Factors and processes in children's transitive deductions

Barlow C. Wright & Jennifer Smailes

Published online: 17 Aug 2015.

Submit your article to this journal ⤤

Article views: 712

View Crossmark data ⤤

Routledge
Taylor & Francis Group

Factors and processes in children's transitive deductions

Barlow C. Wright and Jennifer Smailes

Division of Psychology, Brunel University London, Uxbridge UB8 3PH, UK

Transitive tasks are important for understanding how children develop socio-cognitively. However, developmental research has been restricted largely to questions surrounding maturation. We asked 6-, 7- and 8-year-olds ($N = 117$) to solve a composite of five different transitive tasks. Tasks included conditions asking about item-C (associated with the marked relation) in addition to the usual case of asking only about item-A (associated with the unmarked relation). Here, children found resolving item-C much easier than resolving item-A, a finding running counter to long-standing assumptions about transitive reasoning. Considering gender perhaps for the first time, boys exhibited higher transitive scores than girls overall. Finally, analysing in the context of one recent and well-specified theory of spatial transitive reasoning, we generated the prediction that reporting the full series should be easier than deducing any one item from that series. This prediction was not upheld. We discuss amendments necessary to accommodate all our earlier findings.

Keywords: Children's reasoning; Gender; Markedness; Mental Seriation; Transitive deductions.

If we know John is taller than David, and David is taller than Eric, then we can deduce Eric must be shortest of all three, that John is tallest and also that David is intermediate in height. This thought process is often termed relational reasoning, linear syllogistic reasoning or transitive reasoning (Bonnefond, Castelain, Cheylus, & Van der Henst, 2014; Clark, 1969; Guez & Audley, 2014; Piaget, 1965; Sternberg, 1980; Wright, 2001). In more abstract terms, we can say "if $A > B$, and $B > C$, it then follows that $A > C$". This kind of deductive reasoning is basic to the development and normal functioning of many socio-cognitive processes—from mathematical and text-processing skills, through friendships and the trusting of other people, to generalisations of racial prejudice (Coleman et al., 2010; Favrel & Barrouillet, 2000; Kim & Song, 2011; Markovits, Dumas, & Malfait, 1995; Ragni & Knauff, 2013; Sedek, Piber-Dabrowska, Maio, & Von Hecker, 2011). For example, one might generally trust Mary more than Sally, but might trust Sally more than Trudy. Applying what Halford and Andrews (2004) have called "the transitivity principle", it follows that in a choice between Mary and Trudy, one should generally trust Mary.

Three-term problems such as our earlier examples became significant to psychologists in around 1921, when Piaget investigated the age at which they tended first to be solved by children (see Guez & Audley, 2014; Lazareva & Wasserman, 2010; Piaget, 1965; Piaget & Garcia, 1991; Piaget, Grize,

Szeminska, & Bang, 1977). However, generally, these tasks meet with a different fate in cognitive research versus in developmental research. In cognitive research, three-term tasks are embraced alongside other tasks, as useful for assessing the simplest case of relational reasoning (Clark, 1969; Evans, Newstead, & Byrne, 1993; Knauff, 2009; Wright & Dowker, 2002; Wright, Robertson, & Hadfield, 2011).

In that tradition, Knauff and May (2006) introduce a theory of reasoning that utilises a spatial array framework (cf. De Soto, London, & Handel, 1965) with mental models theory (cf. Johnson-Laird, 1983), and espouses that there are three basic stages of reasoning with such problems: (1) Visualising the items in each of the two premises and also the relational-comparison used (e.g. "taller than)", achieved via processing in occipital lobe. (2) Spatially representing the premises within a single array regarding the relation in question, largely due to processing in parietal lobe (e.g. John > David > Eric). (3) Describing the array symbolically, evaluating the inferential question asked (e.g. "Who is tallest of all three males?)" and then inspecting the symbolic series in memory in order to reach the required answer, achieved by processing in prefrontal cortex within the frontal lobe. Neuroscience evidence shows that, intriguingly, the final stage does not call on visuo-spatial information involving occipital and parietal lobes (Knauff, 2009; Krawczyk, 2012; but see also Wright, 2012). Additionally, this stage may sometimes lead to constructing a mental model that is indeterminate or invalid (e.g. A blood relative of B, B blood relative of C, might lead to A blood relative of C, when in fact the inference is indeterminate—Lazareva & Wasserman, 2010; Wright, 2001).

In line with this theory, Reyna and Brainerd (1990) had previously found that children's transitive inference operates by reasoners using the premises to generate gist and flow information for the whole series (e.g. "things get larger to my right)", but upon doing so, the verbatim premises are no longer retained. Memory for antecedents is less durable than gist representations; and so gist rather than deductive coordination of premises as such is utilised to solve for inferential comparisons. Knauff and May (2006) could be seen as adding to this distinction, insofar as they found that if the verbatim premises are visualised, this can actually interfere with the reasoning process; the implication being that visualisation can make the premises more durable and hence can cause interference with gist.

However, despite initial developments such as Reyna and Brainerd's theory, some developmentalists argued that three-term tasks are invalid as a test of children's deductive inference-making (Bryant, 1998; Bryant & Trabasso, 1971). To again use our initial relation, when premise $A > B$ is presented, the child labels item-A as tall and item-B as short. When $B > C$ is presented, item-B is labelled tall and item-C labelled short. Regarding inferring the tallest, item-A was never labelled short in its premise, and so the child always realises it is the tallest, but the child did not call on transitivity or any kind of reasoning to reach this conclusion (cf., Bryant & Trabasso, 1971). Aside from the curious implication that three-term tasks can be valid in research with adults, but not research with children, we list just three of the many further problems with the labelling conjecture.

