious of the use of apparent-coordination of premises.

We have seen above that it is likely that neither children nor adults tend to rely on deduction, when giving their responses to transitive questions following highly extensive training. We have also learned that one of the originators of the extensive-training paradigm as we know it today, had actually already acknowledged this (Trabasso, 1977). However, Trabasso had additionally claimed that Piaget was mistaken to hold that children's transitive reasoning is ever logical (but contrast Trabasso, van den Broek & Suh, 1989 on this issue). But it is highly important to make clear two things here. First, although Bryant continued to claim transitivity is logical but Trabasso maintained it is never logical, both theorists did continue to insist that there is one and only one transitive capacity (contrast Trabasso, 1977 v Bryant, 1977). Second, there cannot be many problem-solving instances in real life when an individual would get between around 50 and 500 exposures of a small number of premise items (neither adults nor children), followed by the insistence that all transitive responses involving those items must be made via deduction. Indeed, one has to ask oneself just how valuable to everyday reasoning or the great technological human advancements such a gradual and time-intensive transitive competence would be? We will see below that such a competence will have its uses, but those uses are not necessarily about problem-solving and do not seem to call for deduction.

Transitivity in Non-Humans

As intimated above, soon after Bryant and Trabasso's original demonstration with young children, it was realised that the extensive-training paradigm was potentially of use in assessing whether other species also "are logical" (for discussion see McGonigle & Chalmers, 1992; Wynne, 1998). Allen (2006) points out that, particularly for non-humans, great evolutionary importance may be attached to being able to place oneself along a graded continuum such as the social rank of maybe 80 potential competitors. Transitivity is an ideal way of embodying this skill. Indeed, in non-humans, transitivity is said to underpin phenomena ranging from efficient food-foraging behaviour, to establishment and maintenance of social hierarchies (Wynne, 1998).

MacLean, Merritt and Brannon (2008) cite evidence in support of their speculation that transitivity probably developed because of social pressures faced by individuals living in groups. In groups, social hierarchies invariably develop and it is advantageous for the individual to be capable of determining its place in important hierarchies (Cheney & Seyfarth, 1990). In such a group, it both is unlikely and un-adaptive for an individual to observe all the possible pair-wise interactions between all other individuals, plus interactions involving itself, and then store all of these as separate unconnected instances. It is far more adaptive for the individual to be able to construct the fully connected series from relatively few piecemeal (e.g., pair-wise) instances. To put this another way, any species that arrived at this ability for whatever reason, would now be able to begin to form stable social groups and readily insert itself into such groups. Indeed, this is commensurate with the minimisation of personal risk from one's peers and hence self-preservation. The suggestion here is that transitivity, therefore, must be a highly important basic capacity for all social beings.

In support of this speculation, a plethora of studies have clearly intimated that with the right kind of training and enough of that training, virtually no participant group fails to demonstrate transitivity when tested using extensive-training tasks. Species demonstrating it include monkeys, elephants, lemurs, rats, mice, hooded crows, pigeons, sea lions, honey bees and more (Couvillon & Bitterman, 1986; Daisley et al., 2009; Davis, 1992; DeVito, Lykken, Kanter & Eichanbaum, 2010; Lazareva, Smirnova, Zorina & Rayvesky, 2001; MacLean et al., 2008; McGonigle & Chalmers, 1992; Wittemyer & Getz, 2007; Wynne, 1998). For instance, males of a small colourful species of fish called the African Burtoni fish have been demonstrated to make transitive responses (Grosenick, Clement & Fernald, 2007). Even infant non-humans reason transitively (e.g., baby chicks). Indeed, the 1 day-old domestic chick not only has transitivity, but also has it lateralised in a human-like way to the right hemisphere (Daisley et al., 2009). Here, training was via auditory and visual signals from the mother hen and the relational comparative was food-location preferences along a linear transitive dimension. Humans also have transitivity lateralised to the right hemisphere, so long as the task is not a fully verbal one and the participant is typically-developing (e.g., does not have schizophrenia - Hanlon et al., 2011).

We note that, although Daisley et al. conclude that their respective participant groups clearly demonstrated "transitive responding" (Siemann & Delius, 1996; Rabinowitz & Howe, 1994), they avoid any claim to have demonstrated deductive transitive reasoning. However, those theorists who follow Bryant and Trabasso's (1971) claims that the extensive-training paradigm is the best index of deduction, would seem to have to concede infant chicks as having a deductive transitive competence (but see Markovits & Dumas, 1992; Russell et al., 1996).

Continuing research with birds, other comparative theorists have compared the performance of two sub-species of the same species in a transitive inference task. Consistent with the social intelligence hypothesis we intimated at the start of this section, Bond et al. (2003) have reported that with enough training, both the less social species of jay (Western Scrub-Jays) and the much more social species of jay (Pinyon-Jays) demonstrate the same high level of transitive responding. The tasks used arbitrary visual stimuli presented using extensive training and a five term series. Bond et al. also found that, when training was less optimal (their experiment 1), the Pinion Jays demonstrated a much better transitive capacity than did Scrub Jays. Just as Russell et al. (1996) had done for children, they concluded that one species of jays solved their transitive tasks associatively whilst the other solved it using the same (higher level deductive) ability humans have. The same dissociation of solution strategies was found in a more recent study with lemurs (MacLean et al., 2008). The more social Ringtailed Lemurs perform better on a five-term transitive task than did the less social Mongoose Lemurs. However, when training was made more appropriate (following proportional feedback as introduced by Perner & Aebi, 1985), the Mongoose Lemurs performed at the same level as the Ringtailed Lemurs. Again this was interpreted as showing that both species have equal transitive capacities but the Ringtailed Lemur applies this capacity more readily; probably because of living in more socialising groups (i.e., large social groups with roughly linear dominance hierarchies).

The implication to what Russell et al. (1996) term the logical necessity of the generic extensive-training paradigm should by now be clear - one cannot simply use the fact that a participant group passes an extensive-training transitive task as proof that the group did so via deduction. Indeed, Trabasso's (1977) theory held that transitive inferences are always reached primarily because of "memory" rather than because of "deduction". But for further clarity regarding subsequent claims about the deployment of deduction, let us compare the actual pass rates in some well-known studies of human and non-human groups tested. To aid this, Figure 2 shows the relative performance of five different participant groups as assessed and published independently of each other. One anomaly present in Figure 2 is that adults' performance can be lower than both 4 year-olds and 6 year-olds. Clearly this cannot be right. It also shows 6 year-olds can do worse than 4 year-olds, which is also questionable. But the five studies also show two other anomalies: Non-human deductive transitive performance (here Squirrel Monkeys) can exceed the performance of typically-developing child and adult human groups. But are we really to conclude that non-humans may be more logical than humans? Finally, Figure 2 shows that the group providing the highest performance on extensive-training task has actually been humans who have very serious learning disabilities and behavioural limitations (Maydak, Stromer, Mackay & Stoddard, 1995). Here, the conclusion would have to be that deductive transitivity is superior in humans with learning difficulties compared to those with no such difficulties; and yet learning difficulties, almost by definition involve a deficit in deductive capacities.

Insert Figure 2 about here.

Our own conclusions from comparing the five studies of Figure 2 really must be that it is highly dubious to make any claims about the logical and/or deductive capacities of any participant group using only the generic extensive-training task. For non-humans, it seems it is possible to respond transitively but this does not mean animals are doing so using logic, deduction or rationality (contrast McGonigle & Chalmers, 1992). But how would such a non-logical mechanism work?

Recall that Trabasso (1977) denied that the extensive-training paradigm needs to be logical in order to account for human transitive reasoning. Recall also that Trabasso's theory captured the relative performances on inferential comparisons compared to their antecedents (the premise pairs). Regarding the first issue, some comparative investigators (e.g., Wynne, 1998) accept Bryant's (1977, 1998) view that transitivity is a logical entity; but simultaneously accept Trabasso's (1977) view that the mechanisms for it lay in "associative memory" rather than in "logical mechanisms" (cf. Von Fersen et al., 1991). Matters of logic are a specifically human phenomenon yet to be applied to non-humans (Premack, 2007). Hence, by claiming that apparently-logical performance can be explained without recourse to actual logical mechanisms, Wynne was opening up the way for a quite compelling argument that transitive reasoning is the same in humans and non-humans. Regarding the second issue, comparative investigators then began to produce models which in some sense went beyond those posited for humans (for discussion see Vasconcelos, 2008). For example, these animal models could actually "predict" the levels of performance on specific trained (premise) versus un-trained (inferential) pair-wise comparisons, as opposed to Trabasso's theory which was claimed to merely capture relative (rather than absolute) performance (see previous section).

It should be mentioned at this point, that Vasconcelos misattributes regarding Trabasso's theory. Contrary to Vasconcelos, Trabasso did actually produce not one but two predictive models alongside the relative models he proposed to explain data from his lab (see especially Trabasso, 1975, 1977; Trabasso et al., 1975). Indeed, at least one of the models subsequently proposed by Wynne (1998) was actually already proposed in Trabasso's writings.

Wynne (1998) did discuss a larger number of competing mathematical models of transitive responding. At their essence was the notion of an associative mechanism, reducing to LTM as discussed earlier. The transitive responses arrive out of this mechanism because of a value transfer rule (cf. Trabasso, 1975): If a stimulus is rewarded (e.g., item-A in the context of B) then its value is increased relative to the other stimuli (principally item-B). After enough training, a representation of the worth to the participant of all the stimuli in the series is built up in memory. The participant can then respond to any inferential question by simply consulting this integrated representation. The form of the access may be absolute or relative position (e.g., in animals and young humans - Russell et al., 1996; Vasconcelos, 2008), or might just as well be a voluntary mental scan through all items from both ends (e.g., in cognitively well-developed humans - Breslow, 1981; Wright, 2006b).

Wynne's theory is more about eventual series representation rather than the capacity that actually integrated the premise information to form that representation. But what if “the process of constructing the ordered set representation is an important part of the reasoning process, because it is there that the transitivity principle has to be applied” (Halford & Andrews, 2004, pp.126)? The implication both for human and non-human performance would be either that the extensive-training paradigm tends to assess transitive responses at a point in time well after any deductive mechanism would have been invoked and has become now unnecessary (Chapman, 1999; Wright, 2001, 2006b), or else there really is no logical capacity to explain at all (Trabasso, 1977; Vasconcelos, 2008). We will show that the first possibility seems true about humans, with the second possibility seeming true of non-humans (Premack, 2007).

To illustrate indexing deduction too late, in a human sample, consider a study by Wright (2006b). Transitive performance was measured all the way through each of five increments of training on a five-term extensive-training task. Incremented training provides a set of training pairs (premises) and then a large number of test pairs, followed by training pairs, and so on. The test profiles indicated premise acquisition and inferential responses were initially achieved by an associative process. For example, this was pivoted at the small end of the series, as item-E is most unique in associative terms (it is the only one never ever reinforced - Vasconcelos, 2008; Wynne, 1998). Indeed, in the first two increments, the B:D response was better than the B:C and C:D trained pairs (Breslow, 1981). But by three increments of training, both transitive and indeed verbatim performance was now consistent with the engagement of deduction (e.g., the B:D inference was now less certain than its antecedents - see Trabasso, 1977). Finally, by five increments of training, performance was based on a voluntary search through a spatial representation that had now been built up in memory (agreeing with Trabasso, 1977). Now, the B:D inference was again better than its antecedents and the series was pivoted at the large end rather than the small end.

In an additional study it was shown that a further two increments of training produces performance at ceiling (e.g., involving an even stronger symbolic distance effect - Wright, 1998a). But the focus here is on the fact that the transitive responses after five increments of training were not indexing actual deduction, because deduction had already come and gone well before the fifth increment. Wright (2006b) argues that this set of findings shows that, if we know we are going to need the same answer over and over again, it is actually "more logical" to store the answer than to deduce it anew every time we need it (i.e., prolonged extensive-training renders deduction increasingly redundant). Indeed, if deduction was the aim, then performance would have to be taken from the third increment rather than on conclusion of all training.

Our evaluation here is contested by some researchers. For example, Van Elzakker, O'Reilly and Rudy (2003) argue that non-humans can sometimes use a logical mode (they call this top-down processing). In order to test this, they used a six-term series rather than the more usual five-term series. It was argued that a five-term series confounds logical and associative solutions, but a six-term series does not. This is because in a six-term series, items-B through E (i.e., all apart from the end items A and F) are presented as more valuable (>) and less valuable (<) with equal frequency and so a logical solution strategy should lead to identical performance on say B:D compared to say B:E. But for a six-term series, B:E involves two items very near to the absolute ends and so this should attract higher performance. B:D involves the D item which is nearer the centre of the series and so will be linked less with the low value end of the series; hence it is more confusable with item-B (see also D'Amato & Colombo, 1990 and compare Breslow, 1981). Van Elzakker et al. (2003) observed that rats did better on B:E than on B:D, which they took to indicate participants were using value-transfer rather than deduction (note, this is a symbolic distance effect). But this response profile is not confined to non-humans. Using a six-term implied series, human adults have also been reported to use value-transfer when tested after extensive training (Frank et al., 2005). Deductive transitivity then, at least when assessed using extensive-training tasks is probably not species general, and may even be questioned in human adults just as Trabasso (1977) had theorised.

