Running Head: UNDER-7S CAN MAKE TRANSITIVE REASONING

Transitivity for Height Versus Speed: To What Extent do the Under-7s Really Have a Transitive Capacity?

Barlow C. Wright

Department of Psychology, Brunel University, Uxbridge, UK

Suzanne Robertson

Department of Psychology, Brunel University, Uxbridge, UK

Lucy Hadfield

Department of Psychology, Brunel University, Uxbridge, UK

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

B. C. Wright, Department of Psychology, Brunel University, Uxbridge, UB8 3PH, England. E-mail: Barlow.wright@brunel.ac.uk.

Abstract

Transitive inference underpins many human reasoning competencies. The dominant task (the "Extensive-Training-Paradigm") employs many items and large amounts of training, instilling an ordered series in the reasoner's mind. But findings from an alternative "3-Term-Paradigm" suggest transitivity is not present until 7+ years. Interestingly, a second alternative paradigm (the "Spatial-Task"), using simultaneously-displayed height relationships to form premise-pairs, can uphold the 4 year estimate. However, this paradigm risks cuing children and hence is problematic. We investigated whether a height-task variant might correspond to a more ecologically-valid 3-term task. 222 4-6 year-olds either completed a modified height task, including an increased familiarisation phase, or a computer-animated task about cartoon characters running a race in pairs. Findings confirmed both tasks were functionally identical. Crucially, 4 year-olds were at chance on both; whereas 6 year-olds performed competently. These findings contrast with estimates from all three paradigms considered. A theoretical evaluation of our tasks and procedures against previous ones, leads us to 2 conclusions. First, our estimate slightly amends the 7-year estimate offered by the 3-Term-Paradigm, with the difference explained in terms of its greater relevance to child experiences. Second, our estimate can coexist alongside the 4-year estimate from the Extensive-Training-Paradigm. This is because, applying a recently developed "Dual-Process" conception of reasoning, anticipates that extensive-training benefits a species-general Associative System, whilst the Spatial-Paradigm and 3-Term-Paradigm can potentially index a genuinely-Deductive System which has always been the target of transitive research.

Key Words: Children's Reasoning; Dual-Process Theory; Height Task; Spatial Reasoning; Transitive reasoning

Transitivity for Height Versus Speed: To What Extent do the Under-7s Really Have a Transitive Capacity?

Deduction allows us to engage in quite complex problem-solving activities, but may sometimes form part of more routine unconscious or automatic thinking (Deneault & Ricard, 2006; Klaczynski, 2009; Muller, Sokol & Overton, 1999). Perhaps one of the most basic forms of deductive inference-making both in adults and in children is Transitive Reasoning (Bara, Bucciarelli & Lombardo, 2000; Halford, Wilson & Phillips, 1998; Lazareva & Wasserman, 2010). A child is said to possess transitive reasoning when he or she can deduce a latent relationship between two items (say A and C), after being given information about the relationship of each of these items to a third item (B) that just happens to be intermediate between the other two in some respect (Lee & Freire, 2003; Rabinowitz, Grant, Howe & Walsh, 1994).

Our transitive reasoning capacity is said to be important for acquiring many cognitive concepts (see Krackhardt & Kilduff, 1999). It is also an integral part of children's educational progression, such that "in history class, one has to determine which event or personality preceded another; in geography, altitudes are represented by different colors; in geometry, measurement operations are performed;..." (Artman & Cahan, 1993, pp.753). Transitivity even features in social and interpersonal contexts, such as in assessing possible health risks, gambles and friendships (Birenbaum & Gutierrez, 2007; Markovits & Dumas, 1999; Reyna & Farley, 2006). Initially, its importance to child cognition was very much a Piagetian pursuit (e.g., Piaget & Inhelder, 1967; Piaget, Inhelder & Szeminska, 1960; see also Muller et al., 1999 for a sound introduction to Piaget's framework for reasoning research). For Piaget, transitive reasoning was a matter of logical deduction. Hence his basic task for assessing it relied on three-term tasks as featured in syllogistic reasoning generally (“linear syllogistic reasoning” - Sternberg, 1980).

Following recommendations from theorists working with the psychology of logic (e.g., Clark, 1969), Piaget gave his transitivity tasks via the minimum "two" premises, each containing 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; Hooper, Toniolo & Sipple, 1978;

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 required to deduce the inference between items A and C (Hong & Chond, 2001; Sternberg, 1980).

According to Piaget and colleagues, three-term tasks revealed that children have a well developed deductive transitive capacity by 7 or 8 years. But as with so many Piagetian claims, this age estimate was aggressively challenged. In this instance, Bryant and Trabasso (1971) reported that the correct age estimate is 4 years (for similar views see Russell et al., 1996). In support of their claims, Bryant and Trabasso cited findings from a task that was rather different from the three-term paradigm, which we briefly outline below (see Extensive-Training-Paradigm). For now it is sufficient to note that questions were raised about the validity of both paradigms for assessing their target deductive transitive capacity (e.g., contrast Halford & Galloway, 1977; Riley & Trabasso, 1974; Siegal, 2003 with Bryant, 1998; Russell et al., 1996). The present study aimed to resolve this issue by offering methodological improvements to three-term tasks and placing the resultant findings within a conception of human reasoning that seems capable of accepting the three-term findings and the findings from advocates of the Bryant and Trabasso paradigm at the same time (Dual-Process theory - e.g., see Evans, 2009; Klaczynski, 2009; Markovits & Thompson, 2008).

Bryant and Trabasso (1971) raised two issues about the three-term task. Firstly an issue of false-positives: A child not having deductive transitive reasoning can still give "right answers" via simply repeating the labels given to him/her (Bryant, 1998). The second issue is an issue of false-negatives: Here, three-term tasks do not establish the extent to which the child has retained the premises, and so there is the possibility that poor transitive responses result from poor memory rather than from poor inferential reasoning (for critique regarding this issue see Brainerd & Reyna, 1992; Siegal, 2003). However, it is important to note that when these theoretical concerns were eventually empirically tested, they were shown not to be factors after all (e.g., see Lazareva & Wasserman, 2010; Wright & Dowker, 2002). Nevertheless, in response to the false-positives as they conceived them, Bryant and Trabasso (1971) had advocated five as the minimum number of items (A> B> C> D> E). This ensured that the three central items (B, C and D) did not receive an unique label, and hence children could not give correct answers by simply parroting back to the experimenter the uniquely positive item (item-A in their five-term length series). The critical premises in this task were now B:C and C:D, and "it therefore follows that the crucial transitive comparison will be between B and D" (Bryant & Trabasso, 1971, pp.457).