First, advocates of labelling never presented empirical evidence for their view. The only developmental study directly testing it found it not to feature unless magnitude differences between items are highly pronounced (Wright & Dowker, 2002; see also Guez & Audley, 2014 regarding non-humans).

Second, the labelling view predicts that three-term tasks are easier to solve than other competing tasks said to index transitivity (e.g. Bryant & Trabasso's own five-term task using extensive training). However, Bryant's own lab intimates that this prediction is not upheld (compare 3-term paper of Bryant & Kopytynska, 1976 versus paper on 5-term extensive-training task of Bryant & Trabasso, 1971; see also Ameel, Verschueren, & Schaeken, 2007; Markovits & Dumas, 1999; Wright, 2012). Note, tasks avoiding training just like Piaget's three-term task, but which utilise five-terms or more, do appear more cognitively demanding than Bryant and Trabasso's task with 4–6-year-olds (e.g. Andrews, 2010; Markovits et al., 1995; Wright & Howells, 2008). That said, the Bryant and Trabasso (1971) findings themselves have proven highly difficult to replicate with children at or near the age of 4 years (Holcomb, Stromer, & Mackay, 1997; Riley & Trabasso, 1974; Wright, 2012).

Third, in well-controlled transitive studies, each premise is typically presented first 50% of the time. Thus, 50% of the time item-B of premise $B > C$ is first to be labelled tall, and so item-A of the next ($A > B$) premise to be presented would simply cause confusion, or lead to deletion of item-B (Bonnefond et al., 2014), unless of course the child already understands transitivity and can re-order the premises and items deductively (Ameel et al., 2007; Piaget & Garcia, 1991; Piaget et al., 1977; Riley & Trabasso, 1974; Wright & Dowker, 2002).

Labelling disputes aside for now, there are three important issues in transitive research that have received little attention to date: These are ecological validity, gender and markedness. Regarding ecological validity, transitive studies have tended to rely on tasks with a single dimension such as height, speed or weight (Markovits et al., 1995; Wright, Robertson, & Hadfield, 2011). An over-reliance on only one transitive relational-comparison, (e.g. "is better than)" means that study findings could be tied to one specific content (e.g. height but not weight), rather than applying across transitive deductions more generally (a distinction perhaps first captured by Piaget's contrast between concrete versus formal operational thought—e.g. Piaget & Garcia, 1991). One solution is to rely on a large number of different transitive relations and contexts across different three-term tasks, to make possible a composite measure of overall transitive capacity. However, this precaution has rarely been taken (for useful examples see Knauff & May, 2006; Markovits & Dumas, 1999).

Gender may also impact on transitive performance, and yet it seems not to have been a focus of any published study of transitive reasoning. However, gender has been investigated in other areas of cognitive development. One consistent finding is that girls tend to outperform boys on tasks that tap into verbal abilities (Mills, Ablard, & Stumpf, 1993; Strand, Deary, & Smith, 2006). Sternberg (1980) found that transitive reasoning has a strong verbal component. Studies have also investigated spatial reasoning, which was Sternberg's second component to transitive reasoning. Here, a slight advantage to boys increases towards adolescence (Hegarty, Keehner, Kooshabeh, & Montello, 2009; Strand et al., 2006).

Given that transitive reasoning is thought to be partly reliant on children's verbal processing (e.g. solving out aloud—Sternberg, 1980; Trabasso, Riley, & Wilson, 1975; Wright & Dowker, 2002), we wondered whether, during development of deductive transitivity, girls might enjoy an advantage in transitive reasoning. However, given that boys gradually begin to enjoy an advantage in spatial reasoning, and some theorists maintain that spatial competencies are important in transitive reasoning (Brunamonti, Genovesio, Carbe, & Ferraina, 2011; Knauff & May, 2006; Trabasso, 1977; Wright, 2012), it may well be that it is boys who show a transitive reasoning advantage as transitivity increasingly matures.

Kallio (1988) proposed that there are four reference points when making a transitive inference. To use again our initial example of people's heights, "tall" is the main adjective and is the primary reference point from which children make deductions. "Short" is the secondary reference point, as it is the converse of the primary reference point and is defined in relation to that point (Knauff & May, 2006). Short is a "marked" relational adjective compared to tall, which is described as unmarked (Andrews, 2010; Clark, 1969; Maybery, Bain, & Halford, 1986; cf., Piaget, 1965; Sternberg, 1980; Wright, 2001).

Generally speaking, markedness is a term psycholinguists and logicians use to refer to the relationship between two adjectives which can be taken to be the polar opposite of one another (e.g. happy versus sad, big versus small, fast versus slow, heavy versus light, etc.—Chen, Lu, & Holyoak, 2014; cf., Clark, 1969). In the context of transitivity, a marked adjective may form part of the relational-comparative (Andrews, 2010; Maybery et al., 1986). However, usually, its purpose will be to attach to the noun given in the pairwise comparison (e.g. "David is short" relative to John; or "the little mouse versus the big elephant"—Chen et al., 2014; Wright, 2001). Following the latter, as here the relational adjective is specifically used to describe (i.e. is attached to) a concrete entity (i.e. an actual item which in linguistic terms constitutes a noun), we here refer to the associated items as most marked (item-C) versus most unmarked (item-A) in terms of that relational adjective, in the interests of both space and convenience.