Another reason for doubting the claim that deductive transitive reasoning, as targeted by most researchers using extensieve-training tasks, both is concerned with deduction and is species general, is our preoccupation with the symbolic distance effect as the hallmark of transitive reasoning (Breslow, 1981; Trabasso, 1977; Wynne, 1998; Van Elzakker et al., 2003). For example, Browne et al. (2010) concluded they had observed the same transitivity as in humans, based on a symbolic distance effect both for response accuracy (housing preferences for their non-human participants - adult hens) and for RT (Response Times for how fast hens made their choice of housing). Indeed, when the symbolic distance effect is not precisely enough modelled, some theorists question the theoretical framework itself. For example, questioning theories such as Halford et al. (1998) and Trabasso (1977), Vasconcelos (2008) states "the cognitive models have serious difficulties... the way they incorporate the serial position effect, the SDE [Symbolic Distance Effect] and error in the typical TI tests seriously undermines their theoretical robustness" (Vasconcelos, 2008, pp.329 - square brackets added). However, as Bryant and Trabasso themselves had noted, in a five-term series, the B:D inference alone (not the symbolic distance effect) is most indicative of deductive solution.

A prominent study serves to illustrate just how significant this issue can be. McGonigle and Chalmers (1992) completed the planned training with their primate participants (taking some 6 months), observing that the B:D-inferential response was often lower than its antecedents. However, rather than investigating this finding further (it actually suggested monkeys really might make deductive transitive inferences), they felt obliged to continue training for around a further 2 months until they had thoroughly trained away this profile. Only then did they conclude that their participants were exhibiting deductive transitive reasoning (they now offered the term "rationality").

Now, consistent with the originators of the extensive-training paradigm, virtually all the above studies used implied transitive series of at least five items. However, the finding of non-humans being transitive reasoners is not just restricted to extensive-training tasks using five or more items (Browne et al., 2010). For example, it is perhaps ironic that one of the very few tasks seeming to unambiguously show monkeys able to make transitive deductions, in fact used a three-term extensive-training task but it had elements of a non-training task for the critical comparisons (Addessi et al., 2008). Here, Addessi et al. (2008) first established a transitive relationship between three foods (A preferred to B, B preferred to C). Each food was then mapped to a class of symbolic tokens. It was found that, when presented only once with pairs of tokens (i.e., not the foods), not only were the values of each token honoured, but the A:C inference was made far more strongly than with the actual foods. We point out that the food preferences were likely to have been absolute preferences, in agreement with Bryant and Trabasso's initial false positive objections to three-term non-training tasks. The greater contrast with the tokens may therefore have simply reflected an enhancement of the absolute value of food-A in the absence of all the foods. Thus, as Premack (2007) now concludes, "in animals, transitive inference may not be based on deductive reasoning" (Premack, 2007, pp.13866).

Evaluating a Purely non-Training Account for Transitive Reasoning

Irrespective of one's views regarding the validity of the extensive-training paradigm for assessing deductive transitivity in either human adults or in non-humans, there is independent evidence that, actually, below 6 years children have great difficulty coordinating item information (item-A and item-C) about a middle term (item-B). For example, Rabinowitz and Howe (1994) tested 5 year-olds on eight different two-dimensional shapes on computer, in each case presenting three sizes of a given shape and requiring the children to indicate which of the three shapes was the middle-sized one of that shape. This revealed that 5 year-olds cannot reliably judge "middle-ness" when shown three items (A, B and C) simultaneously. So we should anticipate it really is unlikely that 4 year-olds can make this coordination in a piecemeal way, mentally rather than physically, and across five rather than three-items as required in the original extensive-training task.

In line with Rabinowitz and Howe's demonstration for perceptual discrimination of three items, studies using three-term non-training tasks reveal that deductive transitive reasoning does not actually reach competence at 4, 5 or even 6 years. For example, Wright and Dowker (2002) ensured memory immediately before giving the transitive response was 100%. We also ensured each child stated the premises as s/he saw it, did not allow labelling cues, did not cue the child either linguistically or non-linguistically, and did not require the child to give any verbal justifications for his/her transitive response. Thus, every possible rational objection to three-term tasks was addressed (e.g., the false-positives and false-negatives of Bryant, 1998; Bryant & Trabasso, 1971). Yet in this study, 5 and 6 year-olds did not perform at competence for the inference (around 52% indicating guessing between the two items A and B), and even 7 year-olds were only just approaching competence (their 60% is below the 66% point which indicates more correct than guessed responses when item's A, B and C are all potential answers).

In a further study, Wright (2006b) presented 5, 6 and 7 year-olds with three-term tasks that honoured transitivity, plus other tasks which required premise integration but did not involve a fully transitively-connected series (e.g., A>C, B>C. Which is greatest?). The aim here was to determine whether children apply transitivity when it is appropriate but withhold it when it is not appropriate (i.e., discriminate appropriate from inappropriate applications of transitivity). Because not all comparatives that can be used in a transitive-like way actually permit a logical conclusion (e.g., consider A parent of B, B parent of C: or A likes B, B likes C: or A larger than B, A larger than C - see Wright, 2006a and Goodwin & Johnson-Laird, 2006 for discussion), such a discriminative capacity is crucial. It was found that this discriminative transitive competence (being able to tell when transitivity should versus should not be applied) is actually quite well developed at 5 years. However, it was also confirmed that children still do not apply transitivity reliably, until after 7 years. This suggests the capacity at issue is not about knowing that transitivity applies, but rather it is about being able to readily integrate the pair-wise information to yield the implied transitive series (Andrews & Halford, 1998; Greene et al., 2006; Perner & Mansbridge, 1983; Viskontas, Holyoak & Knowlton, 2005).

Then, in what may be the very first direct comparison between a three-term non-training task modified to avoid Bryant and Trabasso's false-positives and false-negatives, and a five-term extensive-training task also avoiding these criticisms, Wright (submitted) asked 7.5 year-olds to solve transitive problems using either task. It was found that these children could do the three-term non-training task but were not reliably above chance on the five-term extensive-training task. Indeed, Wright (1998a) previously reported that children do not attain competence on such a five-term extensive-training task until they are around 12 years. Then, it was argued that the reason why five-term extensive-training tasks are harder than three-term non-training tasks is that the former involve the integration of 10 possible pair-wise relationships as opposed to only 3 in the latter case.

One might object that the above non-training studies involved tasks which might simply have been too abstract, and this is why children tended not to do well until after 7 years. However, Markovits and Dumas (1999) used everyday objects placed in pairs in an office stationary holder. In one task there was a paintbrush that was taller than a comb (the A:B premise). In a second premise-pair the comb was taller than a pen (B:C). Children then had to work out which would be the tallest out of the paintbrush and pen (A:C). Using this general paradigm, Markovits and Dumas arrived at an estimate of children's deductive-transitive competence of around 8 years (see also Ameel et al., 2007; Wright et al., 2011 for other concrete demonstrations).

Non-training tasks involving more than three items concur with the 7-8 year+ estimate for first arrival of a deductive transitive competence. For example, Bouwmeester and Sijtsma (2006) tested 6 to 13 year-old children of mean age 10.1 years on 16 transitive tasks without requiring training. Four of these tasks used three-terms, with another four using five-terms. Just as in Wright (submitted) when using a five-term task from the extensive-training paradigm, Bouwmeester and Sijtsma's data summaries showed that five-term non-training tasks are harder than three-term non-training tasks. The implication is that some integration capacity is essential regardless of whether or not training had been used (see also Goodwin & Johnson-Laird, 2006; Wright & Howells, 2008).

It was interesting to note that whilst Bouwmeester and Sijtsma's (2006) three-term tasks were readily solved from around 7 years, their five-term tasks tended not to be solved until after 10 years (see also Wright, 1998a). Applying Latent Class Modelling to additional data from 7.5 to 12.5 year-olds, Bouwmeester and Sijtsma (2007) replicated these findings. They also found that children fell into three classes. The low transitive ability children performed poorly across all transitive tasks. We must stress that this finding was based on around 10 times the number of children in the Bryant and Trabasso study (615 children). Hence, we have confirmation that even between 8 and 12 years, some children still find transitive reasoning difficult. Other children were intermediate for transitivity, in that they solved tasks presented in ascending or descending order but did poorly if any unsystematic ordering was used (Riley & Trabasso, 1974). Children in the high transitivity class did well across all transitive task orders.

Additional non-training studies have revealed that, so long as we avoid the need for extensive-training, transitive reasoning is linked to WM (e.g., see Oakhill, 1984 with 8.5 year-olds). This tendency can be contrasted against what we have seen for extensive-training tasks. For example, as noted earlier, dual-process theory maintains that system-1 tends to draw little if at all on WM (Morsanyi & Handley, 2008; Stanovich & West, 2000). Van der Lely's (1997) finding that children having less well developed WM capacity did better on an extensive-training transitive task compared to children whose WM was much more developed, taken together with Russell et al.'s (1996) finding that 6 and 7 year-old children did well on an extensive-training task which was stored in LTM and not in WM (both outlined earlier), seems very much in line with this system-1 assertion. The fact that both these studies produced findings that supported experimental (H1) hypotheses rather than simply having to accept null (H0) hypotheses, adds to our view that extensive-training invokes LTM far more than it invokes WM. However, we concede that, ideally, we need additional studies based on individual differences plus studies based on tracking the strength of relationship between transitivity and WM as training progressed, in order to be more definitive on this issue (e.g., see Reyna, Estrada, DeMarinis, Myers, Stanisz & Mills, 2011 for use of such triangulation regarding data analyses).

Notwithstanding this point, we can also be more certain about WM usage by considering how findings differ regarding system-2 versus system-1 in respect of WM. In contrast to what it says about system-1, dual-process theory (e.g., Morsanyi & Handley, 2008; Stanovich & West, 2000) maintains that tasks which do draw on WM are more likely to utilise system-2 (analytic) processing. Relevant to this assertion, Halford, Bain and Maybery (1984) used a dual task paradigm to investigate the relationship between transitivity and WM. Adults were presented with a number of tasks employing binary relations, with participants simultaneously required to do a secondary task known to tap WM. This was a digit encoding task, with recall tested after the binary task was complete. The main binary task was a three-term transitive task. The other binary tasks did not necessitate coordination of information. They found that the transitive task had a detrimental impact on digit task recall, whereas the other binary task did not. These findings were also obtained when the digit task was replaced with either an easy or hard arithmetic task. The implication is that it is the need to coordinate the two premises about their common term during transitive reasoning, which necesitates drawing on a WM resource in common with numeric processing. It is important to recall here that Trabasso (1977) had himself maintained that the amount of extensive-training used in his tasks enable task solution without the need for this very coordination (see also, Breslow, 1981; Wright, 2001).

Halford et al. (1986) subsequently confirmed that there is greatest processing load precisely at the point where the second premise is perceived. Thus, the processing load is greatest at the point in time when a premise begins to be coordinated with another premise - i.e., during premise integration. Then, building on Halford and colleagues' work, other researchers have repeatedly confirmed that young adults with lower WM also experience increased difficulties during cognitive integration (Crone et al., 2009). They also showed that typically-developed older adults show decreases in transitivity commensurate with decreases in WM (Viskontas et al., 2005). Thus here we have support from studies of individual differences as well as studies comparing group-wise performance. However, we are aware of one robust and well-controlled study which may not sit well with our WM dissociation (Reyna & Kiernan, 1994 - see later for outline). At this point, it should be noted that this study used non-training together with delayed testing by 1 week, which meant that the representation of the taught information was in LTM similar to extensive-training tasks. Hence, it is not clear whether we should categorise Reyna and Kiernan's task more with non-training tasks or more with extensive-training tasks.

Five-term non-training tasks suggest transitive reasoning is still not much better than the 84% Bryant and Trabasso (1971 experiment 2) claimed for 4 year-olds, even by the end of adolescence. For example, in a recent study (Wright & Howells, 2008), we presented 17 year-olds with a five-term task but one that did not require training. Four separate premises were presented simultaneously at different locations on a computer screen. Participants could see all four premises and were only asked to answer one question for each new five-term transitive series. The critical B:D inference was found to be 89%, which is around 2% lower than Bryant and Trabasso (1971) had claimed for their 6 year-olds. However, our exact estimate has been repeatedly found for 17 to 19 year-olds using alternative transitive designs within the non-training paradigm (e.g., Goodwin & Johnson-Laird, 2006; Viskontas et al., 2005). The 17 to 19 age estimate has also been reported (and/or intimated) from studies of other domains of reasoning (Amsel et al., 2008; De Neys & Vanderputte, 2011; Reyna & Farley, 2006).

For example, De Neys and Vanderputte (2011 - experiment 3) tested two adolescent groups (13.1 years v 16.3 years) on 8 tasks requiring deduction (based-rate syllogisms). They found that whether or not there was a potential conflict between the reasoners' stereotypical experience (heuristic output) and the conditional or probabilistic solution (analytic output), 16 year-olds did equally well (around 95%). However, although matching this for congruent problems, the 13 year-olds were only at 20% for conflict problems. Note, if we took the heuristic route (intuitive route) as interchangeable with the analytic route, we could mistakenly conclude that the 13 year-olds showed a deductive competence equal to the 16 year-olds. We suggest that the insistence by some, that the extensive-training task should completely replace the non-training paradigm, may essentially reduce to this erroneous situation within the transitive literature.