To prevent false-negatives, Bryant and Trabasso simply began intensively training the children on the four premise pairs needed for their five-term task, in order from largest to smallest pair, or the converse. "Half of the children in each group were trained with each order." (Bryant & Trabasso, 1971, pp.457). This training regime is arguably the most critical aspect of the task (Markovits & Dumas, 1992; Siegal, 2003), and it already subsumes the fact that series sizes are large (typically five or more items – Frank et al., 2005; Halford & Andrews, 2004; Martin & Alsop, 2004; Whelan, Barnes-Holmes & Dymond, 2006). To keep this critical (training) aspect salient, Yamazaki (2004) suggests we call this paradigm the "Extensive-Training-Paradigm".

Unfortunately, Bryant and Trabasso gave no details about how long they trained each child for, and so one cannot definitively comment on whether their training regime inadvertently bypassed any need for the children to deduce the required inferences. However, we note that in a replication conducted by Trabasso himself (Riley & Trabasso, 1974, experiment 2), 4 year-olds needed 4 days training, before they reached criterion for memory. Other theorists have confirmed that training can take days or even weeks with some child/adult groups (Berens & Hayes, 2007; Holcomb, Stromer & Mackay, 1997; Kallio, 1982; Stromer et al., 1993). Notwithstanding training issues, Bryant and Trabasso reported that, after extensive-training on the four premises, 4 year-olds pass their task at 78% for the critical B:D inference. From this finding, Bryant later concluded that "young children, even before they go to school, can (i) make a deductive inference; (ii) make a spatial inference; (iii) make a transitive inference" (Bryant, 1977, pp.63).

However, in the closest replication to date, Riley and Trabasso (1974, experiment 2) found that, even after several days of training, 4 year-olds did not pass the task reliably above chance (see also their experiment 3 for a further failed demonstration). There seem no direct empirical attempts to dispel the doubts raised by Trabasso and other theorists, about 4 year-olds on extensive-training tasks (Siegal, 2003; Wright, 1998). Instead, the general paradigm is simply assumed to assess its target phenomenon. But this same paradigm routinely reports deductive transitive reasoning even in child and adult groups with various impairments or disorders (e.g., Titone, Ditman, Holzman, Eichenbaum & Levy, 2004). For example, Stromer et al. (1993) successfully trained participants with serious learning difficulties, and showed that it was possible to obtain apparently-transitive responses between two five-term transitive series that had not actually been connected in any way. Even infant non-humans apparently reason transitively. 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. This latter example highlights that, with enough training, virtually any species tested with the Extensive-Training-Paradigm might demonstrate transitivity (Grosenick, Clement & Fernald, 2007; Shafir, Waite & Smith, 2002; Wittemyer & Getz, 2007).

We surmise that, the finding of apparently-transitive responses is not necessarily proof that reasoners (human or non-human) are doing so by applying logical deduction (for similar arguments see also Markovits & Dumas, 1992; Premack, 2007; Russell et al., 1996; Van Elzakker, O’Reilly & Rudy, 2003). In the context of 4 year-olds, perhaps Trabasso best summarises this contention: "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).

Given the disputed nature of the Extensive-Training-Paradigm, it is prudent scientific practice to use Bryant and Trabasso's points about the three-term task in order to improve that paradigm, instead of solely relying on the Extensive-Training-Paradigm (Markovits & Dumas, 1992, 1999). In one three-term study, Wright (2006a) showed 5 to 7 year-olds objects (rods, spheres & cubes) with such small differences in size that their relative sizes were quite hard to discern when two objects were held at arms length (around 2m apart but also around 3m away from the child). Once children correctly encoded which was bigger within each of two pairs, they were asked to restate the two premises for themselves and then deduce and report the biggest object. Afterwards, they again stated the two premises. Although memory before and after giving the transitive inference was better than for similar premise pairs in the Extensive-Training-Paradigm (greater than 90% – Holcomb et al., 1997), 5 year-olds only gave the correct solution an average of 40% of the time, with 6 year-olds at 49% and 7 year-olds at 65% (see also Markovits & Dumas, 1999) for an even more ecologically valid demonstration with everyday items (e.g., combs).

Given the difficulty reconciling conclusions from such three-term tasks (e.g., see Markovits & Dumas, 1999; Oakhill, 1984) with the Extensive-Training-Paradigm (e.g., Bryant & Trabasso, 1971; Russell et al., 1996), it was prudent to try to resolve the controversy by relying on a new paradigm. In 1988, Kallio introduced a spatial paradigm using the transitive relations of heights (of rockets in the sky) and speed (planes racing each other). He also managed two manipulations perhaps for the first time. First, he presented a total of six items all at one time, so that all were simultaneously in view. Thus, here was a task using even more than five items, yet it had no need for training (see also Wright & Howells, 2008). Second, Kallio used an experimental procedure that allowed children to focus on only three items out of the six on display. On introducing these design features with 5 to 10 year-olds, Kallio found that children did poorly on transitive questions until around 8 years. This finding is in close agreement with three-term studies (e.g., Markovits & Dumas, 1999; cf. Oakhill, 1984; Wright, 2006a).

What was essentially a simplified variant on Kallio's design was subsequently employed by Pears and Bryant (1990). The main difference was the use of three-dimensional blocks which could be manipulated by the children, instead of them having to reason about drawings of rockets/planes as in the Kallio study. Pears and Bryant (1990) showed 24 4 year-olds small piles of blocks on a table, with two blocks stacked one on the other forming each pile. The transitive series were between four and six-terms but we explain using the minimum three-term format (see Figure 1). For example, one pile might consist of a blue block stacked on top of a red one (say A above B). In another pile, there might be a red block on top of a green block (say B above C). The child was asked to state which out of the blue block and the green block (i.e., A and C) would be higher up, if they had to build a new pile out of just these two blocks. Findings upheld the conclusions from Bryant and Trabasso, with 4 year-olds apparently showing a well developed capacity for transitive reasoning. For example, for the critical B:D trial of Pears and Bryant's five-term series, we calculated performance at 71% across their two experiments.

Insert Figure 1 about here

The Pears and Bryant task was not conclusive, because correct answers could be given by the child simply reporting whichever of the blocks actually sits on top of its own pile, whilst ignoring the other pile ("Supportive-Cue" - Markovits, Dumas & Malfait, 1995). For example, when asked to deduce the highest of A and C, all the child needs to do is note that A is on top of pile A:B, and hence report block-A as his/her answer (see Figure 1 for depiction of the block piles and the non-logical shortcut strategy). To avoid children being "cued" to the correct answer via the absolute height position of item-A, Markovits et al. introduced a filler-block. Essentially, they varied the relative heights of A and C by strategically putting white-filler-blocks below either pile A:B or pile B:C. Now, with one filler-block below the B:C pile, the height cue is neutralised ("Neutral-Cue" condition). Indeed, with two filler-blocks added to B:C, the height cue actually militates against the transitive reasoning answer, by making item-C most salient when what the reasoner needs to do is ignore this and give the usual answer of A ("Conflict-Cue" condition).