In agreeing that markedness is a factor that affects the difficulty of reasoning with transitive relations, Andrews argues that this is "presumably because marked forms are first converted to the unmarked form" (Andrews, 2010, p. 935). In line with this view, when studies pose the transitive question in terms of the unmarked item, this leads to higher accuracy during premise acquisition as well as transitive responding (Acuna, Eliassen, Donoghue, & Sanes, 2002; Andrews, 2010; Carmesin & Schwegler, 1994; Holcomb et al., 1997; Titone, Ditman, Holzman, Eichenbaum, & Levy, 2004). Statistical models, mathematical models and computational simulations encapsulate this unmarked advantage (Breslow, 1981; De Lillo, Floreano, & Antinucci, 2001; Wu & Levy, 2001). That said, when half of adults are taught the series in terms of the unmarked relation, with the other half taught in terms of the marked relation, there is no difference in either learning or transitive responding between these two groups (Lazareva & Wasserman, 2010). Clearly, these two sets of findings are

discrepant with one another. Markedness, then, is in need of further investigation.

Two further elements of Kallio's theory are tertiary versus quaternary reference points. The tertiary reference point is the understanding that, for example *A* is taller than *B* or *B* is taller than *C*. Last and most demandingly, a quaternary reference point is the coordination of three items $A > B > C$. For example, coordinating John is taller than David, but David is taller than Eric (see Halford, Wilson, & Phillips, 1998 for similar conception). Kallio's theory intimates that reporting of all three items should be no better than reporting the marked or unmarked item, because both involve integration and the quaternary reference point. This prediction runs contrary to what we may derive from Knauff's (2009) theory. Knauff (2009) asserts that premise reordering occurs in stage 2 of 3, "before" the premises are formed into a symbolic mental model for scanning, but it does not itself involve deduction (see also Bonnefond et al., 2014). Thus, all three items are already present in the correct order at stage 2, but the answering of a question about, say, which item is shortest must await stage 3, where the mental model is constructed and consulted.

Summary of aims

For a more ecologically valid index of children's transitive reasoning, we used a composite measure comprising five three-term tasks which referenced different contents. We considered gender in our analyses, for the first time in published transitive research. Additionally, we included a condition asking children about item-*C*, additional to a condition asking about item-*A*, to test the commonly held belief that performance on the unmarked item-*A* is superior, again for the first time. Finally, we included a further condition asking children to report the entire transitive series, to determine if reporting all three items in order (mental seriation) is easier or harder than reporting the end items (Kallio, 1988; Knauff, 2009).

METHOD

Participants

Participants were 117 children of 5–8 years, from schools local to the research institution. The children were predominantly Caucasian and from working and middle-class backgrounds. Children from Year 1, Year 2 and Year 3 took part, with 39 children in each of these groups. The associated ages were 6 years ($M = 6.32$, $SD = 0.39$), 7 years ($M = 7.16$, $SD = 0.33$) and 8 years ($M = 8.25$, $SD = 0.26$). The 6-year-old group contained 19 girls and 20 boys, for the 7-year-olds these were 20 and 19, respectively, and for the 8-year-olds, these were 14 and 25. The mean ages for the two genders were the same (+ or − 0.10 years).

Materials

For each of our five tasks, there were two photo picture cards, each one showing the relationships between two objects. One picture always showed an Object *A* in relation to Object *B* ($A > B$) and the other picture always showed Object *B* in comparison to Object *C* ($B > C$). The five tasks were about animals, household items, cars, balls and Finding Nemo; and each one is briefly described later. The photo-picture cards were made by photographing a background and then photographing the objects. For the animal, household items and car tasks, the pictures were cropped digitally using the computer programme PhotoShop. Pictures for the balls task were made by photographing the objects on to white paper and the photos were simply digitised and printed. The pictures for the Finding Nemo task were made by photographing the objects against an Under-The-Sea cartoon background, which had come with the toys when bought.

Additional to the two pictures, we presented one instance of each of the actual toy objects themselves. The physical objects were used by the children to assist their responses. Within each of the five tasks, the actual physical objects (items *A*, *B* and *C*) as presented to the child were always the same ones depicted in the pictures, but were all the same in actual size. This was intended to help ensure that the child relied on what the pictures told them and not on the absolute visual information of the actual objects (see Wright & Dowker, 2002).

Task 1 was adapted from Kallio (1988), and was about animals (the items) and races (speed being the transitive comparison). One picture showed a sheep and a horse in a race, and the sheep won ($A > B$). In another picture, the horse had a race with the pig, and the horse won ($B > C$).

Task 2 was adapted from Markovits and Dumas (1999) and was about household items and tallness. Here, the three objects were a ruler, a comb and a toothbrush, respectively. In the pictures, the objects were seen in pairs in a stationery holder, with a

different portion of each protruding outside the stationery holder.

Task 3 was also adapted from Markovits and Dumas (1999) and used cars and length. In our task, there were three cars: A red car, a black car and a white car. Each picture showed one car partly driven into a garage (partial occlusion) whilst the other car was outside the garage (no occlusion). The children were verbally told that the car in the garage is longer than the car outside the garage.

Task 4 was adapted from Kallio's (1988) task about relative heights. It used balls which were depicted as bouncing. Here, the experimenter wanted to see which of the balls bounced the highest. The balls were blue, pink and silver, respectively.

Lastly, task 5 was a task about cartoon characters in a race. Here, the cartoon characters were from the film Finding Nemo. The characters used were Bubbles, Nemo and Dory.

Design

The study used a multi-factorial design. Independent measures were the age group of the child and the child's gender. The level of markedness (unmarked versus marked) were two levels of a single repeated-measures factor. Separate analyses considered full-series reporting (i.e. asking for the whole series $A > B > C$) both in isolation and in comparison to the average of the marked and unmarked items. In each case, the Dependent Variable was the participant's score.

Procedure

Children were tested in a quiet space reserved just outside the classroom. The five tasks were given in a pre-randomised order. For each task, the left–right spatial location of the two premises and also the order of giving the three questions on the markedness factor were also randomised, and the earlier randomisations were achieved by following instructions on a pre-randomised ordering sheet. About 200 sheets were created, with a sheet selected at random for each child at the start of that child's participation.