Non-training theorists (e.g., Markovits et al., 1995; Wright et al., 2011) have not only repeatedly replicated their own findings regarding 4 year-olds (unlike extensive-training theorists), they have additionally tested links to maturation, education and cognition. For example, Verweij et al. (1996) gave 7.8 to 11.7 year-olds (grades 2 to 5) 10 computer-presented three-term and four-term tasks based on the non-training paradigm. Children were tested twice around 16 weeks apart, whilst in a given grade. It was found that transitive performance in this retested group did improve from November to March of the following year. When this group was compared to two new groups who were tested only once each (November v March, respectively), the retested group performed identical to the November- once group on their first test, but identical to the March-once group upon their second test. This shows that training in transitivity does not improve task performance long term, and/or that children's transitivity development is improved by maturation or general learning at school.

Throwing some light on the maturation versus schooling issue, Castle and Needham (2007) tested 7.8 year-olds according to Piaget's three-term length and space tasks. They found that the better a child's transitive capacity indexed in this way, the better was his/her ability to spontaneously use various tools for measuring in a school setting. Also, better transitive capacity came before, and predicted, improvements in unit-iteration (repeated application of a small object to accurately measure a large object). This can be taken to suggest maturation of one's relational integration capacity plays a role in transitive development (see Halford et al., 1998 for a well specified theory and computational model embodying this finding). Also, Castle and Needham (2007) found individual differences in children's transitive reasoning and also confirmed that there are real changes in transitivity over the course of a school year within any given child (see also Andrews & Halford, 1998; Bouwmeester & Sijtsma, 2006). We noted earlier that such individual differences are in fact an indicator of analytic processing (Evans, 2011).

These above findings mirror findings with more formal tasks of deduction given to children of similar ages, such as the Wason selection task and other conditional tasks (Amsel et al., 2008; Chapell & Overton, 2002; Markovits & Thompson, 2008). For example, Markovits et al. (1996) investigated conditional reasoning in 7 to 12 year-olds. The tasks were framed using pretend contexts, such as the likes and dislikes of people on a pretend planet. Taking both their experiments together, it was found that children's deductive capacities reach competence at around 8 years. Similarly, Crone et al. (2009) gave children and adults a number of variants on the Raven's Progressive Matrices, and concluded that the necessary deductive competence is reached soon after 8 years. Morsanyi and Handley (2008) generalise this conclusion still further. They tested 6.6, 8.5 and 10.5 year-olds on a battery of reasoning tasks plus WM and I.Q. tasks. The WM and I.Q. tasks were combined into a single cognitive development score. When Morsanyi and Handley entered age and cognitive development together as predictors of normative reasoning, they confirmed that even more-so than age, cognitive development was a positive predictor of reasoning proficiency.

So not only do we have further confirmation of the correct age for deductive transitive reasoning, but we also have evidence of its links to other aspects of cognition, plus similar age estimates from other reasoning domains. If and only if we use the non-training paradigm, we see a different side to transitivity, a capacity which both links to wider work on deductive reasoning in children and also to skills children acquire and improve during their schooling (Chapell & Overton, 2002; Rabinowitz & Howe, 1994; Wright, 2006b).

In the face of such compelling evidence from so many diverse research groups, it really is time to take seriously the view that the deductive form of transitive inference is not trivial, and takes quite some years before it is fully on-line (contrast Bryant, 1998; Trabasso, 1977). This conclusion does not mutually exclude the finding of transitive responding in a plethora of human and non-human groups. Rather, it simply challenges comparative theorists to find ways of including non-training aspects, if they are aiming to demonstrate a specifically deductive transitive competence in non-human groups (e.g., see Addessi et al., 2008; Premack, 2007), or wish to demonstrate that human and non-human transitivity are based on the same process (Wynne, 1998; Vasconcelos, 2008).

Brain Research and a Dual-Process Account of Transitivity

Above, the argument was made that tasks based around the extensive-training paradigm versus non-training paradigm seem to index different transitive capacities, one more associative (intuitive) and the other more deductive (analytic). However, we also got the first glimpse that transitive tasks have not only been applied to many species other than humans, but are increasingly being used to help us understand atypical human development over and above typical human development (Lutkus & Trabasso, 1974; Stromer et al., 1973). Transitivity both in typical and atypical groups has been studied using neuroscience experiments (human brain research). Some studies concern brain function as assessed by Magnetic Resonance Imaging (MRI). These have been structural in nature, or functional (FMRI) or even Event-Related (EVR-MRI). Other studies have concerned Positron Emission Tomography (PET) or even Magneto-Encephalography (MEG). Studies of all these types rely on brain-related equipment and operator skills which not only can be highly time-intensive but also tend to be highly expensive. This makes it all the more important that brain research theorists are aware of the nature of the transitive capacity they will likely be assessing, for example, are they supposing it to be "logical" or "associative". In turn, this makes it all the more important that the debate about which paradigm is better able to index deduction be resolved once and for all. Fortunately, when taken alongside studies of non-human brain function, human brain research can actually be as helpful in answering that question as it can be in using transitive tasks as tools to help answer questions about atypical groups. Below, we consider what human and non-human brain research tells us about the picture of transitive reasoning that has emerged thus far.

Research based on the extensive-training paradigm tends to implicate one brain region in particular, for our ability to form transitive inferences - the hippocampal formation. Arguably, this realisation was borne out of studies with non-humans (e.g., with rats - Dusek & Eichenbaum, 1997). In their seminal comparative study, Dusek and Eichenbaum (1997) trained rats on a series of premise pairs, in which one odour was always rewarded when presented with a specific other odour. The odours were connected and implied a transitive series +A:B, +B:C, +C:D and +D:E. After extensive training on these overlapping pairs, Dusek and Eichenbaum lesioned the rats' hippocampus, a small area quite central in the brain, which they had previously shown to be important for rat learning. The rats were subsequently tested on the premise pairs and the inferential pairs. It was found that ability to make inferential responses was very much reduced compared to intact rats. This confirmed that hippocampus is crucial for the relational learning involved in inferential responses on a transitive task. Interestingly, hippocampal abolitions did not adversely affect the rats on the premise pairs (see Fuzzy-Trace theory later).

Following this and subsequent investigations, Eichenbaum (2001) presented a very detailed model of the role of the hippocampus in verbatim and relational memory, informed by single unit recording of rats on tasks that used extensive-training. According to the model, transitive generalisations are built up within associative neural networks within the hippocampus, which encode spatial location information independent of identity and identity information independent of spatial location. Eichenbaum (2001) tested the model by disconnecting pathways between the hippocampus and cortical/sub-cortical areas. As predicted, this resulted in intact premise learning but disrupted inferential generalisation.

In humans, Greene et al. (2006) tested adults on a five-term extensive-training task. Participants were trained outside the scanner and then tested whilst under event-related FMRI. This not only showed up the hippocampus as specifically active during testing but also it was especially highly activated for the inner pairs compared to the outer pairs. The inner pairs are the ones most requiring relational integration (Perner & Mansbridge, 1983; Wright, 2006b). It was also found that a distinct pattern of hippocampal activity for these inner pairs differentiated those able to perform the inferential discrimination, B>D, at test. Other FMRI studies have confirmed with stimuli from geometric shapes to faces, that the hippocampus as critical to the relational reasoning required for inferring transitive responses when testing follows extensive training (Dickins & Dickins, 2001; Wendelken & Bunge, 2009). The critical importance of the hippocampus in human transitive reasoning has been further corroberated by MEG studies with humans (Hanlon et al., 2011) and PET studies on human participants (Nagode & Pardo, 2002).

The hippocampus is not just implicated in transitive responding but also in other tasks which used long-term learning (e.g., repeated presentations of the source information). This suggests it is involved in acquiring, representing or retrieving well-learned information. For example, it is implicated in the acquisition of slow probabilistic learning (Shohamy, Myers, Kalanithi & Gluck, 2008) and in the gradual learning of transverse patterning problems (Moses et al., 2006). In the transverse patterning task, premise +A:B and +B:C are learned using extensive-training; however, a further premise is also trained (+C:A) which makes the transitive series circular. Not only does hippocampal damage impair humans on transverse patterning tasks, it actually makes those tasks almost impossible to learn for other species (e.g., rats - Driscoll et al., 2005; nonhuman primates - Alvarado & Bachevalier, 2005).

One might wish to argue that it is not so much that the hippocampus is critical to extensive-training tasks as compared to non-training tasks, but rather, it is that the tasks relied on in hippocampal research invariably use five or more terms as compared to only three-terms. But the above study (Moses et al., 2006) begins to address this issue, as it used three-term series although the series was circular rather than fully-transitive. Hanlon et al. (2011) further address this issue in so far as they used three-term tasks which were fully transitive in their neurological study. They used MEG to compare human transitive performance in persons having schizophrenia, a disorder which involves hippocampal damage, to their performance on a standard transitive task. It was found that transitive premises could be readily acquired but the premises of the circular series (transverse patterning series) could not be acquired at all. After around 180 trials, in patients, the premises of the transitive task were learned as well as non-patient controls. However, patients could not learn the transverse patterning task. This suggests two things. First, the hippocampus is essential in order to construct/maintain an interlinked series (transverse patterning), even when only three items are involved in that series. Second, it suggests that such an interlinked representation is essential in order to learn the premises in the first place, both for transverse patterning and transitive three-term series (Bouwmeester & Sijtsma, 2006; Brainerd &

Kingma, 1984; Wright, 2006a, b). For test trials, Hanlon et al. (2011) found that, whereas control participants showed greater right hippocampal activation for visually-posed series but left activation for verbally-posed series, patients showed heightened left hippocampal activation for visual tasks as well as verbal tasks. This was taken as implying the left hippocampus is hyperactivated in schizophrenia.

Just as unconscious transitive reasoning was noted in the experimental transitive literature earlier, hippocampal learning also does not have to be conscious learning. In their FMRI study, Greene et al. (2006) gave participants a post-scan questionnaire, asking questions about awareness of the transitive series which had been taught using extensive-training. They found that those participants who were classified as unaware B:D performers actually showed greater parahippocampal activation to the B:D pairs than did the aware B:D performers. This suggests that not only is the hippocampus critical to transitive performance under extensive-training paradigms, but it is also implicated in non-conscious performance (Siemann & Delius, 1996; Wright, 1998a); supporting the thesis that the performance looked at with extensive-training tasks can be achieved with unconscious processing (i.e., via the intuitive route).

Thus far, we have seen findings from studies which used the extensive-training paradigm; even studies doing so using a three-term extensive-training task (Hanlon et al., 2011; Moses et al., 2006). But as explained earlier, it is perfectly feasible to use non-training tasks of three or more terms. When such transitive tasks avoid extensive-training, brain regions other than the hippocampus now seem to become primary to the inferential responses. The brain region most unique to these studies is the frontal lobe. More specifically it is areas within Prefrontal Cortex (PFC). PFC is important for WM, including inhibitory control, executive function and attention (Wood & Grafman, 2003). In non-humans, Addessi et al. (2008) summarise evidence showing that in Rhesus Macaques and in Capuchin Monkeys, neurons in specific regions of the frontal cortex fire during the making of the A:C inference after being presented with A:B and B:C without extensive training (Padoa-Schippa, Jandolo & Visalberghi, 2006).

In humans, Goel et al. (2004) used FMRI with a transitive task using training (familiarity) and another one using novel stimuli (unfamiliar). Each task was found to activate an array of areas throughout the brain. However, critically, the familiar transitive task substantially recruited the hippocampus and occipital lobe, whereas the unfamiliar task substantially recruited areas of the frontal lobe and parietal cortex. As well as supporting the dissociation we allude between extensive-training tasks being highly hippocampus reliant, and non-training tasks being more frontal cortex reliant (e.g., anterior PFC - APFC, dorsolateral PFC - DLPFC), this also supports the claim that the extensive-training paradigm tends to index a lower level capacity than does the non-training paradigm. This is because among other things, the parietal cortex holds higher level representations which were built from lower level visual representations from the occipital cortex. The result is construction of spatial information that is not reliant on an absolute end point (e.g., no longer reliant on item-A of the transitive series). Again using FMRI in humans, Golde, von Cramon and Schubotz (2010) showed that a specific area of PFC (APFC) is functionally differentiated from other areas in frontal cortex by the type of processing it does, and is specialised for the integration of abstract connected information such as occurs in transitive inference.

Some recent studies have been able to even more precisely locate the locus of the transitive coordination process said to be basic to deductive transitive reasoning (Crone et al., 2009; cf. Halford et al., 1998; Wendelken & Bunge, 2010). For example, Wendelken and Bunge (2010) tested 16 participants using a simultaneous five-term non-training transitive task, similar to our own simultaneous task outlined earlier (Wright & Howells, 2008). They found that for relational trials (inferential transitive responses using the "greater than" weight relation on screen), an area of the Prefrontal Cortex within APFC (Rostralateral PFC), was differentially activated. However, none of the hippocampal areas were activated. Then, for trials not requiring transitive responses, an area of the hippocampus was now differentially activated instead of PFC. Here then, we appear to observe a functional dissociation that correlates to structural areas.

The importance to relational tasks within the non-training paradigm, of PFC, and perhaps APFC or more specifically RLPFC in particular, holds for children of 8 to 12 years as well as for adults, as assessed using event-related FMRI (Crone et al., 2009). Additionally, in agreement with estimates for the age of acquisition of the deductive mode for transitive reasoning being near 8 years, Crone et al. found that children processing relational tasks requiring level 1 relational complexity (i.e., two overlapping premises - Halford et al., 1998), seem to call on a capacity in some sense equivalent to what adults use for level 2 relational complexity (i.e., integrating three premises). Thus, reasoning is far more challenging for children than adults. This finding held both in the response accuracy data and the RT data. We believe the above conclusion is warranted for the following reasons: Crone et al.'s FMRI data showed that while adults appeared to have to engage DLPFC and RLPFC in a sustained manner, only when faced with level 2 relational integration; such sustained engagement of DLPFC in children was required even for the simpler 1st-order (level 1) relational task.