In Markovits et al.'s study, 4 and 6 year-olds simply followed the height cue: They did not have deductive-transitive-inference after all. When we calculated overall performance restricted to the three-term series, whilst weighting the supportive-cue, neutral-cue and conflict-cue conditions equally, it emerged that the 6 year-olds (52%) were quite indistinguishable from the 4 year-olds (54%). Only at age 8 years did children now consistently choose the correct item (A) regardless of whether filler-blocks were used or not (overall three-term performance = 76%). Ameel, Verschueren and Schaeken (2007) confirmed similar levels for 8 year-olds in two separate experiments, but also concluded that deductive-transitive-inference development is still not complete at age 8 (Kallio, 1988; Rabinowitz et al., 1994; see also Markovits & Thompson, 2008).

However, just as Markovits et al. (1995) had argued a methodological flaw in Pears and Bryant’s study, Ameel et al. (2007) argued there may have been a methodological issue with Markovits et al.’s study. Basically, children in the Markovits et al. (1995) study might have been slightly confused about the role of the white-filler-blocks. For example, some may have treated them as part of the transitive series. But when Ameel et al. gave children practice trials in a classroom setting, 8 year-olds did no better in a less confusable "multi-coloured-filler-block" condition than in the original white-filler-block condition. They also did no better even when the task was presented in the context of flats in buildings (e.g., A:B as upper flat v lower flat). So Ameel et al.'s (2007) findings do not seem to challenge those of Markovits et al. (1995) after all.

It is important here to distinguish "practice" (e.g., as used by Ameel et al., 2007) from "training" (e.g., as used by Bryant & Trabasso, 1971; Russell et al., 1996; see also Berens & Hayes, 2007 for an example of training with two- and three-term problems). In the Extensive-Training-Paradigm, the training is used deliberately in order to instil the entire transitive series in memory (Breslow, 1981). Put another way, training is used to give the participant a head start on the exact problems s/he will face during test trials. This is in stark contrast with practice, which is used to familiarise the participant with the objects and procedures to be used in test trials; but deliberately avoids giving any assistance whatsoever in regard to any transitive series that will be tested. Using increased practice (not training) on the nature of the task and the role of filler-blocks before testing, ensuring older children receive the same amount of training as 4 year-olds, and giving practice on an individual basis instead of collectively within a large group (contrast Bryant & Trabasso, 1971; Ameel et al., 2007; Markovits et al., 1995), should allow children's true transitive capacity to come through.

The aims of the present study were fourfold.

1, We aimed to find out whether paying closer attention to procedural factors that might have led to children doing less well on the Markovits et al. (1995) and Ameel et al. (2007) height tasks, would now result in findings more in line with those of Pears and Bryant (1990) and Bryant and Trabasso (1971). Clearly it is time to revisit the performance particularly of 4 year-olds. But in order to avoid unduly loading our young participants, we restricted the task to three-term series.

2, Should our findings differ either from those of Bryant and colleagues or Markovits and colleagues, we also aimed to evaluate our methodological improvements to three-term tasks and place our findings in the context of a conception of human reasoning (i.e., Dual-Process theory) that has already been noted as capable of accepting the classical three-term findings but also simultaneously the findings from the Extensive-Training-Paradigm (e.g., see Bouwmeester, Vermunt & Sijtsma, 2007; Evans, 2003; Wright & Howells, 2008).

3, Extending Kallio (1988), we aimed to present arguably the first study that made a direct comparison of a three-term height task with a more classical three-term task using dynamic motion. This was also relevant to the issue of whether different transitive tasks (here, height v racing) tend to lead to very different response profiles (e.g., see the location v length tasks of Russell et al., 1996). On this issue, some theorists argue that the different age estimates for acquisition of transitive reasoning may stem as much from the nature of the items and relations being reasoned about (i.e., the "content") as from the validity or ecological validity of the tasks used themselves (Bouwmeester & Sijtsma, 2006; Wright, 2001).

4, Finally, we wanted to use a paradigm offering specifically 4 year-olds every opportunity to show their transitive deductive capacity but which did not risk giving them "non-logical" cues to the correct answers (Riley & Trabasso, 1974; Wright & Dowker, 2002). Additionally, we wanted to assess the view emerging from virtually every relevant study of the past 20 years, that deductive transitive reasoning does not reach a level indicative of competence until after 6 years (Ameel et al., 2007; Artman & Cahan, 1993; Markovits et al., 1995; Markovits & Dumas, 1999; Rabinowitz et al., 1994; Wright, 2006a).

Method

Participants

These were 222 children from 6 primary schools in the south of England, with a roughly equal split of girls and boys in each group and sub-group below. All children were typically-developing and had English as their only language or first language. Half of each gender were allocated to the height task or to the racing task, with each group further divided into 4, 5 and 6 year-olds. For the height task there were a total of 95 children - 31 in our group of 4 year-olds (Mean = 4.8 years, SD = 0.359), 24 were 5 year-olds (Mean = 5.7 years, SD = 0.341) and 40 were 6 year-olds (Mean = 6.5 years, SD = 0.316). For the racing task there were 127 children - 42 were 4 year-olds (Mean = 4.9 years, SD = 0.317), 42 were 5 year-olds (Mean = 5.6 years, SD = 0.301) and 43 were 6 year-olds (Mean = 6.5 years, SD = 0.359).

Materials

The height task used three-dimensional wooden cylinders 5cm high and with a diameter of 4cm. These were painted brown, grey, blue, red, light green, yellow and pink. There were three cylinders of each colour. The filler-blocks were also cylinders but these were white in colour.

The racing task was given on a portable Dell computer with a 1.83 GHz Core 2 Duo T5550 processor and 15.4 inch screen. The task was presented within Adobe Flash Player 9.0. With each participant's consent, testing was recorded with an Olympus VN6500PC Digital Voice Recorder.

Design

The basic design was between-subjects, with two tasks (height v racing) and three age groups (4, 5 and 6 years). Each task lasted for four trials. For the height task, there was one trial of each of the following types. One trial used the coloured blocks, but placed a white-filler-block under each coloured pile (i.e., under the A:B pile and also under the B:C pile). One trial used the coloured blocks but added one filler-block under the A:B pile, in order to raise A two blocks higher than C. A further trial again added only one filler-block but this time it was under the B:C pile, in order to render A and C of the same height. The last trial-type utilised two filler-blocks under the B:C pile, in order to raise C so that it was now one block higher up than A. The first two trial-types here, if anything, would cue the participant towards answering the transitive question about A and C, with block-A, whilst the last two of the trial-types would either give no cue or cue away from answering A. For the racing task, all three items (A, B and C) were seen side-by-side on the computer screen, about to have a race. A and B were seen, then B with C, and then the A:C question was asked.