For each task, the two photo-picture cards were placed on the table in front of the child simultaneously. The child was asked to describe each picture, and stated what the objects were and what the relation between them was. Thus, the experimenter did not have to present any premises verbally, and could reserve verbalisations for conversational reasons only, whilst encouraging the child's own verbalisations (a technique that teachers will be familiar with). The child was then asked for the most unmarked item (e.g. which ball bounces highest of all three balls). Here, the child responded via a combination of voice, gesturing and touching of the concrete objects or the items in the photographs. Children were also asked for the most marked item (e.g. which ball bounces the lowest) and what the whole series was (e.g. from highest to lowest).

Questions were asked in random order and each answer was recorded on a response sheet. The test session lasted between 10 and 15 minutes. Throughout their participation, all objects and photographs were continually available for inspection, thus any memory loads were minimal. After completing all five tasks, the child was thanked for his/her participation, and given a treat as previously agreed with the class teacher.

RESULTS

For the unmarked condition, a score of 1 was given if that condition had been responded to correctly, or a 0 if not. These scores were then summed over all five tasks, to give a maximum cumulative score of 5. This process was repeated for the marked condition and the full-series condition. Unmarked versus marked performance is summarised in Table 1, according to gender and age group. A three-way mixed-model analysis of variance (ANOVA) used factors of markedness (2 levels), *gender* (2 levels) and *age group* (3 levels). This and our other analyses used a 2-tailed hypothesis with an Alpha level of 0.05 unless otherwise specified.

Table 1 shows that performance in the unmarked condition (identifying item-*A*) was around 5% lower than for the marked condition (identifying item-*C*). The difference was statistically significant ($F(1, 111)$ = 5.14, $p = 0.03$, *Partial Eta*$^2 = 0.04$). Regarding gender, boys tended to perform around 7% higher than did girls (Table 1), with this difference also statistically significant ($F(1, 111) = 4.22$, $p = 0.04$, *Partial Eta*$^2 = 0.04$). Turning to age groups, there was around a 10% improvement in performance between the 6- and 7-year-olds, and a further 6% improvement between ages 7 and 8 years, leading to a significant main effect of age ($F(2, 111) = 7.51$, $p < .01$, *Partial Eta*$^2 = 0.12$).

TABLE 1
Mean transitive performance by age-group gender and markedness

	6-Year-olds (%)	7-Year-olds (%)	8-Year-olds (%)	All years (%)
Gender female				
Unmarked (A)	2.79 (0.26) 56	3.70 (0.25) 74	3.71 (0.30) 74	3.40 (0.16) 68
Marked (C)	3.16 (0.21) 63	3.85 (0.20) 77	3.93 (0.24) 79	3.65 (0.13) 73
Overall	2.97 (0.21) 59	3.78 (0.20) 76	3.82 (0.24) 76	3.52 (0.13) 70
Gender male				
Unmarked (A)	3.30 (0.25) 66	3.74 (0.26) 75	4.32 (0.22) 86	3.79 (0.14) 76
Marked (C)	3.85 (0.20) 77	3.68 (0.21) 74	4.32 (0.18) 86	3.95 (0.12) 79
Overall	3.58 (0.20) 72	3.71 (0.21) 74	4.32 (0.18) 86	3.87 (0.11) 77
Gender both				
Unmarked (A)	3.05 (0.18) 61	3.72 (0.18) 74	4.02 (0.19) 80	3.59 (0.10) 72
Marked (C)	3.50 (0.15) 70	3.77 (0.15) 75	4.12 (0.15) 82	3.80 (0.09) 77
Overall	3.27 (0.14) 65	3.74 (0.14) 75	4.07 (0.15) 81	3.70 (0.09) 74

N for each age group = 39. Maximum average score for each cell = 5. Numbers in parentheses are standard errors.

Table 1 shows that a difference of 9% between unmarked and marked conditions at age 6 years was reduced to less than 2% at age 7 years. However, it also shows a tendency for the difference between unmarked and marked conditions to remain more or less stable between 7 and 8 years. The result of these two differing profiles was an overall two-way interaction effect that did not reach significance ($F(2, 111) = 2.05$, $p = 0.13$, Partial $Eta^2 = 0.03$). None of the remaining interactions were statistically significant (Markedness × Gender − $F(1, 111) = 0.19$, $p = 0.67$, Partial $Eta^2 < 0.01$; Gender × Age − $F(2, 111) = 1.55$, $p = 0.22$, Partial $Eta^2 = 0.03$; Markedness × Gender × Age − $F(2, 111) = 0.52$, $p = 0.59$, Partial $Eta^2 < 0.01$).

Turning now to the full series, performance is summarised in Table 2. We conducted a two-way ANOVA with Age and Gender as factors, and the full-series performance as the dependent variable. The main effect of Gender was marginally significant ($F(1, 111) = 3.56$, $p = 0.06$, Partial $Eta^2 = 0.03$). Age was statistically significant ($F(2, 111) = 8.13$, $p < .01$, Partial $Eta^2 = 0.13$). However, there was no statistically significant interaction between age and gender for full-series performance ($F(2, 111) = 0.65$, $p = 0.52$, Partial $Eta^2 = 0.01$).

In order to address the question of whether generating the full series was more demanding or less demanding than answering for unmarked/marked items, a further ANOVA compared performance on the full-series condition versus mean performance of the unmarked and marked items. Table 3 shows that overall performance for mean markedness was around 13% higher than for full-series reporting. Gender was not included in this analysis because our earlier analyses already confirmed that gender differences exist for the unmarked item, marked item and the full series, but gender does not interact with age or markedness. The present analysis showed a statistically significant main effect of condition (Markedness versus Full-Series − $F(1, 114) = 97.83$, $p < .01$, Partial $Eta^2 = 0.46$). There was also a statistically significant main effect of Age ($F(2, 114) = 9.07$, $p < .01$, Partial $Eta^2 = 0.14$). Furthermore, there was a significant two-way interaction between condition and age, whereby the disadvantage of the full-series condition compared to mean markedness was around 18% at age 6 years and decreased to 10% by age 8 years ($F(2, 114) = 4.35$, $p = 0.02$, Partial $Eta^2 = 0.07$).