Additionally, children performed far less accurately on level 2 tasks, although they tended to respond in identical time to adults (typically less than 8 seconds). This was taken to imply that on level 2 tasks, children either give up after around 8 seconds of thinking, or they solve problems using different (less deductive) strategies to adults. Finally here, Crone et al. concluded that the evidence that activation in RLPFC associated with relational integration increases as a function of age suggests that development of the RLPFC integration mechanism occurs over the age range (8–12 years). This is the same age range as identified as critical to the initial emergence of transitivity as a deductive competence and also many other domains of deductive reasoning (Bouwmeester & Sijtsma, 2007; Chapell & Overton, 2002; De Neys & Vanderputte, 2011; Markovits et al., 1996; Markovits & Thompson, 2008; Morsanyi & Handley, 2008; Verweij et al., 1996; Wright, 1998a, b).

Our final study here is particularly illuminating: Dickins and Dickins (2001) trained adults using extensive-training in order to establish equivalence classes between six items (any item =

any other item). They did this separately for three sets of six items; such that A1=A2=A3=A4=A5=A6, B1=B2=B3=B4=B5=B6, and C1=C2=C3=C4=C5=C6. The following day, Dickins and Dickins (2001), presented participants with a number of individual three-term transitive series. Crucially, each series made use of one item from each of the three sets which had been trained the previous day. This testing essentially used a non-training procedure but with substitutable items which had been established through extensive training. Thus, item-A from each series was interchangeable with any of the other five item-As (see Yamazaki, 2004 for discussion of equivalence classes). This procedure bears some similarity with mapping training as used by Andrews and Halford (1998) with young children. The important finding for our purposes was that, when performance was assessed within each item set (extensively-trained 6-term series e.g., A1...A6) the hippocampus was especially activated, but when the items were part of untrained three-term series (A, B and C), the PFC was now especially activated. Here then we have the clearest demonstration of extensive-training tasks tapping hippocampus but non-training tasks tapping PFC, in a set of tasks which actually used the same set of stimuli with the same set of participants. Again, the duality being proposed here is re-confirmed. Intriguingly, PFC only showed pronounced activation for participants who had performed competently on the three-term transitive task. Thus, unlike in monkeys (Addessi et al., 2008), it even seemed to be discriminating valid transitive reasoning from non-valid transitive reasoning (i.e., indexing successful deduction).

Towards an Integrated Dual-Process Theory of Transitivity

The distinction between a transitive capacity based on associations in LTM and drawing on the hippocampus, versus a second transitive capacity based more on WM and drawing heavily on PFC, was in general not facilitated by the developmental studies. Indeed, many developmentalists still tend to accept extensive-training studies but deny that non-training studies are even relevant to any kind of transitivity (e.g., Bryant, 1998; Russell et al., 1996). It may be speculated that this seriously affected the length of time it has taken for us to settle on the following view: Whether we are talking about humans or non-humans, young children or older persons, individuals with learning difficulties or typically-developing individuals, or even normal brains or abnormal brains, two distinct processes of transitive responding (not just one) are necessary to adequately explain transitive performance.

Indeed, by the time of the Pears and Bryant study in 1990, Trabasso (i.e., an architect both of the extensive-training paradigm and the information-processing theory specifically associated with it) had himself signalled a change in view. For example, Trabasso et al. (1989) found that some procedural formats result in the usual memory dependent profile, with these instances leading to inferential responses that actually get better the greater the distance between the items to be reasoned about in the transitive series (the Symbolic Distance Effect - Vasconcelos, 2008). However, Trabasso et al. (1989) additionally reported that when highly extensive-training is not relied on for a five-term series, the accuracy of inferential responses now deteriorates (rather than improving) with increasing distance (the reverse Symbolic Distance Effect - Barrouillet, 1996; Wright, 2006b). Crucially, in his 1977 paper, Trabasso himself had previously argued that the reverse symbolic distance effect would imply the use of logical deduction. Trabasso's own work then, now supports not one but two modes for reaching transitive conclusions; one non-deductive and the other deductive (Trabasso et al., 1989). Fortunately, thanks to additional evidence from the two apparently competing and mutually exclusive research paradigms, some of which comes from theorists who have relied both on extensive-training and non-training tasks, Trabasso's more recent realisation and the case for a dual-process theory of transitivity development is now almost incontrovertible (Ameel et al., 2007; Bouwmeester & Sijtsma, 2007; Wright & Howells, 2008).

A mode of transitive responding present by 4 years-old, may well be demonstrable only under very extreme and favourable training procedures (see Holcomb et al., 1997; Riley & Trabasso, 1974). This said, the plethora of studies demonstrating it in atypically-developing groups and also in non-humans, suggests it may even be an innate capacity (Allen, 2006; Siemann & Delius, 1996; Wittemyer & Getz, 2007). Importantly, a different transitive capacity (the one most have been interested in) develops over several years, maturing from 7 years onwards (Markovits & Dumas, 1999; Wright et al., 2011). Mirroring PFC development, it does not begin to reach its full adult potential until some time during adolescence (Perner & Mansbridge, 1983; Viskontas et al., 2005; Wright & Howells, 2008).

It does actually make good sense to speak of two general ways for reaching transitive solutions. One (analytic route) for situations that might be one-off, and the other (intuitive route) for situations likely to be repeated over and over again: One for new or novel situations and the other for heavily memorised information. Or even one for conscious/wilful inferences and the other for unconscious/automatic inferences (Martin & Alsop, 2004). The claim is not that one of these routes ever obviates the other, but that they are both continually activated (e.g., in parallel), but with the analytic route guiding decisions if it has produced output (Amsel et al., 2008). Evans (2003) summarises evidence that indicates the analytic system (system-2) tends to lead to much slower decisions than the intuitive system (system-1) and is more taxing on linguistic and WM processes (see also De Neys & Glumicic, 2008). Consistent with this view, regarding transitive reasoning, Wright (2006b) trained adults on a five-term extensive-training task, and timed RTs for the critical B?D inferential comparison at around 950ms by the end of all training given. But when Wright and Howells gave participants a transitive task that again used five-terms but this time did not require any training because all premises were familiar and continually in view, we found that responses now took much longer (around 3000ms) even though performance was similar to the first of these studies. We have since closely replicated these findings within a single participant group where each participant was given both tasks (in preparation). In line with this roughly 2 second slowing for a non-training task compared to an extensive-training task, De Neys and Dieussaert (2005) also reported an analogous 2 second slowing for participants with good WM who did their syllogistic task analytically versus heuristically (see also Goodwin & Johnson-Laird, 2006).

The two reasoning processes can be almost completely independent of each other. For example, for probabilistic reasoning, priming seems to affect the associative (intuitive) process but not the deductive (analytic) process; whereas secondary tasks, change of reasoning context, or giving unrelated training to induce formal reasoning, each affect the analytic process but not the intuitive process (Ferreira, Garcia-Marques, Sherman & Sherman, 2006). However, for certain reasoning domains (e.g., syllogistic reasoning), the analytic process may actually need to be able to suppress the output of the intuitive process (De Neys & Vanderputte, 2011). This suppression draws on the executive processing domain of WM, which further slows system-2 relative to system-1 (De Neys & Dieussaert, 2005). On still other occasions, the two processes may work together rather than competing. For example, the deductive process might generate inferential solutions and then store these over the long term via an associative integrated mental representation (Ferreira et al., 2006; Wright, 2006b). Crucially, since in practise the modes are not always mutually exclusive of one another, there is no reason at all for the paradigms that call up these modes to be mutually-exclusive of one another. Indeed, as achieved by Verweij, Sijtsma and Koops (1999) and Dickins and Dickins (2001), there may even be ways of combining both paradigms into new hybrid designs, capable of revealing much more about how transitive reasoning is embodied in human thought (see also Addessi et al., 2008 regarding monkeys).

Notwithstanding Piaget's theory which had already alluded to two routes to transitive responses in children (for discussion see Barrouillet, 2011; Chapman & Lindenberger, 1988; Piaget et al., 1977), the first modern account relevant to transitivity which explicitly argued for two processes came from Brainerd, Reyna and colleagues (e.g., Brainerd & Kingma, 1984; Brainerd & Reyna, 1990, 2001, 2004; Reyna & Brainerd, 1990; Reyna & Ellis, 1994; Reyna & Kiernan, 1994). Initially, Brainerd and Kingma reported six transitive three-term and five-term experiments with children of around 5 to 9 years. Two findings emerged across all the experiments. First, performance on the inferential pairs did not seem to be dependent on the premise pairs. But second, both premises and inferences seemed to be based on a single integrated representation of the entire implied transitive series. Building on these findings, Brainerd and Reyna (1990), after a reanalysis of several existing studies (e.g., Bryant & Trabasso, 1971; Chapman & Lindenberger, 1988), reached two further conclusions. Third, the integrated representation is based on a fuzzy-trace of the entire implied series, and results in a representation capturing the gist and flow of the series (e.g., "things get bigger to the left" Brainerd & Kingma, 1984, pp.336). Fourth, they concluded that verbatim recall and inferential responding are independent but parallel processes. This theory resembles a dual-process theory, but with the potential advantage of being based on processes which both are more basic and which can be very precisely conceptualised and specified (Reyna & Brainerd, 2011; but contrast claims by Vasconcelos, 2008).

Regarding Brainerd's and Reyna's notion of memory independence, this thesis would seem at odds with the memory-dependence view of Trabasso (1977) and Bryant (1998). Indeed, it would also seem at odds with the vast literature on transitivity in non-human participants (e.g., Bond et al., 2003; MacLean et al., 2008; Vasconcelos, 2008; Wynne, 1998). On memory dependence versus independence in children, Chapman and Lindenberger (1992) found that older child groups exhibit better premise retention alongside better reasoning; and that within any age group, those participants better at retaining premises tend to exhibit better reasoning. This is in line with Trabasso's (1977) early conception of reasoning being highly dependent on memory. However, Brainerd and Reyna (1992) found that whether or not children give a particular correct transitive response does not depend at all on whether they could correctly recall the premise information logically sufficient for reaching that inference. This finding has been shown to be robust and validates the Memory Independence Effect (cf. Brainerd & Kingma, 1984).

Recently (submitted), we tested both memory theses against both the non-training paradigm and the extensive-training paradigm, within a single study. Seven and a half year-olds completed either a 3-term non-training transitive task or a 5-term extensive training task. Both in analyses of sub-group performance (e.g., good memorisers v poor memorisers) and in analyses of individual differences (e.g., regression analyses), the non-training task invariably revealed memory independence between premises and inferential responses; but the extensive-training task showed memory dependence. This study adds weight to our thesis that the two paradigms have tended to look at rather different capacities. But this leads us to a possible duality between processes called on in the respective tasks, versus a duality of processes within the extensive-training task.

Although we have not as yet found a study that has directly contrasted these two possibilities for a dual-process conception, we do note one study that seems to have offered a solution, by employing a task that could be considered as non-training but also as having relied on LTM representations such as would be formed from training. Here, Reyna and Kiernan (1994) tested 6 and 9 year-olds in two experiments. Although they did not use extensive training as such, they had participants learn premises (sets of three sentences making up a story) and then tested them both immediately (tantamount to non-training) and one week later (tantamount to using the kind of LTM representations induced by extensive-training). For the three sentences, two were relatable to yield an inference (linear or spatial). In their first experiment, children were asked to consider the exact sentence wording but would sometimes be asked to make judgements on other types of sentence, for example, false sentences or differently-worded true sentences. It was found that responses indicated they discriminated between true sentences that were verbatim and true sentences that were paraphrases of true verbatim sentences. Both these sentence types would produce the same gist representation (e.g., the flow of spatial relations in spatial sentences), and hence Reyna and Kiernan concluded that the two above sentence types could not have been generated solely by a single gist representation of all sentences.

In a second experiment where children were explicitly instructed to consider the gist rather than the verbatim information, their responses now indicated that both gist and verbatim sentences were using the same gist representation. Reyna and Kiernan concluded that "we must relinquish a commonly held assumption in research on language, namely, that gist is derived from verbatim representations" (Reyna & Kiernan, 1994, pp.189). Indeed, Reyna and Kiernan's experiments would seem to indicate that, if anything, it is verbatim information (e.g., the premise information in transitive tasks) that can readily be gleaned from the gist and flow, rather than the other way around. This said, gist and verbatim information do begin as two independent processing routes, although gist might assist the more challenging verbatim process. But whether these routes are enough to handle both findings from the extensive-training paradigm and the non-training paradigm is open to debate.

Reyna et al. (2011) consider the verbatim versus gist distinction in fuzzy-trace theory to represent a dual-process theory. We move from a reliance on verbatim representations to gist representations. That is from literal individual encoded statements and other information, to the meaning (or gist) that should be drawn when such information was presented. Thus, gist is not just "inference" but is also in the more wider sense "interpretation". This makes it difficult to directly relate fuzzy-trace theory to our own dual-process theory; although it does not necessarily challenge either theory. Note, another way of looking at the distinction offered by Reyna et al. (2011) is to argue that gist processing develops rather slowly and is not mature until some time during adolescence. In this conception, gist, although seemingly an instance of intuitive reasoning, does start to look like analytic reasoning (Reyna & Brainerd, 2011). Then, we would wonder what younger children who have not yet adequately developed this ability can do whilst waiting for it to arrive. The only option would seem to be to rely on the information that older children can produce gist from. That is to rely on the initial internalisation and initial representation of the stimulus, which might be considered to be verbatim information. So it may not be that verbatim and gist processes represent a productive duality, but simply that verbatim information is relied on earlier but is of lesser utility.