Procedure

Participants were tested in a quiet but familiar room in their school. Ethical approval and parental consent were gained prior to testing and children were continually monitored to ensure they were not stressed and remained eager to take part. Children were given one task only. For the height task, there were first 4 practice trials with the Pears and Bryant procedure (i.e., no filler-blocks). This trial type was not used in test trials. Upon demonstrating familiarity with the coloured blocks, the test context and procedure, the child was then introduced to the role of the filler-block under pile B:C, via one further practice trial. If there was any confusion about the filler-block, a further three trials were to be given. However, no child showed confusion and so the number of practice trials was five for each child. In essence these practice trials constituted "warm-up problems" rather than training (for first use of such trials see Hooper et al., 1978).

In each trial, two piles representing premises A:B and B:C were set up about 20cm apart on the table. The piles were roughly the same distance in front of the child. Around half the time, A:B was to the left of B:C and for the remainder it was on the right. Also, for each side-to-side order of premises, on around one-third of occasions the A:B pile was up to 5cm nearer the child, and around one-third of the time it was around 5cm further away from the child. Such controls have recently been advocated to ensure that the absolute positions and spatial order of items cannot systematically benefit performance (Wright & Dowker, 2002). Two blocks, one of colours A and C respectively, were then placed between the two piles but closer to the child. It should be noted that in this task and also the racing task explained next, we followed protocols introduced by Wright and Dowker (2002), which made sure that the experimenter never introduced any colour, any other label or indeed any information to the child. Instead, the child was asked about the colours of the blocks forming each pile, and answered by narrating on their colours and respective height-positions within their respective piles. The child was now asked to make a third pile with two of the loose blocks A and C. However, it was impressed upon the child that the aim was to make sure that the height relationships within the first two piles was honoured, when deciding on block placement for their own A and C blocks:

“So that one (experimenter pointing to colour A in the loose block) was on top of that one (pointing to loose block B and then referring the child to pile A:B). And that one (loose block B again) was on top of that one (pointing to loose block C then to pile B:C). And what we need to do is make a new pile with these two (pointing to loose blocks A and C several times) which way do we have to put them?”

The child attempted this and was given feedback. Each new trial used different colours and/or different orders of colours and positions. In trial five which involved a filler-block, the child was informed that the white-filler-block was just there to bump up one of the two piles, but does not actually change what the child has to do with the loose blocks. Note, the absence of a white block in the child’s own cluster of two coloured blocks helped to reinforce the neutral status of the filler-block. The child took loose blocks A and C as before and placed one on top of the other forming the required new pile. In any practice trial, if A and C were not ordered correctly then this was explained to the child, who was then encouraged to correct the order.

Upon completing the five practice trials, the task proper was then begun. The task comprised four trials without feedback. If we use 1 to denote the presence of one filler-block and 0 to denote its absence, and we refer to A:B then B:C regardless of which side each had actually been displayed, then the four test trials were 11, 10, 01 and 02 respectively. These trials were given in a random order, and are depicted in Figure 2. The task took around 15 minutes including briefing, debriefing, instructions and warm-up trials.

Insert Figure 2 about here

For the racing task, children viewed short cartoons containing dynamically-life-like movements of cartoon characters. These were presented on computer in full-screen mode, and were set within a three-dimensional scene presented on a two-dimensional screen. Markovits and Thompson (2008) have previously used cartoon characters in developmental reasoning research. Here, one cartoon was of three boys, another of butterflies and a third of girls. The fourth trial was a repeat of one of the first three, with added verbal justification of whatever response the child gave. The first three trials were given in random order, with the fourth selected randomly from the first two.

To use the trial with boys as an example, the boys were dressed in separate colours (red, blue and green). This was shown in three stages. First the children saw the red boy racing the blue boy (the red boy won) and at this point children were asked who was in the race and who won. Next they saw the blue boy racing the green boy (the blue boy won) and again they were asked to confirm who was in the race and who won. They were then told that the red boy would be racing the green boy next and were asked who they thought would win and if appropriate, why? These answers were recorded and the children were then shown the final race. Thus, the child was given feedback regarding response correctness, directly after giving each A:C response. This task took around 10 minutes in all, and hence more children could be tested in the available time. Children were thanked for their participation and were returned to their classroom. For both the height task and the racing task, a child was given a mark of 1 if the answer on a trial had been correct and 0 if incorrect. The cumulative score was taken as a fair estimate of transitive reasoning capacity.

Results

In the racing task, the difficulty level of individual trials should be the same, whereas in the height task the four test trials used filler-blocks in different numbers or positions. Therefore, as a prelude to the main analyses below, we considered variations in proportion of overall correct responses on the height task. Proportions were 0.57 for trial 01, 0.62 for trial 11, 0.66 for trial 10, and 0.65 for trial 02. Thus, there was a tendency for trial 01 to be hardest. In that trial, one filler was below pile B:C to make A and C appear the same height in absolute terms (i.e., no height cue either for or against the correct answer). There was a slight tendency for performance to improve with a more helpful cue (i.e., trial 10 > trial 11 > trial 01 - see Ameel et al., 2007; Markovits et al., 1995). The only trial not neatly fitting this picture was trial 02. This would seem to indicate a tendency for the controls we implemented here (e.g., providing filler-blocks but ensuring the child did not use them in his/her answers), if anything, appearing to make the children think slightly more carefully about the inferable relationship between A and C (see Markovits et al., 1995 on 8 year-olds). Importantly, when we analysed these four trial-types statistically, we found no significant difference between them (Cochran's Q test - $N = 95$, $df = 3$, $Q = 3.178$, $p = 0.365$). Clearly as there was no overall reliable difference between individual trial types, it is acceptable for us to combine our trials into one cumulative transitive score.

The cumulative scores were analysed according to task (height v racing) and age group (4 v 5 v 6 years). All tasks were two-tailed unless stated otherwise. A summary of the data is provided in Table 1. Table 1 shows that overall the performances on the two tasks were quite similar. However, when either task was viewed according to the different age groups, the 6 year-olds did much better than the 5 year-olds with 4 year-olds doing least well.

Insert Table 1 about here

The trends shown in Table 1 were assessed using a two-way Analysis of Variance (ANOVA) with factors of task (height v racing) and age group (4 v 5 v 6 years). This confirmed that the rather slender overall difference between the two tasks was not reliable, as shown by a main effect that was not statistically significant (F (1, 216) = 0.005, $p = 0.946$, Obs.Power = 0.051). The slender advantage of either task over the other did not alter significantly from any age group to the next, as indicated by a non-significant interaction effect (F (2, 216) = 0.036, $p = 0.964$, Obs.Power = 0.055). However, the overall superior performance with increasing age group was statistically significant (F (2, 216) = 9.620, $p < 0.001$, Obs.Power = 0.980). The observed power being near its maximum of 1.000, is an indication that our age group differences had not arisen just because of the specific sample sizes used here.