Our final set of analyses assessed children's performance against two criteria—random guessing

TABLE 2
Full-series performance by age and gender

	6-Year-olds (%)	7-Year-olds (%)	8-Year-olds (%)	All years (%)
Female	1.95 (0.31) 39	3.20 (0.30) 64	3.36 (0.36) 67	2.84 (0.19) 57
Male	2.80 (0.30) 56	3.37 (0.31) 67	3.76 (0.27) 75	3.31 (0.17) 66
Genders combined	2.37 (0.22) 47	3.28 (0.22) 66	3.56 (0.22) 71	3.07 (0.13) 61

N for each age group = 39. Maximum average score for each cell = 5. Numbers in parentheses are standard errors.

TABLE 3
Summary of mean items versus full-series performance

	6-Year-olds (%)	7-Year-olds (%)	8-Year-olds (%)	All years (%)
Item A,C	3.28 (0.15) 66	3.76 (0.15) 75	4.14 (0.15) 83	3.73 (0.09) 75
Full Series	2.39 (0.22) 48	3.28 (0.22) 66	3.62 (0.22) 72	3.10 (0.13) 62
Overall	2.83 (0.18) 57	3.52 (0.18) 70	3.88 (0.18) 78	3.41 (0.10) 68

N for each age group = 39. "Item *A, C*" refers to average of items *A* and *C*. Maximum average score for each cell = 5. Numbers in parentheses are standard errors.

chance (Bryant & Trabasso, 1971) and restricted guessing chance (Wright & Dowker, 2002). Additionally, we adopted a basic psychophysics method for determining whether the levels of performance observed against our two criteria actually represent a deductive transitive competence. This is based around the notion that competent reasoners should be at least as close to perfect inference performance as to chance performance (a competence threshold —Mueller & Pickering, 1970). If a participant getting an answer wrong does so because they confuse item-*A* with item-*C*, as maintained by Bryant (1998), then random chance performance is 50%. The competence threshold therefore lays midway between 50% and 100%, which is 75%. Alternatively, if the problem is between item-*A* and item-*B*, rather than *A* versus *C* (Wright & Dowker, 2002), then we still end up with a competence threshold of 75%. However, if we make the common assumption that children who are not competent in transitive reasoning are not able to even partially order the items of the implied series (an assumption of Bryant & Trabasso, 1971), then we must hold that they guess among all three items, rather than between only two of them as described earlier. Consequently, random chance performance is 33.3% rather than 50%, and its associated competence threshold is now midway between 33.3% and 100%, which is 67%.

We compared each group's item-*A* performance against these assumptions about chance performance and competence thresholds, using a series of one-sample *t*-tests. The 6-year-olds were significantly above random chance (df = 38, *t* = 6.20, *p* < .01). When we adopted the 50% 2-choice chance criterion as outlined earlier, the 6-year-olds were still significantly above it (df = 38, *t* = 2.47, *p* = 0.02). However, at the same time, they were significantly below the 75% competence threshold as defined against 2-choice chance (df = 38, *t* = −3.13, *p* < .01).

For the 7-year-olds, their item-*A* performance did not differ significantly from the competence threshold defined against 2-choice chance (df = 38, *t* = −0.19, *p* = 0.85). The 8-year-olds were the only

group performing significantly above the competence threshold defined against 2-choice chance (df = 38, *t* = 2.53, *p* = 0.01).

We repeated the earlier analyses for the full-series questions, as these were the only questions requiring transitive inferences to be computed within mental space, with little potential cuing from the premise pairs continually on display (Brainerd & Reyna, 1992; Wright & Dowker, 2002). We again contrasted performance against both random chance and 2-choice chance. Note, as there are 6 ways of arranging items *A*, *B* and *C*, random chance here was 1 in 6 (or 16.7%). However, if, as our previous analysis of item-*A* performance leads us to suspect, reasoners tend to dismiss item-*C* from being a candidate for biggest item and then the issue is whether they realise that they already have the means to select between item-*A* and item-*B*, a more appropriate comparison is against 2-choice chance just as before.

The 6-year-olds were significantly above random chance performance on the full series (df = 38, *t* = 2.80, *p* < .01), but were not significantly below 2-choice chance (df = 38, *t* = −0.45, *p* = 0.66). They were, however, significantly below the competence threshold set against 2-choice chance (df = 38, *t* = −5.33, *p* < .01).

The 7-year-olds were significantly above 2-choice chance (df = 38, *t* = 4.18, *p* < .01). However, they were also significantly below the 75% competence threshold set against 2-choice chance (df = 38, *t* = −2.50, *p* = 0.02). The 8-year-olds were also significantly above 2-choice chance (df = 38, *t* = 5.58, *p* < .01). But, unlike the 7-year-olds, they did not perform significantly below the 75% competence threshold based on the 2-choice chance criteria (df = 38, *t* = −0.67, *p* = 0.51).

DISCUSSION

Children in the present study had the two premise pairs continually in view, reducing mental processing requirements (Ameel et al., 2007; Kallio, 1988; Riley & Trabasso, 1974; Trabasso, van den Broek, & Suh,

1989; Wright & Howells, 2008). Nevertheless, we found transitive reasoning to be quite demanding for 6-year-olds. This finding is at odds with labelling theorists (e.g. Bryant, 1998), who would have expected item-*A* performance to approach ceiling. In line with developmental three-term task advocates (e.g. Piaget et al., 1977), we found a substantial improvement in resolving item-*A* between 6 and 8 years (Artman & Cahan, 1993; Castle & Needham, 2007; Wright, 2006; Wright et al., 2011).