This contrasts with our own above conception of dual-process theory for transitivity, in which both the intuitive route and the analytic route are highly important and always used to complement each other (Reyna & Farley, 2006). Additionally, it is important to note that in our conception, Reyna et al.'s "gist" would seem to have to span both our intuitive process and our analytic process. If gist representations can routinely be computed with little verbatim training, then we hold that analytic thinking is taking place. If gist representations can only be achieved by highly extended training, then we would hold that the individual is still dominated by intuitive thinking. The difference between verbatim versus gist representations may therefore essentially reduce to asking whether the individual can actively arrive at gist or whether gist would have to be passively induced via highly extensive-training.

According to Brainerd and Reyna (2004), young children generally reason less well than older children because they cannot rely enough on gist processing. Thus, they have to fall back on coordination of verbatim information, which both is more difficult and less durable than gist generalisations (Reyna & Kiernan, 1994; Reyna et al., 2011). In agreement with this view, Bouwmeester and Sijtsma (2004) found with children of 6 to 13 years, that those children who "reduced the premises", that is to say they mentioned the premise information and gist as part of their answer or their explanation of why they reached their answer, these children tended to do better than children who simply repeated visual aspects of the information they had seen or drew on external information (the latter were near chance). Also, younger children may not have as well developed rules of reasoning nor well developed inhibitory processes (e.g., Brainerd & Reyna, 1990; Kavale et al., 1995). Bouwmeester and Sijtsma (2007) argue that verbatim memory is relied on simply because it is relatively well developed by 5 years, whereas gist memory might take a few more years to develop. However, Perner and Mansbridge's (1983) study required only memorisation of the premises (i.e., needed no inference) and yet 8 year-olds did poorly when interlinked premises were involved. In other words, it might not simply be that young children have difficulty with gist, but rather that they have difficulty with the integration process that results in gist (Chapman, 1999; Halford & Andrews, 2004). Improvements to integration both will improve gist and improve ability to remember the verbatim information (Wright, 2006b).

Brainerd and Reyna (2004) conclude from further studies, that both verbatim and gist memory develop mostly during middle childhood but their development follows very different trajectories. Notably, the gist process takes years longer than the verbatim process, to develop near adult levels. One implication is that younger children will require more assistance in order to achieve gist processing which approaches or even improves on older children on transitive tasks (de Boysson-Bardies & O'Regan, 1973; Kallio, 1982; Wright & Dowker, 2002). Note, Bond et al. (2003) and MacLean et al. (2008) have reached essentially the same conclusion regarding within-species comparisons of non-human species (see earlier).

So if gist takes longer to develop and children therefore rely on verbatim premise representations, resulting in poor inferential performance on transitive tasks, does this help us explain finding such as those of Trabasso and colleagues? Recall Trabasso's studies revealed that even 4 and 5 year-olds have little difficulty reporting transitive inferences when trained extensively enough (Bryant & Trabasso, 1971; Holcomb et al., 1997; Riley & Trabasso, 1974; Trabasso et al., 1975). The answer to our question might be in the nature of the assistance given to children in acquiring premises in studies such as Trabasso et al. (1975). That is to say, by repeating each premise many times over, and doing so often in the order those premises fit into the global transitive array and for as many sessions or days as necessary, we help the child to very gradually build-up the gist of the transitive relations given across all premises, regardless of how well developed or poorly developed are the child's gist processes, and regardless of whether any premise is also encoded in a verbatim manner. The more such training we give the better the transitive performance; although as Trabasso (1975, 1977) himself noted, there is now no need for the kind of logical coordination that Bryant (1977, 1998) or Piaget (1954, 1970) aimed to investigate.

In line with our theorising, the 4 and 5 year-old children in Bryant and Trabasso's (1971) experiment 2 were intimated to be some or all of those from experiment 1. This means that the already passively-trained gist representations could be further consolidated in experiment 2, rather than children having built up the higher levels of performance in experiment 2 from scratch. This is really the only credible explanation of why critical B:D performance rose from 78% in experiment 1 to 84% in experiment 2, despite experiment 2 using the more demanding (non-visual) feedback regime. Note, Trabasso (1975, 1977) also retrained and retested some 17 of the same 4 year-olds from the Bryant and Trabasso (1971) study across three quite separate sets of training occasions. Performance improved from 78% to 84% (matching the Bryant & Trabasso reported levels with 25 children) and then to 93%; the latter being even higher than B:D performance of 6 year-olds in the original Bryant and Trabasso study.

In short, the view that the gist process develops more slowly than the verbatim process can help explain the finding that children (whether 4 or 7 years) tend to do relatively well on extensive-training tasks but poorly on non-training tasks (Wright, submitted; Wright & Dowker, 2002; Wright et al., 2011). A relevant prediction from Reyna and Kiernan's study is that, if we can isolate performance on premise pairs during the training phase of extensive-training tasks from premise reporting during the testing phase that occurred close in time to the conclusion of training, we should find superior performance in the latter case, because it is more likely that premises in the test phase will be reported from the gist of the entire transitive series. This prediction has in fact now been confirmed by Reyna et al. (2011), who found that testing immediately after adolescents and young adults were presented with the respective premises produced a bias towards participants reporting verbatim information over gist inferences, but testing a second time after 1 week had elapsed led to a bias now in favour of reporting gist.

However, this prediction has not, to our knowledge, been investigated specifically regarding a transitive reasoning task. We are currently testing it using an extensive-training task both with adults and children; using an incremented learning paradigm. On Reyna et al.'s (2011) findings, we should expect that during test phases the reporting of premises plus the reporting of inference would both be drawn from the same global gist representation. But during the training phases, as there is no need to consult the entire transitive array, there should be a greater tendency for participants to rely on verbatim representations instead of the global gist. We hope to report our findings on this issue in the near future.

As we have just seen above, fuzzy-trace (gist) theory will be critical to further expounding just how transitivity operates. But as well as generating some quite expected findings, it does also make some quite counterintuitive ones. For example, it makes the unusual prediction that under specific conditions, performance can sometimes appear to be less logical in an older group than a younger group. We have just seen this in a comparison between Trabasso (1975) and Bryant and Trabasso (1971) for 4 versus 6 year-olds (see also de Boysson-Bardies & O'Regan, 1973; Halford, 1982; Van der Lely, 1997). But as an example from wider cognition, Reyna and Ellis (1994) gave 4.7, 8.0 and 11.1 year-olds 18 problems embodying risk reasoning (probabilities). Half the problems were based on gambling relative to a small sure gain, with the other half on gambling relative to a small sure loss. The framing effect refers to adult reasoners generally choosing the risk option for problems posed in the context of losses compared to problems posed in the context of gains. Reyna and Ellis theorised that, on Piagetian theory or a generic dual-process theory, children should be expected to move from a more intuition-based implicit decision strategy (which would tend to show biases for loss context), towards a more logical strategy as they get older (embodying less and less difference between the gain condition relative to the loss condition).

In fact, it was the 4 year-olds who showed the least difference between the two contexts (i.e., no framing effect); and it was the 11 year-olds who showed the greatest overall tendency towards taking greater risk when a problem was framed in a loss context. This finding might imply children become less logical in adolescence! But whether this is true or not, the finding also suggests children reason more like adults in adolescence (Reyna et al., 2001). The findings for Reyna and Ellis' (1994) middle child-group further complicated the interpretation. Specifically, the 8 year-olds actually showed a reverse framing effect, tending to take greater overall risk for problems framed in a gain context compared to problems in a loss context (i.e., generally the reverse of 11 year-olds).

Again, in transitive reasoning, there are apparent developmental reversals here too (Kallio, 1982; Wright & Dowker, 2002). We saw another of these earlier, where Van der Lely's (1997) 8 year-olds outperformed their 10 year-olds. But it is now apparent that the extensive-training paradigm above should already have explicitly anticipated and documented this finding: In providing additional training for the 8 year-olds on their transitive task, Van der Lely was ensuring that the 8 year-olds were relying more on a build-up of gist information in LTM than were the 10 year-olds. Thus, Van der Lely had essentially appealed to intuition more in the 8 year-olds than in the 10 year-olds (Piaget, et al., 1977). The result was that, much as de Boysson-Bardies and O'Regan (1973) had found, "Children implement transitivity differently, and sometimes better than adults" (1973, pp. 533). However, de Boysson-Bardies and O'Regan had long since noted that "the children in these experiments were not using deductive transitivity at all" (1973, pp. 533).

The above notwithstanding, one can explain developmental reversals without recourse to fuzzy-trace theory. In the Loss-Gain problem, we note that if children do not appreciate the concept of "gain" and "loss" well enough, then they would tend to treat them the same (note that child-friendly task presentations meant that these terms were not actually explicitly used within the tasks). This alone can account for the 4 year-old's null comparison between gain and loss contexts. Indeed, Brainerd and Reyna (2004) themselves note "in these cases, young children will not extract salient event meanings from target items because those meanings are not yet understood, and many years must pass before they are." (Brainerd & Reyna, 2004, pp.410). Thus, in agreement with non-training theorists who propose 4 year-olds do not tend to reason logically, Reyna and Ellis' (1994) data may simply reflect that this group did not take account of level of risk but simply considered the potential payoff (taking the most attractive outcome - see Reyna et al., 2011 for similar argument). This will not have been due to children failing to understand the task or failing to comprehend the information presented: We note sufficient controls were present to prevent such issues. Rather, we argue that the younger children would simply have a different (less mature) conception of the phenomena (i.e., the contexts) employed in the tasks.

By 8 years, children will now have a better-developed appreciation of the concept of loss relative to gain. This may stem from the greater impact of negative feedback (teaching us about when we should alter our thinking) compared to positive feedback (telling us nothing new apart from that it is safe to carry on as we are). For example, children will learn about loss from experiencing forced-changes in their behavioural options (e.g., not being able to play with a particular toy any more) when things are lost or taken away from them by peers, teachers, parents, accidents etc (again we are talking of the concepts of loss and gain rather than the specific words "loss" and "gain"). This new understanding notwithstanding, their logical capabilities have only just started to be drawn on routinely in problem solving (Piaget, 1954, Piaget et al., 1977). Thus, here, we should expect children to treat the loss and gain contexts differently, although their greater conceptual appreciation of "loss" (loss-aversion) leads them to take fewer risks to secure uncertain gains in the loss context and take greater risks to secure uncertain gains in the gain context. This is the reverse framing effect.

In Reyna and Ellis' data, 8 year-olds did this only when the risk was high (0.75); suggesting that they were trying to take probability structure into account but were still most led by the payoff as with the youngest children. As both their understanding of loss and their deductive competencies reach maturity, children should be expected to take the probability structure more into account, reducing the bias (greater risk) of the gain context; and also they should be expected to apply this knowledge discriminately depending both on the level of risk and the level of potential outcome. This explains why in relative terms 11 year-olds took greater overall risk in the loss context when the potential payoff was small or intermediate (1 v 4 respectively) but took greater risk in the gain context when the potential payoff was large (30 - see also Reyna et al., 2011 regarding 15 year-olds). What we are saying is that Reyna & Ellis' fuzzy-trace account is perfectly plausible, but so is an alternative account in terms of a more standard dual-process theory (intuitive v analytic route); so long as we take into account the likelihood that children's semantic knowledge (here it was about the concept of loss) increases with age.

Fuzzy-trace theory does highlight and more neatly solve some intriguing phenomena such as memory independence and developmental reversals (Bouwmeester & Sijtsma, 2007; Brainerd & Reyna, 1990; Reyna & Brainerd, 2011). However, this does not mean there are no issues with it. For example, in their original account of fuzzy-trace theory, Brainerd and Kingma (1984) argue that "The advantage of fuzzy-traces is that they do not drain off large amounts of working-memory capacity because their structure is impoverished." (Brainerd & Kingma, 1984, pp.336). But we saw that for transitive reasoning and also for dual-process theory, one hallmark of deduction and the analytic processing route is that it tends to be WM intensive (De Neys & Dieussaert, 2005; Halford et al., 1984; Morsanyi & Handley, 2008; Stanovich & West, 2000). By contrast, associative transitivity and the intuitive processing route appear not to draw heavily on WM (Ellenbogen et al., 2007; Moses et al., 2006; Russell et al., 1996; Siemann & Delius, 1996; Stromer et al., 1993; Van der Lely, 1997). Therefore, if gist processing is not WM intensive, we should suspect that it may not yet offer a complete account of deductive transitivity (Evans, 2011).

Another potential issue with fuzzy-trace theory as a complete dual-process theory, at least as applied to transitivity, is that it does not as yet distinguish in any obvious way between the transitivity routinely indexed by extensive-training tasks and the transitivity that can be indexed by non-training tasks (e.g., see Brainerd & Kingma, 1984; Brainerd & Reyna, 1990). The experimental evidence, learning disabilities evidence, comparative evidence and brain research evidence each point to two ways of reaching transitive solutions; but fuzzy-trace theory currently does not distinguish them. Instead, it seems to apply mainly to performance on tasks which either used the extensive-training paradigm or which permitted ready-ordered visual tokens for the items in the transitive series. As with our WM argument above, this can be taken to imply that fuzzy-trace theory is more attuned to associative transitivity than to deductive transitivity.