As children had to make a decision based on placement of two blocks rather than all three blocks (as in most transitive research), ours was a two-alternative forced-choice design, and hence chance performance was 50%. We see from Table 1 that 4 year-old children were just above the chance level for each task and for the two tasks combined. In order to determine whether this slender above chance performance was nevertheless robust, a one-sample t-test was carried out for all the 4 year-olds (i.e., both tasks pooled). This showed that the roughly 2%

overall difference was not statistically significant (t (72) = 0.466, p = 0.643). A similar test for 5 year-olds showed that they were reliably above chance (t (65) = 3.566, p = 0.001), as were the 6 year-olds (t (82) = 9.119, p < 0.001).

Wright (2006a) suggests that the notion of a competence threshold might be useful in developmental research on transitive reasoning. Borrowing the concept from general psychophysical designs, he asserted that the competence threshold should be defined as that level laying as close to perfect performance as it is to chance performance. For our two-alternative forced-choice design this is the 75% level. A competence threshold was useful here for two main reasons. First, it would allow us to determine whether any group's performance that was significantly above chance, might still need to be viewed with a certain amount of caution. Second, it would allow us to determine whether any group did so well that it tended to give correct answers more often than answers that might indicate the use of a guessing strategy (i.e., beginning to approach perfect performance).

However, it could be argued that considering the number of children reaching criterion is preferable to considering the group means. We therefore counted the numbers of children in each group reaching the 75% level and converted these into percentages. For the 4 year-olds across both tasks, we obtained 39%. For the 5 year-olds we obtained 55%, and for the 6 year-olds this was 71%. As these values are not as favourable as the group-mean values shown in Table 1, we focus our remaining statistical analyses on the group means across both tasks.

Next, three t-tests analogous to those outlined earlier were conducted against the 75% level (mean per child of 3 out of 4 correct responses). The 4 year-olds were significantly below the competence threshold (t (72) = -5.204, p < 0.001). The 5 year-olds were also significantly below this threshold (t (65) = -2.972, p = 0.004). However, the overall difference between the 6 year-olds' correct responses and the 75% level was less than 0.5% and consequently this difference was not statistically significant (t (82) = -0.225, p = 0.822). We conducted one further test to confirm whether or not the 6 year-olds' performance was reliably less than perfect performance (i.e., 4 out of 4). This confirmed their level of performance was significantly below perfect performance (t (82) = -9.569, $p < 0.001$).

Discussion

The findings of the present study showed that paying closer attention to procedural factors led to an age estimate for transitive reasoning that is neither in full agreement with advocates of the Extensive-Training-Paradigm (e.g., Pears & Bryant, 1990) nor the three-term alternatives (e.g., Markovits & Dumas, 1999); although much closer to the latter perspective. Our findings were identical for two very different tasks, suggesting that content may not be as big a problem for developmental transitive studies as previously thought. We also revealed a more gradual developmental profile for transitive reasoning, with it not being fully developed even at 6.5 years. We now discuss each of these issues, plus the apparent disparity between the Extensive-Training-Paradigm and its main alternatives. We show below that the difference between our own findings and the Extensive-Training-Paradigm can readily be accounted for by Dual-Process theory.

We believe our height task represents an ecologically-valid demonstration of transitivity in 4 to 6 year-olds. This is because, although it might be regarded as being quite abstract stemming from it concerning coloured blocks, it is correct to say that children gain experience of playing with blocks just like ours from a very early age (a claim already made by Pears & Bryant, 1990). Also, children have much experience of watching cartoons, and have no difficulty at all mapping real life attributes (e.g., a desire to win) onto dynamically-moving cartoon characters. Thus, we contend that our racing task and our height task were of equal or indeed higher real-world relevance (i.e., task concreteness) than previous tasks such as Wright (2006a). We could not compare our 4 year-olds to Wright's study directly, because the groups in that study were 5, 6 and 7 years. However, we can compare our findings for 5 and 6 year-olds. Here we found that 6 year-olds performed very well at around 74%, which is far better than the 49% reported by Wright in his condition most similar to our own (labelled "Fully-Transitive").

The higher levels of performance observed here are in line with the view that the more relevant, interactive and perhaps even more fun context in which our child participants were tested, together with the warm-up regime we employed, permitted 4, 5 and 6 year-olds to give a better account of their true deductive abilities, as compared to previous developmental transitive studies (e.g., Markovits & Dumas, 1999; Wright & Dowker, 2002). However, it is possible that familiarisation and increased context relevance may not have had a direct effect on children's demonstrated transitive capacity, but instead simply promoted the generation of an heuristic that could result in above chance performance at least for the 5 year-olds (heuristic here refers to the more general use not to be confused with "heuristic" as used in Dual-Process theory below - see Smith & Collins, 2009 for discussion of both uses). But as long as task familiarity and/or context-relevance did not do this either via inducing the solution as a memory by-product (i.e., associatively - Russell et al., 1996) or by biasing towards correct answers (Wright & Dowker, 2002), such an heuristic might well qualify as a deductive process or deductive procedure (for related discussion see Markovits & Thompson, 2008).

A few investigators have noted different capacities in different transitive tasks. For example, Russell et al. (1996) found 6 year-olds exhibited higher performance in length versus location tasks. This is in line with the thesis that there are content effects inherent in transitive reasoning (note, they did not discuss this in their paper and instead alluded to "logical necessity"). By contrast, our findings for a height versus racing task are at least suggestive that, when tested in ways fully appropriate to children, deductive-transitive reasoning might well stem from a single capacity that can be deployed to a variety of contents with similar effectiveness (Berens & Hayes, 2007; Kallio, 1988; Perner, Steiner & Staehelin, 1981; Piaget & Inhelder, 1967; Siemann & Delius, 1996). Admittedly, this conclusion may pertain to children more than to adults. Also, even for children, our use of only two tasks cannot be absolutely definitive on this issue. Thus, our conclusion here will require future replications with other contents that can be presented in experimental contexts with which children are highly familiar or intrigued.

As well as content, there was another potentially-influential difference between our height task and our racing task - one was "simultaneous" and the other was "successive". Concerning simultaneous premise presentations, one of the advantages (and indeed a major rationale) of the

Pears and Bryant (1990) task was that it avoided the need to hold the premise information in memory while a further premise is considered or while the transitive solution is being deduced (Ameel et al., 2007; cf. Kallio, 1988). However, to our knowledge, no study has yet directly contrasted performance on such a task (i.e., transitivity of height) with a task which is logically the same but does require premises to be held in mind. The two tasks we used allowed direct comparison perhaps for the first time (although see Siemann & Delius, 1996 and Perner et al., 1981 for similar arguments). Here, we found no reliable difference between our two tasks across our three age groups, nor any reliable tendency towards any advantage at one particular age. Just to clarify, our height task did not call for our child reasoners to retain in mind one premise-pair whilst the second pair was added (Maybery, Bain & Halford, 1986) but our racing task did require such memory retention, and yet the findings at each of our three ages were indistinguishable from one task to the other. Thus, although concerns about memory lowering transitive reasoning performance was one of the main justifications for turning to the Extensive-Training-Paradigm and abandon three-term tasks (Bryant, 1998; Pears & Bryant, 1990), it would seem that this theoretical concern is not borne out by the empirical evidence (Brainerd & Reyna, 1992; Maybery et al., 1986; Markovits & Dumas, 1999; Wright, 2006b).