Markovits et al. (1995) showed that 4- and 5-year-olds tend to guess randomly when solving for item-*A* on three-term tasks. In an additional study (Wright et al., 2011), we confirmed this for 5-year-olds but 6-year-olds employed an intermediate strategy yielding better performance. Our present data reconfirm that by 6 years, children are not basing their transitivity judgements on random chance. Rather, they appear to resolve item-*C*. From that point, they use a combination of *A*:*B* guessing and genuine transitive judgements. By 8 years, children now routinely appreciate the need to coordinate both given premises in order to fully solve for item-*A*; and their increasing ability to successfully do this leads to performance now starting towards perfect performance. This is suggestive of phase-like development in transitive reasoning—1, random chance; 2, an influence of *A*:*B* guessing; 3, transitive competence (Wright, 2006; Wright & Dowker, 2002; see also later).

Turning to gender, we report for the first time in transitive research, that there was an overall advantage to one gender (boys). However, there was no systematic change in the advantage of boys across our three age groups either with age or with level of markedness. The overall advantage of boys on our tasks may be due to spatial abilities taking priority over verbal abilities between ages 6 and 8 years (Brunamonti et al., 2011; Hegarty et al., 2009; Strand et al., 2006). An importance of spatial processing to adult's transitive reasoning has recently been confirmed in two experimental studies of reaction-time (Brunamonti et al., 2011; Demarais & Cohen, 1998) and two brain imaging studies, one using a visual transitive task and the other using an auditory transitive task (Fangmeier & Knauff, 2009; Fangmeier, Knauff, Ruff, & Sloutsky, 2006). An implication for cognitive experimental research with children is that spatial tasks utilising high visual processing of the relation can interfere with construction of the mental model of the transitive series (Knauff & May, 2006), and hence at least in theory, it should be possible to reverse the gender advantage we have found here. This will be the subject of our future developmental research. But what is already clear is that transitive studies should begin to include analyses of gender effects.

As well as gender, another largely previously underresearched issue in transitive reasoning is markedness (which we operationalised here in terms of the relation supporting item-*A*'s position in the series versus that of item-*C*). On this issue, we acknowledged Trabasso's (1977) unmarked-item superiority position. That said, neither our findings here nor those of a number of other investigators seem to support that view (De Lillo et al., 2001; Wu & Levy, 2001). Presently, we have found that it was the most marked item of our transitive series (item-*C*) that resulted in highest performance, rather than the unmarked item (item-*A*). This finding appears not to have been explicitly reported previously. However, a number of published studies do present data which at least intimate a marked item superiority (e.g. see data tables in Favrel & Barrouillet, 2000; Frank, Ruby, Levy, & O'Reilly, 2005; Lazareva & Wasserman, 2010; Moses, Villate, & Ryan, 2006; Siemann & Delius, 1996; Wright & Dowker, 2002; Wright & Howells, 2008). The greater prominence of the marked (*C*) item implies that the transitive series is constructed starting from this item, rather than from the most unmarked end of the series as previously assumed.

Comparative transitive theorists could object that markedness is a decidedly human phenomenon and yet equivalent effects have been found in non-humans (e.g. Eichenbaum, 2001; Higa & Staddon, 1993). However, such effects may be due, not to deduction as such, but rather to perceptual or associative capacities fed by extensive training, as present in many species (Coleman et al., 2010; Premack, 2007; Siemann & Delius, 1996). Both in studies with humans and those with non-humans, associative (also known as reinforcement) accounts may be controlled for (e.g. Allen, 2006; Lazareva & Wasserman, 2010; Yamazaki, 2004). Importantly, though, reinforcement issues cannot be ascribed to the present study, because we did not train (reinforce) children on any premise, and both premises were actually in full view throughout each task.

Earlier we summarised Knauff's (2009) transitive theory as stating that deduction occurs in stage 3, where the mental model is formed symbolically. Here, transitive responses are reached by traversing this model and inspecting it to answer the specific question asked (e.g. "Who is the tallest?"). This theory intimates that the entire series is set out both spatially and symbolically at a point in time

that is before any question is asked, and so, simply reporting the entire series should be easier than the reasoner having to inspect it to answer a question about any one item (e.g. item-*A* or *C*). ANOVA showed that reporting the full series was far more difficult than answering for item-*A* or *C*. This finding is more in line with Kallio's (1988) theory, which proposed that integrating both premises to realise the transitive series constitutes a quaternary level of representation, the most demanding (highest) level for transitive relations (see also Halford et al., 1998).

One might argue that any direct comparison between single items versus the full series is invalid, because random chance performance is different in these respective cases. We would first state that it is not possible to test the predictions we generated from Kallio's versus Knauff's theory, without somehow directly comparing single item against full series. Next, we would point out that, in any case, our final analyses indicate that children do not simply go from random chance to competent performance; and indeed, random chance is not even an issue for 6–8-year-olds. Rather, in the case of both individual items and the full series, the issue is about moving from a level reflecting 2-choice chance (i.e. the child having difficulty deducing between the two items that had been given a positive label). For both individual items and the full series then, the most appropriate level of chance for 6–8-year-olds is the same (50%), and the competence threshold set against this level is 75%. Thus, we would argue that it is legitimate to directly compare single items versus the full series to these levels.