Bouwmeester and Sijtsma (2007) do successfully align fuzzy-trace theory (gist theory) to deductive transitivity and not to associative transitivity; but this only re-emphasises our point: At least up until now, fuzzy-trace theory seems to have to choose one route, and seems unable to embrace both routes simultaneously (but see Reyna et al., 2011 earlier).

A further issue is that fuzzy-trace theory is concerned with "how verbatim and gist traces are stored in the first place, of how they are subsequently retrieved, of how they are preserved over time, of the characteristic phenomenologies that result when they are accessed, and of how they vary with development." (Brainerd & Reyna, 2001, pp.433). But what seems absent from this list is what we are most concerned with here - the process by which information is actually organised, actually coordinated, actually manipulated in the first place - analytic/deductive reasoning (Halford et al., 1986; Halford et al., 1998; Wright, 2006b; Wright & Howells, 2008). Fuzzy-trace theory then may currently be more about memory than it is about reasoning. Its gist representations enable retrieval but currently it is not clear how they in themselves constitute actual coordination of the premises (Halford et al., 1986). Although Trabasso's early work did hold this to be enough to explain transitive reasoning (Trabasso, 1977), we have seen that his later work drew back from this position (Trabasso et al., 1989).

Concerning fuzzy-traces and deduction, recognition/extraction of the gist and flow information has been found to be easier if the premise information is presented ordinally rather than scrambled (Brainerd & Reyna, 1990; cf. Kallio, 1982). Similarly, Bouwmeester and Sijtsma (2007) argue that reporting from the gist representation is also easier if the items are presented simultaneously and ordinally. In other words, gist processing is better when the child actually does not have to actively coordinate/manipulate the information for him/herself. But if deduction can be said to feature within any given time frame, it really has to be the time during which coordination actually occurs (i.e., integration).

In agreement with our current view, Chapman (1999) argues, the mental representation for the implied transitive series (i.e., gist trace) is important, but what is even more important are the cognitive mechanisms whose operations result in this mental representation in the first place. As an alternative to gist, we might well hold that transitive responses are deduced via the reasoner constructing and reading off from mental models (Bara et al., 2001; Goodwin & Johnson-Laird, 2006). But just as with fuzzy-trace theory, even here, what we really need to know is what knowledge, experience or ability leads the reasoner to choose to set up mental models (the representations) in the first place? What is the basis of the reasoner's realisation that constructing and reading off from such models might solve transitive problems anyway? We would argue that it is these abilities, over and above any other abilities, that approximate to any deductive competence. It is crucial to note that none of these abilities are readily gleaned from the mental representation of the entire transitive series itself.

Summary and Conclusions

The child development perspective is what first popularised the psychology of transitivity and caused it then to move forward radically. However, many of the great strides forwards have been in many other research domains (e.g., animal/comparative, learning disability, mental ill-health, neuroscience). These not only have increased our understanding of the various groups, but also our understanding of the nature of transitivity itself. Meanwhile, the development of transitivity in children (especially young children), has unfortunately not progressed much since Bryant and Trabasso's original study. It is as though the original positive findings with 4 year-olds propagated a reticence, not only to three-term non-training tasks globally, but even towards five-term tasks as used by Trabasso himself. The result was that progress towards a more shared developmental perspective has been quite sporadic (e.g., Bouwmeester & Sijtsma, 2006; de Boysson-Bardies & O'Regan, 1973; Halford et al., 1998; Markovits et al., 1995; Rabinowitz & Howe, 1994; cf. Riley & Trabasso, 1974; Trabasso, 1977; Trabasso et al., 1989; Verweij et al., 1999; Wright, 2006b; Wright & Dowker, 2002).

Arguably, the most critical issue in transitive research remains the disagreement about the age at which transitivity as a deductive competence is routinely available in children. As issues go, this is "one of the most vexed in the field of cognitive development." (Bryant, 1998, pp.266). The solution to this issue carries ramifications to many of the debates and many of the participant groups mentioned above. The studies in this paper confirm that the 4 year estimate derives from one particular paradigm (the extensive-training paradigm) but the 8 year estimate derives from a different paradigm (non-training paradigm). Advocates of each paradigm tend to deal with the age discrepancy in a different way: Extensive-training theorists simply dismiss the competing paradigm as incapable of assessing transitive reasoning at all (e.g., Bryant & Trabasso, 1971; Russell et al., 1996). But Trabasso's work aside, they seem yet to actually retest 4 year-olds using their own paradigm (e.g., see Bryant, 1998; Russell et al., 1996). By contrast, non-training paradigm theorists and also Trabasso, tend to accept the 4 year-estimate but claim it is a valid transitive capacity but not the deductive capacity that is supposed to be at issue (e.g., see Wright, 1998b). Non-training theorists do replicate their findings over and over with 4 year-olds (e.g., Markovits et al., 1995; Wright et al., 2011). Yet for some reason, these findings are simply assumed invalid, because of their non-alignment with what many wish to believe about transitive logic in 4 year-olds.

In order to help correct this unfruitful imbalance in transitive research, the present paper offered a myriad of empirical evidence, confirming that the debates about age and any issues about indexing a deductive mode of transitivity, stem primarily from a reliance on one or other of our two competing research paradigms (extensive-training paradigm v non-training paradigm, respectively). Then, rather than attempting to suppress either of these alternative paradigms or ignore their disparate findings, we sought to embrace both paradigms/claims, and instead see where this would lead. Where it leads is to a conception of transitive reasoning that is in some sense new (although for anticipation of our position see Brainerd & Kingma, 1984; cf. Piaget et al., 1977; Wright, 1998b). This is a dual-process conception of the transitive reasoning process. On this theory of transitivity, one paradigm (the extensive-training paradigm) tends to access an associative mode that captures transitive generalisations as the premise information is learned, with the other paradigm (the non-training paradigm) being better placed to tap a genuinely deductive mode of transitive reasoning.

Trabasso's systematic change in theoretical position about transitivity was always based on the available empirical evidence of the time, and hence it is unsurprising that he eventually ended up with a view actually consistent not only with dual-process theory as we propose here, but also consistent with the very theory he had originally opposed (i.e., Piagetian theory). Thus, by 1989, Trabasso had come to realise both that logic can indeed play a part in reasoning about transitive relations, and that therefore, any adequate theory must contain two routes - an analytic (system-2) route for deductive transitivity and an intuitive (system-1) route for associative transitivity. Trabasso (1989), then, may be regarded as intimating a dual-process conception of transitivity.

A dual-process perspective on transitivity allows the re-evaluation of a number of interesting findings using the extensive-training paradigm. We conclude the extensive training paradigm typically indexes an associative transitive capacity that is fast although slow to build-up, can be unconscious, which does not rely much on WM, and which is not closely related to educational attainment or even maturation. These are each system-1 (intuitive) constructs. We then show that non-human transitive research is in line with this conclusion. However, our evaluation of non-training tasks is different. This confirms that not only are these linked to slow but fast to set in motion processing and conscious thought, but they are also linked to other system-2 (analytic) constructs such as WM, intelligence and academic performance.

As well as an overwhelming array of experimental and comparative studies, brain research studies and mental-illness research also support our dual-process conception of transitive reasoning. Indeed, they even offer us quite precise loci for the critical neural systems relating to each of the two processes. What is especially important to appreciate here is that all the brain research studies seem to implicate the hippocampus as particularly critical to the intuitive process. Furthermore, all such studies happen to be based on the extensive-training paradigm, just as our dual-process theory would expect. Conversely, brain research studies implicate a completely different area (regions within the PFC as most critical to the analytic route), and only do so for humans. Here, the studies reflecting analytic processing and the PFC have in common the use of non-training elements to their design.

So it would seem that the duality of process we claim for behavioural psychological studies is closely mirrored in brain imaging studies. "There are two distinct types of transitive inference, with the wrong one currently being widely viewed as the only one" (Wright, 1998b, pp.848). The extent to which a participant calls on each will depend most crucially on how familiar s/he is with the source information (e.g., how much training has been received - Wright, 2006b).

Although we can be quite certain that a dual-process theory of transitive reasoning offers the best account on current evidence, we do not pretend that there are no further issues to tackle if this account is to exhibit longevity. For instance, we noted that the performance reversal effects such as reported by Reyna and colleagues (e.g., Reyna & Ellis, 1994; Reyna et al., 2011), can be problematic. Dual-process theory does intimate that such reversals are the result of a shifting (but predictable) dynamic between intuitive and analytic processes during middle adolescence. However, a dual-process perspective is yet to offer as precise an account of this phenomenon as already offered from fuzzy-trace theory (Brainerd & Reyna, 1990; Reyna & Brainerd, 1990, 2011; Reyna et al., 2011). Our conception of a dual-process theory of transitive reasoning is, however, quite consistent with fuzzy-trace theory, at least in principle. But we do note that even fuzzy-trace theory is not itself without issues. For example, a fuzzy-trace theory of transitivity does not yet appear to acknowledge that the deductive mode for transitivity might be distinct from the associative mode, with fuzzy-traces (i.e., gist) playing a different role within each mode or at least arrived at by different strategies (fast-acquired and active v slowly-acquired and passive).

What is clear though is that our own dual-process theory must benefit from a pre-existing fuzzy-trace variant of the two processes; and this achievement is an exciting prospect. However, this will still leave at least one emerging issue to be addressed: The issue of what is the mechanism that brings about the dynamic interaction between the intuitive and analytic systems? On this issue, Wright (2001) and Evans (2011) allude to the need to postulate a third "coordinative" system (Evans' "Type 3 process"). However, Amsel et al. (2008) have been more explicit on the nature of the third process: "In dual-process theory, metacognitive skills function to regulate conflicts between analytically and experientially based responses." (Amsel et al., 2008, pp.455). For example, Amsel et al. (experiment 1) found that college students (around 23 years) often report metacognitively that they are aware when a task should be solved analytically; but sometimes solve those very same tasks intuitively ("experientially"). This profile was not found in middle-school adolescents (13 year-olds - experiment 2). At least by adulthood, then, metacognitive skills are in some sense distinct from our intuitive and analytic processes. But it would seem that these metacognitive skills, themselves, would need ultimately to derive from a dynamic mix of the two processes. For example, metacognitive skills are heavily reliant on language, and Evans (2011) notes that language has to be a system 2 process but also has to make extensive use of system 1. What we allude to here, is that theories must work hard to avoid the danger of an infinite regress (e.g., the conflicts between the two systems are managed by a third system that itself reduces to the very same two systems!).

Notwithstanding such emerging issues, undoubtedly, additional to solving the long-standing debate on transitivity development, our dual-process theory of transitive reasoning might address some of the conclusions from more recent areas of research such as mental illness research and non-human research. For example, in schizophrenia research, it should now be possible to determine whether this disorder carries a decrement in relational reasoning from highly trained premises (i.e., a low level associative deficit), or a decrement in reasoning from premises that were given without the need for extensive training (i.e., a deficit in deductive reasoning), or even both modes. Hanlon et al. (2011) have already shown that in schizophrenia, the laterality effects seen in normal adults is replaced by a hyperactive hippocampus, particularly the left side. Interestingly, the left hemisphere tends to be called on for symbolic aspects such as linguistic processing. But symbols are an integral part of analytic thinking too. This suggests that schizophrenia might also involve abnormalities in deductive thought; but such analytic thinking actually implicates the PFC. This raises the question of whether schizophrenia involves an overactive left PFC as well as an overactive left hippocampus. This question can be answered quite simply, by comparing the critical inference on a non-training transitive task against the equivalent inference on an extensive-training task.

In comparative research, our theory opens up the possibility of determining whether the deductive mode of transitivity is just as species-general as the associative mode has been found to be, plus determining just which species can exhibit most human-like reasoning. On this issue, comparative theorists may not always be too optimistic, even regarding the prospect of species such as monkeys being confirmed as having a logical capacity (e.g., contrast Premack, 2007 v Vasconcelos, 2008), partly because of language (absent in non-humans) being so crucial to deductive thought (Chapman, 1999). Chapman does argue that information coordination/manipulation (i.e., our analytic route) must to some extent be founded in linguistic competencies. However, Wright (2001) and Stromer et al. (1993) independently note that it may be our symbolic ability rather than our specifically linguistic ability that facilitates coordination; and at least some other species have already been confirmed as having human-like symbolic ability if not a fully fledged linguistic ability (Addessi et al., 2008). Thus, the question of whether any non-human species exhibits deductive transitive reasoning, just like the question of whether 4 year-olds really do solve transitive tasks using deduction, is still very much an open one.

References

Addessi, E., Mancini, A., Crescimbene, L., Padoa-Schioppa, C., & Visalberghi, E. (2008). Preference transitivity and symbolic representation in Capuchin monkeys (Cebus paella). PLoS ONE, 3 (6), e2414 dio:10.1371/journal.pone.0002414.

Allen, D. N. (2006). Transitive inference in animals: Reasoning or conditioned associations? In S. Hurley, & M. Nudds, (Eds.), Rational animals? (pp.175-185). Oxford: Oxford University Press.

Alvarado, M. C., & Bachevalier, J. (2005). Selective neurotoxic damage to the hippocampal formation impairs performance of the transverse patterning and location memory tasks in rhesus macaques. Hippocampus, 15, 118-131.