Before turning to how our findings are best conceptualised, a core issue is about whether the most accurate estimate of the age at which deductive-transitive-inference reaches competence should be taken as 4 years (Holcomb et al., 1997; Russell et al., 1996; Pears & Bryant, 1990) or some time beyond 7 years (Ameel et al., 2007; Kallio, 1988; Markovits & Dumas, 1999; Wright, 2006a). Relevant to this issue, Goodwin and Johnson-Laird (2006) found that even by the end of adolescence, deductive-transitive-inference has still not fully matured (see also Reyna & Farley, 2006; Wright & Howells, 2008). Goodwin and Johnson-Laird’s participants’ mean performance never exceeded around 89%, although admittedly their task was very demanding, using transitive problems having two premise-pairs but housing four integrated items (i.e., four-term series) and also using a double-relational comparative (e.g., one premise was of the form John is taller than Peter to a greater extent than Robert is taller than Mike). But with our tasks, our findings indicate that deductive-transitive-inference has approached competent performance by around 6.5 years, and is already reliably above chance by 5.7 years of age.

This said, our additional threshold analyses do qualify this conclusion somewhat. Although 5 year-olds may be above chance, they are also below a threshold reflecting a well developed transitive competence. For 6 year-olds, they do approach this competence level. But they are neither above it nor are they close enough to perfect performance to be regarded as having a fully developed transitive capacity. Thus, although our findings do demonstrate that tending to the procedural details of a transitive task, indeed suggests a more well developed deductive capacity than have other recent studies (e.g., Ameel et al., 2007; Markovits & Thompson, 2008; Wright, 2006a); development is clearly not complete until sometime after 6 years (e.g., 7 years, Wright & Dowker, 2002; or even 8 years Markovits et al., 1995; Markovits & Thompson, 2008).

It is possible that the slightly more favourable age estimate from the present investigation stems partly from a "competence v performance" distinction (see Muller et al., 1999 for an introduction to this distinction regarding the development of other reasoning processes). Specifically, it could be argued that our study only indexed children's transitive "competence", whereas Markovits and colleagues (e.g., Markovits & Dumas, 1999; Markovits et al., 1995; Markovits & Thompson, 2008) are more interested in a more explicit demonstration of transitive capacity via "performance". However, we make two points here. First, the main difference between our height task and Markovits et al.'s height task was that we made sure the children were familiar both with the entities being reasoned about (blocks) and the actual activity used in the experiment. Whilst it is easy to see how this might render the children more comfortable with their task and promote better performance for that reason, it is difficult to see how it would reduce "performance" into the more basic "competence". Second, it is important to realise that our findings do show a gradual improvement in 5 and 6 year-olds' transitive capacity to just about midway between a level indicating random guessing and a level indicating perfect performance by 6.5 years (Halford et al., 1998). However, we also found that 6 year-olds still have some way to go before their performance could be said to be highly developed (i.e., significantly over 75%). Thus, our findings should perhaps be seen as qualifying Markovits et al.'s conclusion about the improvements between 6 and 8 years, rather than challenging that conclusion as such.

Accepting that our findings on age of competence in transitivity do not stray too far from those of three-term task advocates, two questions then arise. First, does the finding of a fairly well developed deductive-transitive capacity at or after around 6.5 years (Van Elzakker et al., 2003; see also Ameel et al., 2007) refute findings from within the Extensive-Training-Paradigm which points to children being "logical" by or before age 4 years (Bryant, 1977, 1998; Russell et al., 1996 but contrast Riley & Trabasso, 1994; Wright, 2006b)? Second, regardless of our answer to the first question, is there a way of accommodating both sets of findings (Bouwmeester et al., 2007; Wright, 1998)? We suggest the answer to the first question is NO, and to the second question is YES. Regarding the first question, note that the Bryant and Trabasso (1971) favourable findings for 4 year-olds have never really been closely replicated (e.g., see Riley & Trabasso, 1974); whereas findings from Three-Term-Paradigms which deny a transitive reasoning capacity until near or after 7 years have been well replicated (Bara et al., 2000; Markovits & Dumas, 1999; Maybery et al., 1986; Oakhill, 1984; Wright, 2006a; Wright & Dowker, 2002). This said, the many successes of the Extensive-Training-Paradigm with human groups not thought to have a well developed deductive capacity (e.g., Stromer et al., 1993) or with non-human groups some of which may not be held to experience conscious thought (e.g., Daisley et al., 2009), leave us in little doubt that with age-appropriate training and extensive enough training, the Bryant and Trabasso findings can likely be replicated with perhaps even younger children. But we must also conclude such tasks index transitivity that is not underpinned by deduction (Holcomb et al., 1997; cf. Riley & Trabasso, 1974; Wittemyer & Getz, 2007).

Regarding the second question, we note a recently burgeoning group of theories of reasoning that can accommodate both sets of findings - which may be generally termed Dual-Process theory (Bouwmeester et al., 2007; Bryson & Leong, 2007; Elqayam, 2009; cf. Evans, 2003; Evans, Handley, Neilens & Over, 2010; Markovits & Thompson, 2008; Reyna & Farley, 2006). So far-reaching and persuasive is its scope, that Klaczynski (2009) argues that, 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). It is therefore unsurprising that recent researchers have sought to explain transitivity development from a Dual-Process perspective (Ameel et al., 2007; Bouwmeester et al., 2007; Wright & Howells, 2008).

Dual-Process theories may differ in terms of the exact nature of the two posited processing routes: Examples are System 1 v System 2, 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 Chen & Chaiken, 1999; Evans, 2009; Klaczynski, 2009; Smith & Collins, 2009; Wright, 2001). But in general terms, all theories agree that we have evolved two main systems in order to function, the first is an inherently fast and parallel process that is largely instinctive, automatic and unconscious (Siemann & Delius, 1996), although the reasoner may be conscious of the outcome of its application. This system is thought to be based on the associative architecture of Long-Term Memory (Markovits & Thompson, 2008). The second system is a much slower and partly serial process that is typically voluntary, wilful and conscious in many aspects of its process as well as the outcome of that process (Smith & Collins, 2009). This second process is heavily reliant on Working Memory, and it dominates over the Associative System during abstract or hypothetical reasoning (Evans, 2003). Dual-Process theories can be divided in terms of whether they take the view that these two processes are completely independent of each other, take the view that one of them can influence the processing done by the other, or take the view that a further process or dimension functionally distinct from the two reasoning streams may need to somehow coordinate their contributions (see Elqayam, 2009; Evans, 2009; Wright & Howells, 2008 for discussions of these issues).