Our final analyses showed that the 6-year-olds were better than 2-choice chance for item-*A,* but not for the full series, and did not reach competence in either case. The 7-year-olds, although above 2-choice chance in both instances, reached competence for item-*A,* but not for the full series. The 8-year-olds were again above 2-choice chance, and were also above the competence threshold for item-*A*, although reaching but not exceeding it for the full series. This reconfirms that full-series performance lags behind item-*A* performance. Indeed, our data suggest that this lag is equivalent to approximately 1 year's cognitive development (see Table 3 earlier).

Can we explain the greater difficulty of full series within the three-level structure of Knauff's theory (e.g. Knauff, 2009; Knauff & May, 2006)? For Piaget et al. (1977), deductive transitivity is largely about understanding and embracing that the middle term (e.g. item-*B*) can have two relative values (one

against *A* and the other against *C*), and can therefore be used to link item-*A* indirectly to item-*C*. This represents a cognitive concept that is acquired at around 7 years. In opposition to Piaget's view of transitivity being a challenging concept for some groups, some (e.g. Bonnefond et al., 2014) argue that applying transitivity is trivial, and it is establishing the resultant mental model that is demanding (e.g. Bonnefond et al., 2014, p. 101). Others argue along similar lines to Piaget on this issue. For example, Halford and Andrews (2004) state that "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, p. 126). Knauff (2009) summarised brain research showing that integration occurs at stage 2 rather than at stage 3. One implication is that stage 3 may therefore be a post-transitive stage more to do with re-describing what occurred at stage 2 in symbolic terms plus performing an inspection of the re-described array which itself does not approximate deduction (see Trabasso, 1977; Wright, 2001). An alternative is to posit that deduction actually occurs both across Knauff's stages 2 and 3, rather than being confined to stage 3 only.

We favour the second alternative here, and have previously reported brain research studies which show that both pre-frontal cortex and parietal cortex are involved in the deduction (Wright, 2012). In behavioural terms, it is possible to construct a transitive-like series at stage 2, for a relation such as "is next to"; but it must be accepted that only at stage 3 can the reasoner choose to accept or reject the resultant symbolic model as transitive and therefore valid (Lazareva & Wasserman, 2010). For instance, consider *A* left of *B*, *B* left of *C*; this implies *A* left of *C* (Knauff, 2009). This *A*:*C* relationship holds if *A*, *B* and *C* are along a straight line from left to right. But some reasoners might conceive of a situation where they are not—for example, perhaps you sometimes eat with friends at dinner tables which are round. The *A*:*C* comparison does not hold if *A*, *B* and *C* are equidistant around such a table (Wright, 2001). Indeed, in this situation, the *A*:*C* relationship is *C* left of *A*, not *A* left of *C*. If a reasoner constructs only the linear mental model, then the $A > C$ conclusion follows; but if s/he constructs both these models, then s/he should hold that the *A*:*C* comparison is indeterminate without further disambiguating information.

Our findings regarding single items versus full-series reporting then do not necessarily refute Knauff's (2009) three-stage spatial theory of transitive reasoning. However, our findings do call for a

closer look at the notion that deduction occurs only in one stage: We suggest that it occurs across two stages, one for applying transitivity to yield premise integration, and the other to evaluate the nature of the relation by constructing one or more mental model of the integrated premises (Wright, 2001). The challenge now is to devise both experimental and neuroscience studies capable of testing between the two alternative conceptions offered here.

CONCLUSIONS

We conducted a transitive study with 6–8-year-old children intended to be more ecologically valid than many previous studies in two ways. First, we relied on a composite transitive score from five different contents, rather than using just a single transitive relation. Second, we avoided issues of training and memory, by ensuring that the two premises were visible to the child at all times. Using this procedure, we investigated four main issues. On these issues, first, we found that our more ecologically valid procedure still gave results quite typical of other three-term tasks. Thus, the assumption from labelling theorists (e.g. Bryant, 1998) that such tasks are invalid because of being too easy, with children solving for item-A via non-transitive labelling strategies, was not supported. Second, we found a gender difference for transitive reasoning. Specifically, in line with their supposed superior spatial abilities, boys presented higher levels of transitive performance than did girls. It is therefore worth including analyses of gender in future transitive studies. Third, developmental research has tended to assert that item-A, the item uniquely associated with the marked relation, is the pivot of the transitive series; being solved both first plus at the highest level of all items in the series. However, contrary to this assumption, it was our marked item-C that was associated with higher performance than our unmarked item-A.

Finally, in line with Kallio's (1988) relational theory, we found that ordering the transitive series ($A > B > C$) is more difficult than solving for item-A. This suggests that a transitive deductive competence may be used to integrate the entire series, rather than the series being integrated first and deduction applied purely to inspect that series to solve for any items (contrast Knauff's, 2009 spatial theory). Our finding suggests a minor amendment to Knauff's theory: Deduction may occur across two levels rather than just one. These are the

integration of premises into a transitive-like series (stage 2), followed by the testing of the resultant mental model (stage 3) plus the search for alternative models that potentially render the series invalid or ambiguous. Insofar as these findings and our suggested amendment represent initial inroads, they are open to further confirmation or challenge. However, regardless, it is hoped that our present demonstrations open up further possibilities of other aspects of transitive reasoning which hitherto may have been underinvestigated.

REFERENCES

Acuna, B., Eliassen, J., Donoghue, J., & Sanes, J. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cerebral Cortex*, *12*, 1312–1321.

Andrews, G. (2010). Belief-based and analytic processing in transitive inference depends on premise integration difficulty. *Memory & Cognition*, *38*(7), 928–940

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.

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.

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

Bonnefond, M., Castelain, T., Cheylus, A., & Van der Henst, J. (2014). Reasoning from transitive premises: An EEG study. *Brain and Cognition*, *90*, 100–108.

Brainerd, C., & Reyna, V. (1992). The memory independence effect: What do the data show; what do the theories claim? *Developmental Review*, *12*, 164–186.