Ameel, E., Verschueren, N., & Schaeken, W. (2007). The relevance of selecting what's relevant: A dual process approach to transitive reasoning with spatial relations. Thinking and Reasoning, 13 (2), 164-187.

Amsel, E., Klaczynski, P. A., Johnston, A., Bench, S., Close, J., Sadler, E. Walker, R. (2008). A dual-process account of the development of scientific reasoning: The nature and development of metacognitive intercession skills. Cognitive Development, 23, 452-471.

Andrews, G., & Halford, G. S. (1998). Children's ability to make transitive inferences: The importance of premise integration and structural complexity. Cognitive Development, 13, 479-513.

Bara, B. G., Bucciarelli, M., & Lombardo, V. (2001). Model theory of deduction: A unified computational approach. Cognitive Science, 25, 839-901.

Barrouillet, P. (1996). Transitive Inferences from set-inclusion relations and working memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22 (6), 1408-1422.

Barrouillet, P. (2011). Dual-process theories and cognitive development: Advances and challenges. Developmental Review, 31, 79-85.

Bond, A. B., Kamil, A. C., & Balda, R. P. (2003). Social complexity and transitive inference in corvids. Animal Behaviour, 65, 479-487.

Bouwmeester, S., & Sijtsma, K. (2004). Measuring the ability of transitive reasoning, using product and strategy information. Psychometrika, 69, 123-146.

Bouwmeester, S., & Sijtsma, K. (2006). Constructing a transitive reasoning test for 6- to 13-year-old children. European Journal of Psychological Assessment, 22 (4), 225-232.

Bouwmeester, S., & Sijtsma, K. (2007). Latent class modeling of phases in the development of transitive reasoning. Multivariate Behavioral Research, 43 (3), 457-480.

Boysson-Bardies, de B., & O'Regan, K. (1973). What children do in spite of adults' hypotheses. Nature, 246, 531-534.

Brainerd, C. J., & Kingma, J. (1984). Do children have to remember to reason? A fuzzy-trace theory of transitivity development. Developmental Review, 4, 311-377.

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. Developmental Review, 10, 3-47.

Brainerd, C. J., & Reyna, V. F. (1992). Explaining "memory free" reasoning. Psychological Science, 3, 332-339.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in reasoning, memory, and cognitive neuroscience. Advances in Child Development and Behavior, 28, 41-100.

Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. Developmental Review, 24, 396-439.

Breslow, L. (1981). Reevaluation of the literature on the development of transitive inferences. Psychological Bulletin, 89 (2), 325-351.

Browne, W. J., Caplen, G., Edgar, J., Wilson, L. R., & Nicol, C. J. (2010). Consistency transitivity and inter-relationships between measures of choice in environmental preference tests with chickens. Behavioural Processes, 83, 72-78.

Bryant, P. E. (1977). Logical inferences and development. In B. A. Geber (Ed.), Piaget and knowing. Studies in genetic epistemology. (pp.53-64). London: Routledge & Kegan Paul.

Bryant, P. (1998). Cognitive Development. In Eysenck, M. (Ed.), Psychology an integrated approach. London: Wesley Longman Limited.

Bryant, P. E., & Kopytynska, H. (1976). Spontaneous measurement by young children. Nature, 260, 773.

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. Nature, 232, 465-458.

Castle, K., & Needham, J. (2007). First graders' understanding of measurement. Early Childhood Education Journal, 35, 215-221.

Chapell, M. S., & Overton, W. F. (2002). Development of logical reasoning and the school performance of African American adolescents in relation to socioeconomic status, ethnic identity, and self-esteem. Journal of Black Psychology, 28 (4), 295-317.

Chapman, M. (1999). Constructivism and the problem of reality. Journal of Applied Developmental Psychology, 20 (1), 31-43.

Chapman, M., & Lindenberger, U. (1988). Functions, operations and decalage in the development of transitivity. Developmental Psychology, 24, 542-551.

Chapman, M., & Lindenberger, U. (1992). Transitivity judgments, memory for premises, and models of children's reasoning. Developmental Review, 12 (2), 124-163.

Cheney, D. L., & Seyfarth, R. M. (1990). How monkeys see the world: Inside the mind of another species. Chicago: University of Chicago Press.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. Psychological Review, 76 (4), 387-404.

Couvillon, P. A., & Bitterman, M. E. (1986). A conventional conditioning analysis of "transitive inference" in pigeons. Journal of Experimental Psychology: Animal Behavior Processes, 18 (3), 308-310.

Crone, E. A., Wendelken, C., Van Leijenhorst, L., Honomichl, R. D., Christoff, K., Bunge, S. A. (2009). Neurocognitive development of relational reasoning. Developmental Science, 12 (1), 55-66.

D'Amato, M. R., & Colombo, M. (1990). The symbolic distance effect in monkeys (Cebus Apella). Animal Learning and Behavior, 18, 133-140.

Daisley, J. N., Mascalzoni, E., Rosa-Salva, O., Rugani, R., & Regolin, L. (2009). Lateralization of social cognition in the domestic chicken (Gallus gallus). Philosophical Transactions of The Royal Society B, 364, 965-981.

Davis, H. (1992). Transitive inference in rats (rattus norvegicus). Journal of Comparative Psychology, 106, 342-349.

De Neys, W., & Dieussaert, K. (2005). Individual differences in rational thinking time. Proceedings of the Annual Conference of the Cognitive Science Society, 27, 577-582.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. Cognition, 106 (3), 1248-1299.

De Neys, W., & Vanderputte, K. (2011). When less is not always more: Stereotype knowledge and reasoning development. Developmental Psychology, 47 (2), 432-441.

DeVito L. M., Lykken, C., Kanter, B. R., & Eichenbaum, H. (2010). Prefrontal cortex: role in acquisition of overlapping associations and transitive inference. Learning & Memory, 17 (3), 161-167.

Dickins, T. E., & Dickins, D. W. (2001). Symbols, stimulus equivalence and the origins of language. Behavior and Philosophy, 29, 515-516.

Driscoll, I., Howard, S. R., Prusky, G. T., Rudy, J. W., & Sutherland, R. J. (2005). Seahorse wins all races: Hippocampus participates in both linear and non-linear visual discrimination learning. Behavioural Brain Research, 164, 29-35.

Dusek, J. A., & Eichenbaum, H. (1997). The hippocampus and memory for orderly stimulus relations. Proceedings of the National Academy of Sciences of the United States of America, 94 (13), 7109-7114.

Eichenbaum, H. (2001). The hippocampus and declarative memory: Cognitive mechanisms and neural codes. Behavioural Brain Research, 127, 199-207.

Ellenbogen, J. M., Hu, P. T., Payne, J. D., Titone, D., & Walker, M. P. (2007). Human relational memory requires time and sleep. PNAS, 104 (18), 7723-7728.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. Trends in Cognitive Sciences, 7 (10), 454-459.

Evans, J. St. B. T. (2011). Dual process theories in reasoning: Contemporary issues and developmental applications. Developmental Review, 31, 86-102.

Ferreira, M. B., Garcia-Marques, L., Sherman, S. J., & Sherman, J. W. (2006). Automatic and controlled components of judgment and decision making. Journal of Personality and Social Psychology, 91 (5), 797-813.

Fersen, L. von., Wynne, C. D. L., Delius, J. D., & Staddon, J. E. R. (1991). Transitive inference formation in pigeons. Journal of Experimental Psychology, Animal Behavior Processes, 17, 334-341.

Frank, M. J., Rudy, J. W., Levy, W. B., & O'Reilly, R. C. (2005). When logic fails: Implicit transitive inference in humans. Memory and Cognition, 33, 742-750.

Goel, V., Makale, M., & Grafman, J. (2004). The hippocampal system mediates logical reasoning about familiar spatial environments. Journal of Cognitive Neuroscience, 16 (4), 654-662.

Golde, M., von Cramon, D. Y., & Schubotz, R. I. (2010). Differential role of anterior prefrontal and premotor cortex in the processing of relational information. NeuroImage, 49, 2890-2900.

Goodwin, G. P., & Johnson-Laird, P. N. (2006). Reasoning about the relations between relations. The Quarterly Journal of Experimental Psychology, 59 (6), 1047-1069.

Greene, A. E., Fugelsang, J. A., Kraemer, D. J. M., Shamosh, N. A., & Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. Brain Research, 1096, 125-137.

Greene, A. J., Spellman, B., Dusek, J. A., Eichenbaum, H. B., & Levy, W. B. (2001). Relational learning with and without awareness: Transitive inference using nonverbal stimuli in humans. Memory and Cognition, 29, 893-902.

Grosenick, L., Clement, T. S., & Fernald, R. D. (2007). Fish can infer social rank by observation alone. Nature, 445, 429-432.

Halford, G. S., & Andrews, G. (2004). The development of deductive reasoning: How important is complexity? Thinking and Reasoning, 10 (2), 123-145.

Halford, G. S., Bain, J. D., & Maybery, M. T. (1984). Working memory and representational processes: Implications for cognitive development. In H. Bouma & D. G. Bouwhuis (Eds.), Attention and performance X. (pp. 459-470). Hillsdale, NJ: Erlbaum.

Halford, G. S., Maybery, M., & Bain, J. D. (1986). Capacity limitations in children's reasoning: A dual-task approach. Child Development, 57, 616-627.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. Behavioral and Brain Sciences, 21, 803-864.

Hanlon, F. M., Houck, J. M., Pyeatt, C. J., Lundy, S. L., Euler, M. J., Weisend, M. P., Thoma, R. J., Bustillo, J. R., Miller, G. A., & Tesche, C. D. (2011). Bilateral hippocampal dysfunction in schizophrenia. NeuroImage, 58, 1158-1168.

Holcomb, W. L., Stromer, R., & Mackay, H. A. (1997). Transitivity and emergent sequence performances in young children. Journal of Experimental Child Psychology, 65, 96-104.

Kallio, K. D. (1982). Developmental change on a five-term transitive inference. Journal of Experimental Child Psychology, 33, 142-164.

Kavale, K. A., Forness, S. R., & Lorsbach, T. C. (1995). Fuzzy-trace theory: Implications for research in learning disabilities. Learning and Individual Differences, 7 (2), 119-127.

Klaczynski, P. A. (2009). Cognitive and social cognitive development: Dual-process research and theory. In J. St. B. T. Evans & K. Frankish (Eds.), In two minds: Dual processes and beyond. (pp.265-292). Oxford, UK: Oxford University Press.

Lazareva, O. F., Smirnova, A. A., Zorina, Z. A., & Rayvesky, V. V. (2001). Hooded crows solve a transitive inference problem cognitively. Animal Welfare Supplement, 10, S219-S231.

Libben, M., & Titone, D. (2008). The role of awareness and working memory in human transitive inference. Behavioural Processes, 77, 43-54.

Lutkus, A., & Trabasso, T. (1974). Transitive inferences by preoperational, retarded adolescents. American Journal of Mental Deficiency, 78, 599-606.

MacLean, E. L., Merritt, D. J., & Brannon E. M. (2008). Social complexity predicts transitive reasoning in prosimian primates. Animal Behaviour, 76, 479-486.

Markovits, H., & Dumas, C. (1992). Can pigeons really make transitive inference? Journal of Experimental Psychology, Animal Behaviour Processes, 18, 311-312.

Markovits, H., & Dumas, C. (1999). Developmental patterns in the understanding of social and physical transitivity. Journal of Experimental Child Psychology, 73, 95-114.

Markovits, H., Dumas, C., & Malfait, N. (1995). Understanding transitivity of a spatial relationship, a developmental analysis. Journal of Experimental Child Psychology, 59, 124-141.

Markovits, H., & Thompson, V. (2008). Different developmental patterns of simple deductive and probabilistic inferential reasoning. Memory and Cognition, 36 (6), 1066-1078.

Markovits, H., Venet, M., Janveau-Brennan, G., Malfait, N., Pion, N., & Vadeboncoeur, I. (1996). Reasoning in young children: Fantasy and information retrieval. Child Development, 67, 2857-2872.

Martin, N., & Alsop, B. (2004). Transitive inference and awareness in humans. Behavioural Processes, 67, 157-165.

Maydak, M., Stromer, R., Mackay, H. A., & Stoddard, L. T. (1995). Stimulus classes in matching to sample and sequence production: The emergence of numeric relations. Research in Developmental Disabilities, 16 (3), 179-204.

McGonigle, B. O., & Chalmers, M. (1992). Monkeys are rational. Quarterly Journal of Experimental Psychology, 45B, 189-228.

Morsanyi, K., & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgement. Journal of Experimental Child Psychology, 99, 18-36.

Moses, S. N., Villate, C., & Ryan, J. D. (2006). An investigation of learning strategy underlying transitive inference performance in humans. Neuropsychologia, 44, 1370-1387.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgments of numerical inequality. Nature, 215, 1519–1520.

Nagode, J. C., & Pardo, J. V. (2002). Human hippocampal activation during transitive inference. NeuroReport, 13, 939-944.

Oakhill, J. (1984). Why children have difficulty reasoning with three-term series problems. British Journal of Developmental Psychology, 2, 223-230.

Padoa-Schippa, C., Jandolo, I., Visalberghi, E. (2006). Multi-stage mental process for economic choice in capuchins. Cognition, 99, B1-B13.

Pears, R., & Bryant, P. E. (1990). Transitive inference by young children about spatial position. British Journal of Psychology, 81, 497-510.