Regarding transitive reasoning, Kallio (1982) found that in order to succeed on his task within the Extensive-Training-Paradigm, children first demonstrated internalisation of the entire transitive series given by the pair-wise information, constructing an internal mental representation of the linearly graded scale (Breslow, 1981; Halford & Andrews, 2004; Wright, 2001). Note, children generally tended to be able to generalise the series, only if the premises had been given in both directions and were fully-ordered (see also Riley & Trabasso, 1974; Stromer et al., 1993; Siegal, 2003; Wright, 2006b). In Russell et al.'s (1996) study, they deliberately set out to ensure that the premise information was integrated into an overall representation of the entire five-term transitive series by using extensive-training. They then ensured that the series was in LTM and was not in WM, by introducing a delay of at least 7 minutes and filling this time with an unrelated intervening task designed to occupy WM. In this study and in the Extensive-Training-Paradigm more generally, there is a deliberate attempt to appeal to LTM-based associative representations (contrast Russell et al., 1996), which within a recent Hybrid Dual-Process conception we see as akin to Evans' (2009) Type 1 (non-deductive) processes.

Type 1 processes are to be distinguished from Type 2 processes. For example, in a variant on the Extensive-Training-Paradigm, Wright (2006b) found that adult participants gave responses via an associative mode in under 1 second. This contrasts with Wright and Howells (2008) who used a five-term task that was not reliant on extensive-training, and found responses now took some three-times as long. Also, in Wright (2006b) a deductive "response accuracy" profile was not evident at the start or end of extensive-training (e.g., the B:D inference was more accurate than the premises on which it was based - see also Russell et al., 1996; Stromer et al., 1993), which is indicative of a Type 1 process, whereas for Wright and Howells (2008) a deductive profile did emerge (e.g., now B:D was never better than B:C or C:D - see Clark, 1969), indicative of a Type 2 process. Hence, this contrast supports our view that two rather different types of process can just as well lead to transitive responses (e.g., associative v deductive - see also Russell et al., 1996).

When we need to act more quickly than our deductive skills permit, we may use the output of the associative process only, or even just until we arrive at an alternative course of action via the deductive process. But equally, when an inference reached deductively will be called on repeatedly, it is more economical to deduce it only the first few occasions and from then on simply store the solution for repeated retrieval later. Thus, having both processes will actually confer advantages. We reach this conclusion specifically about human transitive performance, but comparative theorists have recently begun to come to the same conclusion regarding transitive reasoning in non-humans (Premack, 2007). On this view, we should not at all be surprised to find that “transitive responding” via Type 1 or associative processes (Bryant & Trabasso, 1971; Siemann & Delius, 1998) would appear some years ahead of “transitive reasoning” by use of Type 2 or deductive processes applied to verbatim premises (Koutstaal & Cavendish, 2006; Markovits & Dumas, 1999; see also Riley & Trabasso, 1974 and Markovits & Thompson, 2008 for similar conception of transitivity and conditional reasoning development).

The one issue yet to be addressed in our account of transitive reasoning would seem to be the Pears and Bryant (1990) findings. Clearly their task did not require extensive-training or memorisation of anything (Kallio, 1988). Hence we cannot put their finding of transitive reasoning in 4 year-olds down entirely to Type 1 processes of Hybrid Dual-Process theory. Relevant to this issue, Wright (2001) maintained that additional to the two processing routes, there is a Cue-Based Process which can be called on in any memory-saving or process-saving strategy that the reasoner chooses (or is forced) to rely on during reasoning. Essentially, the cue-based system is "preconscious" and facilitates faster or less effortful conclusions from either Type 1 or Type 2 processes (Evans, 2009). The two processing routes plus the cue-based system are then brought together in a Type 3 process, which coordinates them in order to arrive at the decision/action (see Evans, 2009).

Adopting Trabasso's own reflection about his earlier 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). We suspect that in their subsequent experiment, Pears and Bryant (1990) again inadvertently provided precisely the type of non-logical cue (assistive height cue) that Bryant and Trabasso had failed to eliminate in their devising of the Extensive-Training-Paradigm. Recent improvements to height tasks (e.g., Ameel et al., 2007; cf. Markovits et al., 1995) have simply assessed transitivity whilst successfully eliminating the non-logical height cue. So it seems that three apparently separate streams of research (Extensive-Training-Paradigm, Three-Term-Paradigm, and height tasks) can actually be accommodated within a single conception of transitive reasoning in children.

Conclusions

There have been decades of debate over whether or not 4 year-olds possess the ability to routinely reason logically in relation to transitive reasoning. Our findings here relied on a task that is theoretically the easiest to solve and the least demanding on memory. We found that 4 year-olds do not have a well developed deductive transitive competence, 5 year-olds are well on the way to developing one, and 6 year-olds do have one. Estimates from the present study indicate that the actual age by which children first possess a deductive-transitive-inference that is above chance is around 5.7 years. However, even the well developed capacity found in 6 year-olds falls some way short of children being "logical" at this age.

At this stage our findings are somewhat suggestive about exactly why our results may give a slightly more favourable estimate of 6 year-olds' deductive-transitive-inference competence, compared to other tasks which used the Three-Term-Paradigm (e.g., in terms of practice, familiarisation, increased relevance to the children). However, we are more confident that the difference between our findings and those from the Extensive-Training-Paradigm reflect a deductive mode for transitive reasoning versus an associative mode for transitive reasoning, respectively (or perhaps more accurately - a Type 2 or deduction-capable set of processes v a Type 1 or associative-led set of processes). Thus, explaining transitive reasoning findings from different research streams would seem to require a Dual-Process conception of children's reasoning development. We also conclude that, previous transitive tasks based on the comparative of "height" have indeed been assessing the deductive mode. However, some may have inadvertently allowed in "cues", with these cues leading to overly favourable estimates particularly for 4 year-olds. It is hoped that research will now turn to identifying further details regarding how Dual-Process reasoning operates during children's socio-cognitive development.

References

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 & Reasoning, 13 (2), 164-187.

Artman, L., & Cahan, S. (1993). Schooling and the development of transitive inference. Developmental Psychology, 29 (4), 753-759.

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

Berens, N. M., & Hayes, S. C. (2007). Arbitrarily applicable comparative relations: Experimental evidence for a relational operant. Journal of Applied Behavior, 40, 45-71.

Birenbaum, M. H., & Gutierrez, R. J. (2007). Testing for transitivity of preferences predicted by a lexicographic semi-order. Organisational Behaviour and Human Decision Processes, 104, 96-112.

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., Vermunt, J. K., & Sijtsma, K. (2007). Development and individual differences in transitive reasoning: A fuzzy trace theory approach. Developmental Review, 27, 41-74.