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

Brunamonti, E., Genovesio, A., Carbe, K., & Ferraina, S. (2011). Gaze modulates non-professional reasoning: Further evidence for special representation of reasoning premises. *Neuroscience*, *173*, 110–115.

Bryant, P. (1998). Cognitive development. In M. Eysenck, (Ed.), *Psychology an integrated approach* (pp. 260–292). London: Wesley Longman.

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*, 456–458.

Carmesin, H. O., & Schwegler, H. (1994). Parallel versus sequential processing of relational stimulus structures. *Biological Cybernetics*, *71*, 523–529.

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

Chen, D., Lu, H., & Holyoak, K. J. (2014). The discovery and comparison of symbolic magnitudes. *Cognitive Psychology*, *71*, 27–54.

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

Coleman, M. J., Titone, D., Krastoshevsky, O., Krause, V., Huang, Z., Mendell, N. R., … Levy, D. L. (2010). Reinforcement ambiguity and novelty do not account for transitive inference deficits in Schizophrenia. *Schizophrenia Bulletin*, *36*(6), 1187–1200.

De Lillo, C., Floreano, D., & Antinucci, F. (2001). Transitive choices by a simple, fully connected, back-propagation neural network: Implications for the comparative study of transitive inference. *Animal Cognition*, *4*, 61–68.

Demarais, A. M., & Cohen, B. H. (1998). Evidence for image-scanning eye movements during transitive inference. *Biological Psychology*, *49*, 229–247.

De Soto, L. B., London, M., & Handel, M. S. (1965). Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology*, *2*, 513–521.

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

Evans, J. S. B., Newstead, S. E., & Byrne, R. M. (1993). *Human reasoning: The psychology of deduction*. Hove: Psychology Press, Taylor & Francis Group.

Fangmeier, T., & Knauff, M. (2009). Neural correlates of acoustic reasoning. *Brain Research*, *1249*, 181–190.

Fangmeier, T., Knauff, M., Ruff, C. C., & Sloutsky, V. (2006). FMRI evidence for a three-stage model of deductive reasoning. *Journal of Cognitive Neuroscience*, *18*, 320–334.

Favrel, J., & Barrouillet, P. (2000). On the relation between representations constructed from text comprehension and transitive inference production. *Journal of Experimental Psychology*, *26*(1), 187–203.

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

Guez, D., & Audley, C. (2014). Transitive or not: A critical appraisal of transitive inference in animals. *Ethology*, *119*, 703–726.

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

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

Hegarty, M., Keehner, M., Kooshabeh, P., & Montello, D. P. (2009). How spatial abilities enhance, and are enhanced by dental education. *Learning and Individual Differences*, *19*(1), 61–70.

Higa, J. J., & Staddon, E. R. (1993). Transitive inference in multiple conditional discriminations. *Journal of the Experimental Analysis of Behavior*, *59*, 265–291.

Holcomb, W. L., Stromer, R., & Mackay, H. A. (1997). Transitivity and emergent sequence performances in young children. *Journal of Experimental Child Psychology*, *65*(1), 96–124.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press.

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

Kim, Y. A., & Song, H. S. (2011). Strategies for predicting local trust based on trust propagation in social networks. *Knowledge-Based Systems*, *24*, 1360–1371.

Knauff, M. (2009). A neuro-cognitive theory of deductive relational reasoning with mental models and visual images. *Spatial Cognition & Computation: An Interdisciplinary Journal*, *9*(2), 109–137.

Knauff, M., & May, E. (2006). Mental imagery, reasoning, and blindness. *Quarterly Journal of Experimental Psychology*, *59*, 161–177.

Krawczyk, D. C. (2012). The cognition and neuroscience of relational reasoning. *Brain Research*, *1428*, 13–23.

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

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*, 600–613.

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.

Mills, C. J., Ablard, K. E., & Stumpf, H. (1993). Gender differences in academically talented young students' mathematical reasoning: Patterns across age and subskills. *Journal of Educational Psychology*, *85*(2), 340–346.

Moses, S. N., Villate, C., & Ryan, J. D. (2006). An investigation of learning strategy supporting transitive inference performance in humans compared to other species. *Neuropsychologia*, *44*, 1370–1387.

Mueller, J. H., & Pickering, J. P. (1970). Correction procedures in three-alternative verbal-discrimination lists with intertriad associations. *The American Journal of Psychology*, *83*, 401–411.

Piaget, J. (1921). Une forme verbale de la comparaison chez l'enfant. *Archives de Psychologie*, *18*, 141–172.

Piaget, J. (1965). *Judgment and reasoning in the child*. London: Routledge & Kegan Paul.

Piaget, J., & Garcia, R. (1991). *Toward a logic of meanings*. Hillsdale, NJ: Erlbaum (Original work published 1987).

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

Premack, D. (2007). Human and animal cognition: Continuity and discontinuity. *Proceedings of the National Academy of Sciences*, *104*(35), 13861–13867.

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*(3), 561–588.

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

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

Sedek, G., Piber-Dabrowska, K., Maio, G. R., & Von Hecker, A. (2011). Individual differences in prejudice and associative versus rule-based forms of transitive reasoning. *European Journal of Social Psychology, 41*, 853–865.

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

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

Strand, S., Deary, I. J., & Smith, P. (2006). Sex difference in cognitive abilities test scores: A UK national picture. *British Journal of Educational Psychology, 76*, 463–480.

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

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* (pp. 201–230). Hillsdale, NJ: Erlbaum.

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

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

Wright, B. C. (2012). The case for a dual-process theory of transitive reasoning. *Developmental Review, 32*(2), 89–124.

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 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.

Wu, X., & Levy, W. B. (2001). Simulating symbolic distance effects in the transitive inference problem. *Neurocomputing, 38–40*, 1603–1610.

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