Perner, J., & Aebi, J. (1985). Feedback-dependent encoding of length series. British Journal of Developmental Psychology, 3, 133-141.

Perner, J., & Mansbridge, D. G. (1983). Developmental differences in encoding length series. Child Development, 54, 710-719.

Piaget, J. (1954). The construction of reality in the child. New York: Basic Books.

Piaget, J. (1970). Piaget's theory. In P. H. Mussen (Ed.), Carmichael's manual of child psychology. New York: Wiley.

Piaget, J., Grize, J. B., Szeminska, A., & Vinh Bang. (1977). Epistemology and psychology of functions. Dordrecht, The Netherlands: Reidel.

Piaget, J., Inhelder, B., & Szeminska, A. (1960). The child's conception of geometry (E. Lunzer, Trans.). London: Routledge & Kegan Paul.

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. PNAS, 104 (35), 13861-13867.

Rabinowitz, S. M., & Howe, M. L. (1994). Development of the middle concept. Journal of Experimental Child Psychology, 57 (3), 418-448.

Reyna, V. F., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A Fuzzy-trace model. Developmental Review, 31, 180-206.

Reyna, V. F., & Brainerd, C. J. (1990). Fuzzy processing in transitivity development. Annais of Operations Research, 23, 37-63.

Reyna, V. F., & Ellis, S. C. (1994). Fuzzy-trace theory and framing effects in children's risky decision making. Psychological Science, 5, 275-279.

Reyna, V. F., Estrada, S. M., DeMarinis, J. A., Myers, R. M., Stanisz, J. M., & Mills, B. A. (2011). Neurobiological and memory models of risky decision making in adolescents versus young adults. Journal of Experimental Psychology: Learning, Memory, and Cognition, 37 (5), 1125-1142.

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making: Implications for theory, practice, and public policy. Psychological Science in the public interest, 7, 1-44.

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instruction, and retention interval. Developmental Psychology, 30, 178-191.

Riley, C. A. (1976). The representation of comparative relations and the transitive inference task. Journal of Experimental Child Psychology, 22, 1-22.

Riley, C. A., & Trabasso, T. (1974). Comparatives, logical structure, and encoding in a transitive inference task. Journal of Experimental Child Psychology, 17, 187-203.

Russell, J., McCormack, T., Robinson, J., & Lillis, G. (1996). Logical (versus associative) performance on transitive inference tasks by children: Implications for the status of animals' performance. The Quarterly Journal of Experimental Psychology, 49B (3), 231-244.

Shohamy, D., Myers, C. E., Kalanithi, J., & Gluck, M. A. (2008). Basal ganglia and dopamine contributions to probabilistic category learning. Neuroscience Biohavioral Reviews, 32 (2), 219-236.

Siegal, M. (2003). Cognitive development. In A. Slater & G. Bremner (Eds.), An introduction to developmental psychology. (pp. 189-210). Oxford: Blackwell.

Siemann, M., & Delius, J. D. (1996). Influences of task concreteness upon transitive responding in humans. Psychological Research, 59, 81-93.

Smith, E. R., & Collins, E. C. (2009). Dual-process models: A social psychological perspective. In J. St. B. T. Evans & K. Frankish (Eds.), In two minds: Dual processes and beyond. (pp. 197-216). Oxford, UK: Oxford University Press.

Stanovich, K. E. (2009). What intelligence tests miss. The psychology of rational thought. New Haven and London: Yale University Press.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences, 23, 645-726.

Stromer, R., Mackay, H. A., Cohen, M., & Stoddard, L. T. (1993). Sequence learning in individuals with behavioural limitations. Journal of intellectual Disability Research, 37, 243-261.

Titone, D., Ditman, T., Holzman, P. S., Eichenbaum, H., & Levy, D. L. (2004). Transitive inference in Schizophrenia: Impairments in relational memory organization. Schizophrenia Research, 68, 235-247.

Trabasso, T. (1975). Representation, memory and reasoning: How do we make transitive inferences? In A. D. Pick (Ed.), Minnesota Symposia on Child Psychology. Minneapolis: University of Minnesota Press, Vol 9.

Trabasso, T. (1977). The role of memory as a system in making transitive inference. In R. V. Kail & J. W. Hagan (Eds.), Perspectives on the development of memory and cognition. (pp. 333-366). Hillsdale NJ: Erlbaum.

Trabasso, T., Riley, C. A., & Wilson, E. G. (1975). The representation of linear order and spatial strategies in reasoning. In R. Falmagne (Ed.), Reasoning, Representation and Process. Hillsdale NJ: Lawrence Erlbaum Associates.

Trabasso, T., van den Broek, P., & Suh, S. Y. (1989). Logical necessity and transitivity of causal relations in stories. Discourse Processes, 12, 1-25.

Van der Lely, H. K. J. (1997). Language and cognitive development in a grammatical sli boy: Modularity and innateness. Journal of Neurolinguistic 10 (2/3), 75-107.

Van Elzakker, M., O'Reilly, R. C., & Rudy, J. W. (2003). Transitivity, flexibility, conjunctive representations and the hippocampus. I. An empirical analysis. Hippocampus, 13, 334–340.

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. Behavioral Processes, 78, 313-334.

Verweij, A. C., Sijtsma, K., & Koops, W. (1996). A mokken scale for transitive reasoning suited for longitudinal research. International Journal of Behavioral Development, 19 (1), 219-238.

Verweij, A. C., Sijtsma, K., & Koops, W. (1999). An ordinal scale for transitive reasoning by means of a deductive strategy. International Journal of Behavioral Development, 23 (1), 241-264.

Viskontas, I. V., Holyoak, K. J., & Knowlton, B. J. (2005). Relational integration in older adults. Thinking and Reasoning, 11 (4), 390-410.

Wendelken, C., & Bunge, S. A. (2009). Transitive inference: Distinct contributions of rostrolateral prefrontal cortex and the hippocampus. Journal of Cognitive Neuroscience, 22 (5), 837-847.

Wittemyer, G., & Getz, W. M. (2007). Hierarchical dominance structure and social organisation in African elephants, Loxodonta africana. Animal Behaviour, 73, 671-681.

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex function: Processing and representational perspectives. Nature Reviews Neuroscience, 4, 139-147.

Wright, B. C. (2001). Re-conceptualising the transitive inference ability: A framework for research. Developmental Review, 21 (4), 375-422.

Wright, B. C. (Submitted). A missing comparison in transitive reasoning research.

Wright, B. C. (1998a). Psychological mechanisms of logical transitive inference in adults and children. Doctoral Dissertation: University of Oxford, England.

Wright, B. C. (1998b). And if the developmental data doesn't quite fit… Behavioral and Brain Sciences, 21, 847-848.

Wright, B. C. (2001). Re-conceptualising the transitive inference ability: A framework for research. Developmental Review, 21 (4), 375-422.

Wright, B. C. (2006a). On the emergence of the discriminative mode for transitive-Inference. European Journal of Cognitive Psychology, 18 (5), 776-800.

Wright, B. C. (2006b). The information processing task revisited: Investigating profiles from the start to the end of training. Thinking and Reasoning, 12 (1), 91-123.

Wright, B. C., & Dowker, A. D. (2002). The role of cues to differential absolute size in children's transitive inferences. Journal of Experimental Child Psychology, 81, 249-275.

Wright, B. C., & Howells, D. (2008). Getting one step closer to deduction: Introducing an alternative paradigm for transitive inference. Thinking and Reasoning, 14 (3), 244-280.

Wright, B. C., Robertson, S., & Hadfield, L. (2011). Transitivity for height versus speed: To what extent do the under-7s really have a transitive capacity? Thinking and Reasoning, 17 (1), 57-81.

Wynne, C. D. L. (1998). A minimal model of transitive inference. In C. D. L. Wynne & J. E. R. Staddon (Eds.), Models for action: Mechanisms for adaptive behaviour (pp. 269-307). Hillsdale, NJ: Erlbaum.

Yamazaki, Y. (2004). Logical and illogical behavior in animals. Japanese Psychological Research, 46 (3), 195-206.

Figure Captions

Figure 1:

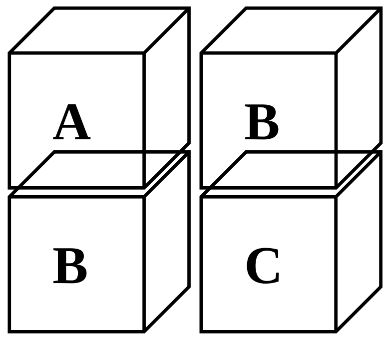
Illustration of the transitive block task, using three-term examples. Top = depiction of Pears & Bryant (1990) task. Bottom = depiction of one of the four conditions of Markovits et al. (1995) task. The Pears & Bryant task can be solved using the non-logical (absolute) strategy rebuked by Bryant & Trabasso (1971). The Markovits et al. task guards against that strategy.

Figure 2:

Depiction of differences between critical transitive B:D performance on various independently conducted five-term extensive-training studies. Comparisons suggest several anomalies. The paradigm is therefore questionable. Note, 4 years = Bryant & Trabasso (1971), 6 years = Russell et al. (1996), adults = Greene et al. (2001), adults LD = Maydak et al. (1995), monkeys = McGonigle & Chalmers (1992).

Figure 1:

Pears & Bryant Task
The A:C comparison can be worked out by coordinating around B, but also non-inferentially by simply noting which of A and C satisfies the description "is high".



Markovits et al. Task
A:C can be deduced inferentially as before, but this time the alternative absolute height strategy favours neither A nor indeed C.

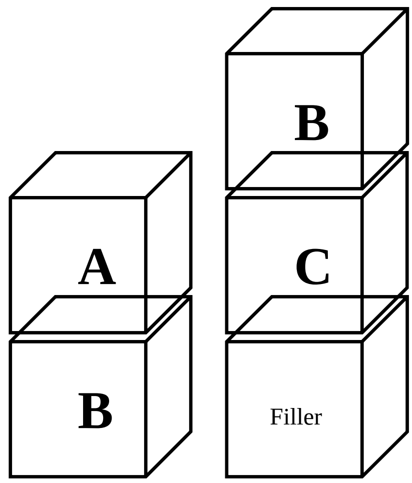
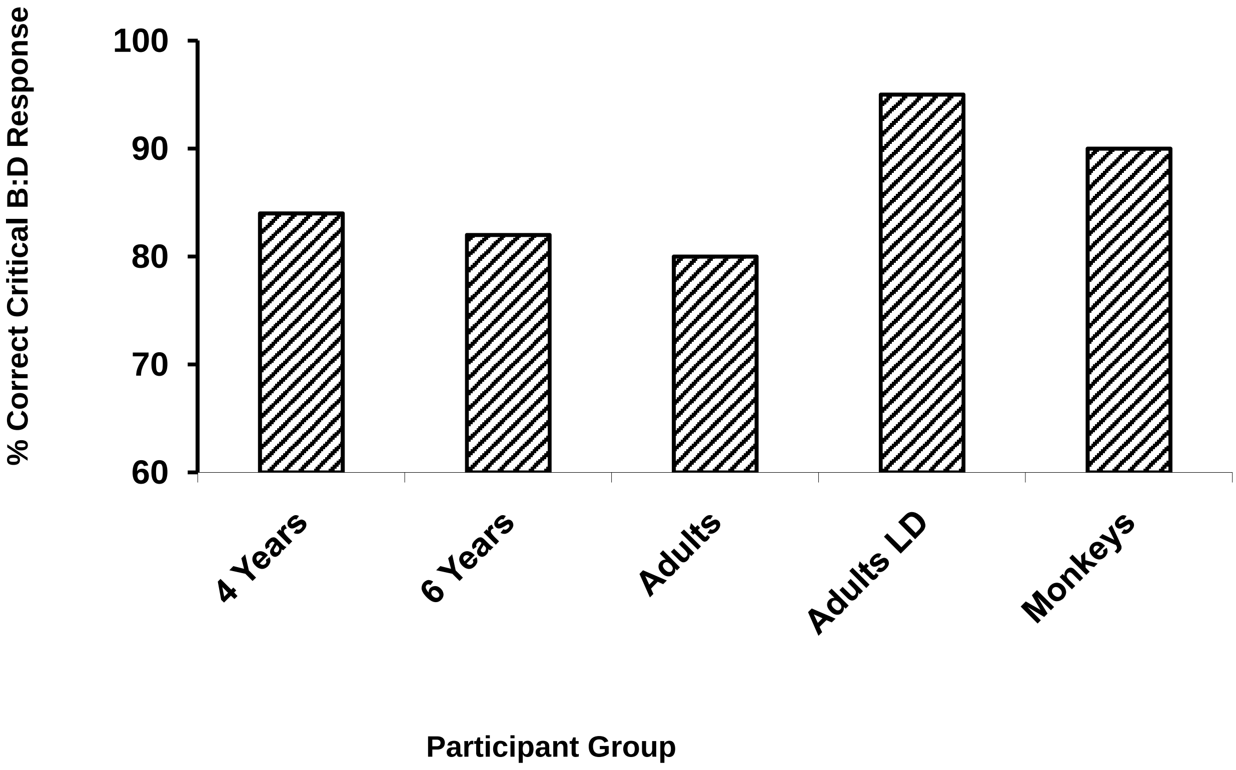


Figure 2:


% Correct Critical B:D Response
100
90
80
70
60
4 Years
6 Years
Adults
Adults LD
Monkeys
Participant Group