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

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

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., & Trabasso, T. (1971). Transitive inferences and memory in young children. Nature, 232, 456-458.

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken and Y. Trope (Eds.), Dual-process theories in social psychology (pp. 73-96). New York: Guilford Press

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

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.

Deneault, J., & Ricard, M. (2006). The assessment of children's understanding of inclusion relations: Transitivity, asymmetry, and quantification. Journal of Cognition and Development, 7 (4), 551-570.

Elqayam, S. (2009). Models of dependence and independence: A two-dimensional architecture of dual processing. Thinking & Reasoning, 15 (4), 377-387.

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. (2009). How many dual process theories do we need? One, two, or many? In J. St. B. T. Evans & K. Frankish (Eds.), In two minds: Dual processes and beyond. (pp. 33-54). Oxford, UK: Oxford University Press.

Evans, St. B. T., Handley, S. J., Neilens, H., & Over, D. (2010). The influence of cognitive ability and instrumental set on causal conditional inference. The Quarterly Journal of Experimental Psychology, 63 (5), 892-909.

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

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

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 & Reasoning, 10 (2), 123-145.

Halford, G. S., & Galloway, W. (1977). Children who fail to make transitive inference can remember comparisons. Australian Journal of Psychology, 29, 1-5.

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.

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.

Hong, L., & Chond, L. (2001). The mental models of solving of three-term series problems by individuals. Acta Psychologica Sinica, 33 (6), 518-525.

Hooper, F. H., Toniolo, T. A., & Sipple, T. S. (1978). A longitudinal analysis of logical reasoning relationships: Conservation and Transitive inference. Developmental Psychology, 14 (6), 674-682.

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

Kallio, K. D. (1988). Developmental differences in the comprehension of simple and compound comparative relations. Child Development, 59, 397-410.

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.

Koutstaal, W., & Cavendish, M. (2006). Using what we know: Consequences of intentionally retrieving gist versus item-specific information. Journal of Experimental Psychology: Learning Memory and Cognition, 32 (4), 778-791.

Krackhardt, D., & Kilduff, M. (1999). Whether close or far: Social distance effects on perceived balance in friendship networks. Journal of Personality and Social Psychology, 76, 770-782.

Lazareva, O. F., & Wasserman, E. A. (2010). Nonverbal transitive inference: Effects of task and awareness on human performance. Behavioural Processes, 83, 99-112.

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

Markovits, H., & Dumas, C. (1992). Can pigeons really make transitive inferences? Journal of Experimental Psychology: Animal Behavior Processes, 18 (3), 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.

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

Maybery, M. T., Bain, J. D., & Halford, G. S. (1986). Information-processing demands of transitive inference. Journal of Experimental Psychology: Learning, Memory and Cognition, 12 (4), 600-613.

Muller, U., Sokol, B., & Overton, W. F. (1999). Developmental sequences in class reasoning and propositional reasoning. Journal of Experimental Child Psychology, 74, 69-106.

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

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

Perner, J., Steiner, G., & Staehelin, C. (1981). Mental representation of length and weight series and transitive inferences in young children. Journal of Experimental Child Psychology, 31, 177-192.

Piaget, J., & Inhelder, B. (1967). The child's conception of space: Translated from French by F. J. Langdon & J. L. Lunzer (originally published 1956). London: Routledge & Kegan Paul.

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. Proceedings of the National Academy of Sciences of the United States of America, 104, (35), 13861-13867.

Rabinowitz, S. M., Grant, M. J., Howe, M. L., & Walsh, C. (1994). Reasoning in middle childhood: A dynamic model of performance on transitivity tasks. Journal of Experimental Child Psychology, 58, 252-288.

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), 1-44.

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.

Shafir, S., Waite, T. A., & Smith, B. H. (2002). Context-dependent violations of rational choice in honeybees (Apis mellifera) and grey rays (Perisoreus Canadensis). Behavorial Ecology Sociobiology, 51, 180-187.

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.

Siemann, M., & Delius, J. D. (1998). Algebraic learning and neural network models for transitive and non-transitive responding. European Journal of Cognitive Psychology, 10 (3), 307-334.

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.

Sternberg, R. J. (1980). Representation and process in linear syllogistic reasoning. Journal of Experimental Psychology: General 109 (2), 119-159.

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.

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.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005). Every-day conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. Memory & Cognition, 33, 107-119.

Whelan, R., Barnes-Holmes, M., & Dymond, S. (2006). The transformation of consequential functions in accordance with the relational frames of more-than and less-than. Journal of the Experimental Analysis of Behavior, 86, 317-335.

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

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

Wright, B. C. (2001). Reconceptualizing the transitive inference ability: A framework for existing and future 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 & Reasoning, 12 (1), 91-123.

Wright, B. C., & Dowker, A. D. (2002). The role of cues to differential absolute size in children's transitive inference. 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 & Reasoning, 14 (3), 244-280.

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

Table 1: Task Performance According to Age Group

	4 Years	5 Years	6 Years	All Groups
Height Task	2.097 (0.226)	2.583 (0.257)	2.950 (0.199)	2.543 (0.132)
	52%	65%	74%	64%
Race Task	2.071 (0.194)	2.524 (0.194)	3.000 (0.192)	2.532 (0.112)
	52%	63%	75%	63%
Overall	2.084 (0.149)	2.554 (0.161)	2.975 (0.138)	
	52%	64%	74%	

Note: Figures in parenthesis are standard errors and the scores are given as percentages in bold.

Figure Captions

Figure 1: Depiction of the Pears and Bryant task using a three-term example. The reasoner could be asked to work out which of blocks A and C would be higher if there was a single tower that had included both these blocks. The experimenters constructed the task such that the reasoner could consider block-pile A:B and block-pile B:C, and then deduce the relationship between block-A and block-C. But the child did not need to understand deductive transitivity. S/he could simply observe that block-A is high in its pile but block-C is not; which gives the right answer to the experimenters' question without ever needing to deduce anything.

Figure 2: Depiction of the 4 test trials used in the height task. Trial-types are – top left = 11, top right = 01, bottom left = 10, bottom right = 02. For each trial-type, the child was provided with blocks A and C only side by side as depicted in the centre of the figure (i.e., was not offered a filler block). The child answered the transitive question by placing either block-A on top of B or placing block-B on top of A.

Figure 1:

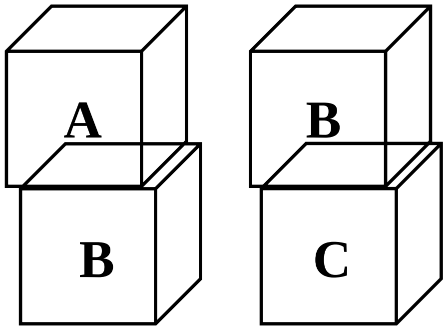

Figure 2: