

**The relationship between numerosity discrimination and arithmetic skill reflects  
the Approximate Number Sense and cannot be explained by inhibitory control**

Stephanie A. Malone

Verena E. Pritchard

Michelle Heron-Delaney

Institute for Learning Sciences and Teacher Education, Australian Catholic University

Kelly Burgoyne

University of Manchester and Australian Catholic University

Arne Lervag

Department of Education, University of Oslo

Charles Hulme

University of Oxford and Australian Catholic University

Author Note:

This research was supported by a grant from the Australian Catholic University.

Correspondence concerning this manuscript should be addressed to Charles Hulme,  
Department of Education, University of Oxford, 15 Norham Gardens, Oxford, OX2  
6PY

Email: [charles.hulme@education.ox.ac.uk](mailto:charles.hulme@education.ox.ac.uk)

### Abstract

Numerosity discrimination tasks (judging which of two random dot arrays contains the larger number) have been widely used as a measure of the efficiency of an Approximate Number Sense and are a correlate of early arithmetic skills. Recently, it has been suggested that the relationship between numerosity discrimination and arithmetic is explained by inhibition rather than the Approximate Number Sense. We assessed this hypothesis in a study of 496 children (mean age = 81.23 months) using numerosity discrimination tasks which manipulated the congruency between surface area and numerosity. Numerosity discrimination for incongruent arrays (which are postulated to require inhibition due to a conflict between judgements based on surface area rather than numerosity) was more difficult than for congruent arrays. However, all numerosity discrimination tasks showed substantial correlations with each other, and correlated with arithmetic. A latent variable path model showed that a general numerosity judgement factor correlated with arithmetic even after controlling for a measure of response inhibition. In contrast, numerosity discrimination for incongruent arrays showed no unique relationship with arithmetic ability. Our results do not support the view that the relationship between numerosity discrimination and arithmetic is largely attributable to inhibition; rather, they are consistent with the view that magnitude discrimination tasks tap the operation of an Approximate Number Sense.

**Key words:** Approximate Number Sense (ANS), arithmetic development, numerosity judgment, inhibition

## The relationship between numerosity discrimination and arithmetic skill reflects the Approximate Number Sense, not inhibitory control

The Approximate Number System (ANS) codes magnitudes in an analogue fashion and the acuity of this system has been claimed to be a constraint on the development of number skills (Booth & Siegler, 2006; Dehaene & Cohen, 1995). Meta-analyses have found a significant, albeit weak, relationship between numerosity judgments and arithmetic ability (Fazio et al., 2014; Schneider et al., 2017). This relationship has often been interpreted with reference to the Triple Code Model (Dehaene & Cohen, 1995), suggesting that representations of numerosity (magnitude) provide the semantic content for Arabic numerals and number words. Recently, however, Gilmore et al. (2013) have put forward an alternative view. Specifically they suggest the inhibitory control component of numerosity judgment tasks is a major driver of this relationship. In this study, we examine this alternative interpretation in a large sample of children during the early years of formal education.

Non-symbolic magnitude judgement ability is typically assessed using a dot comparison task (e.g., Bonny & Lourenco, 2013; Libertus, Feigenson, & Halberda, 2013), where participants are asked to identify the more numerous of two dot arrays. Arrays that are further apart in numerical distance (e.g., 5 vs. 10) are easier to discriminate than those that are closer (e.g., 5 vs. 8) (Moyer & Landauer, 1967). This “distance effect” is commonly manipulated by varying the ratio between arrays. ANS acuity increases during development with the ratio between arrays required for reliable discrimination reducing from 1:2 for infants, to 4:5 for 5-year-olds, and finally, 10:11 for adults (Halberda & Feigenson, 2008; Xu & Spelke, 2000). ANS acuity is a stronger correlate of arithmetic for children than adults (Chen & Li, 2014; Fazio et al., 2014; Schneider et al., 2017).

In the dot comparison task continuous perceptual features (such as the total surface area of an array) may offer additional cues to numerosity (Gebuis & Reynvoet, 2012; Gilmore et al., 2013; Leibovich, Katzin, Harel, & Henik, 2017). For example, when dot sizes are consistent across arrays, the more numerous array covers a larger surface area than the less numerous array. To control for this the total surface area of the arrays and the size and density of dots within a display are often manipulated. For example, in some displays the more numerous set of dots may be composed of smaller dots so that the total surface area of two displays is equated. Other studies have used conditions where perceptual features (such as surface area) are positively (congruent) or negatively (incongruent) correlated with numerosity (e.g., Bugden & Ansari, 2015; Fuhs & McNeil, 2013).

Gilmore et al. (2013) noted differences in the inhibitory control demands of the different types of trial commonly used in dot comparison tasks. For incongruent trials (where the more numerous display consists of smaller dots) children need to inhibit their initial response based on a salient perceptual feature (surface area) in order to select the more numerous array. In contrast, for congruent trials perceptual features and numerosity are aligned, meaning that inhibition is not necessary in order to select the correct response. As inhibitory control is also a predictor of arithmetic (Bull & Lee, 2014; Bull & Scerif, 2001; Kroesbergen, Van Luit, Van Lieshout, Loosbroek, & Van de Rijt, 2009), Gilmore et al. (2013) hypothesised that the shared influence of inhibition on both arithmetic and incongruent dot comparison trials may be responsible for the relationship observed between dot comparison tasks and arithmetic. Taken to the extreme, this view would see the relationship between dot comparison judgements and arithmetic as explicable purely in terms of the role of inhibitory control in both tasks, and not as evidence for the role of the ANS in

arithmetic development. Clayton and Gilmore (2015) however propose a more nuanced “competing processes account” which holds that both ANS acuity and inhibitory control influence children’s numerosity judgements. Specifically, acuity is engaged to determine the more numerous array while inhibitory control is employed to suppress any prepotent response based on salient perceptual features (e.g., total surface area).

In addressing the role played by inhibition in the magnitude-arithmetic relationship, it is important to consider the development of inhibitory skills in general and to distinguish different types of inhibitory processes that may be measurable and applicable to the assessment of the magnitude-arithmetic relation in mid-childhood. Inhibitory control is thought to develop across childhood (e.g., Williams, Ponesse, Schachar, Logan, & Tannock, 1999), with the ability to inhibit a prepotent response developing most rapidly between 4- and 6-years-old (Tillman, Thorell, Brocki & Bohlin, 2008; see also Wiebe, Sheffield & Espy, 2012). This ability has been postulated to involve four separable components: response inhibition, interference control, cognitive inhibition and oculomotor inhibition (Nigg, 2000). Over time, these skills become increasingly differentiated (Howard, Johnson, & Pascual-Leone, 2014), and may vary as a consequence of IQ and age (Jonkman, Lansbergen, & Stauder, 2003; Lee, Lo, Li, Sung, & Juan, 2015; Davis & Anderson, 2001). Of these types, both response inhibition and interference control, which are often referred to interchangeably as forms of inhibitory control, have been found to relate to mathematics ability (St Clair-Thompson & Gathercole, 2006; Szucs, Devine, Soltey, Nobes, & Gabriel, 2013).

To investigate the specific role of inhibition in magnitude discrimination, Gilmore et al. (2013) conducted two studies. First, with a sample of 80 children

(mean age = 7.7 years;  $SD = 1.9$  years, range = 4.7 – 11.9 years) they compared the relationship between arithmetic skills and performance on a numerosity discrimination task consisting of either congruent or incongruent trials. They reported that performance on incongruent, but not congruent, magnitude trials correlated with arithmetic ability (incongruent:  $r = .55, p < .001$ ; congruent:  $r = .03, p = .82$ ). This, along with slower reaction times for incongruent trials, was interpreted as suggesting that the additional processing step of inhibiting the initial response based on perceptual features (such as surface area) was a critical determinant of the magnitude judgment-arithmetic relationship.

To assess the influence of inhibition on the magnitude-arithmetic relation more directly, the second study by Gilmore et al. (2013) ( $n = 71$ ; mean age = 9.4 years;  $SD = .6$ ; range = 7.8 – 10.5 years) included a measure of response inhibition (NEPSY-II Inhibition subtest; Korkman, Kirk, & Kemp, 2007) in addition to a dot comparison (numerosity judgement) task. Although performance on the numerosity judgement task (a measure which combined congruent and incongruent trials) predicted arithmetic ( $r = .35$ ), once the variance associated with response inhibition was removed it was no longer a significant predictor. Although taken together, the findings of these two studies appear to provide initial support for the explanatory role of response inhibition, they should be interpreted with caution. Specifically, Gilmore et al. (2013) posit that it is the shared influence of inhibition on incongruent trial performance and arithmetic that accounts for this relationship. However, this was not explicitly tested; instead the impact of inhibitory control was only examined in relation to a single, general measure of magnitude comparison (see Study 2; Gilmore et al., 2013).

Fuhs and McNeil (2013) compared the influence of congruency in the dot comparison task on its correlation with mathematics performance in preschool children (mean age = 55 months). Three conditions in the dot discrimination task were used: surface area positively correlated with numerosity (congruent), average stimulus size negatively correlated with numerosity (incongruent), and inverse trials whereby surface area was inversely related to numerosity (e.g., if the numerosity differed by 2:3, the surface area between arrays would differ by 3:2; inverse incongruent). Because of the very young age of the children in this study, a pure measure of arithmetic was not used, and instead mathematics proficiency was assessed by the Test of Early Mathematics Skills (TEMA; Ginsburg & Baroody, 2003) which assesses a broad range of abilities including counting, number knowledge and arithmetic. Interestingly only performance on inverse incongruent trials ( $r = .23$ ), and not simple incongruent ( $r = .08$ ) or congruent trials ( $r = .15$ ), was a significant correlate of mathematics. However, once response inhibition was accounted for, inverse trial performance was no longer a significant predictor of mathematics. It is also noteworthy that these correlations between numerosity judgement and mathematics were much lower than is typically observed with this age group ( $r_s = .30 - .40$ ; Fazio et al., 2014; Schneider et al., 2017).

Although the findings of Fuhs and McNeil (2013) appear to lend further support to the influence of inhibitory control on magnitude judgements, it is important to note that they did not have a pure measure of arithmetic ability. Because of the broad range of skills assessed by the TEMA it is not possible to determine the specific role of inhibition in explaining the predictive relationship between magnitude discrimination and arithmetic from this study. In addition this study focused exclusively on children in the very earliest stages of arithmetic development.

The present study investigates the role of inhibition in accounting for the correlation between numerosity judgement tasks and arithmetic skills in children. Most critically, we present the results of a large-scale experimental study where we manipulated the congruency between surface area and numerosity, in a dot judgement task. Our dot-comparison task involves 8 conditions that systematically manipulate the ratio (2:3 and 5:6) and congruency (2 congruent and 2 incongruent per ratio) of stimuli. This allows us to assess the role of numerical distance and congruency on performance on the numerosity discrimination task. According to the arguments of Gilmore et al. (2013) only performance on incongruent trials (where there is a conflict between numerosity and the competing cue of surface area) should correlate with arithmetic ability. According to the more nuanced “competing processes” account of Clayton and Gilmore (2015) we would expect the correlation to be stronger for incongruent trials (but perhaps present, if weaker, for congruent trials). In addition to experimentally manipulating congruency in the dot comparison task we also administered a test of response inhibition to all children (the Head-Toes-Knees-Shoulders task; HTKS). This task is appropriate for 4-8 year olds (McClelland et al., 2014), which is the age group most typically assessed in studies examining the role of inhibition in mathematics (Fuhs & McNeil, 2013; Gilmore et al., 2013).

In a methodological advance, the use of multiple assessments of arithmetic (addition and subtraction) and numerosity judgment allows latent variable models to be constructed, correcting for measurement error (Cole & Preacher, 2014).

## **Method**

### **Participants**

Four hundred and ninety-six children from 11 primary schools (10 fee paying and 1 publically funded) in Brisbane, Australia (225 girls; mean age = 81.23 months,



$SD = 4.25$ , range 71-99 months) were assessed during their second year of formal schooling (i.e. Year 1). Government data (Index of Community Socio-Educational Advantage; ICSEA) indicates that eight of the schools serve areas with an average level of social advantage, while the remaining three have a student population with higher levels of social advantage. The Australian Catholic University Human Research Ethics Committee (2015-141H) provided ethical approval, and consent was obtained from parents and children prior to participation.

### **Materials and Procedure**

All assessments of arithmetic and inhibition were individually administered. Non-verbal IQ and magnitude comparison assessments were group administered to the whole class. Children were also assessed on a range of additional measures (including language, reading, and motor tasks) as part of a larger longitudinal study.

***Magnitude comparison.*** All conditions were presented in a 14.8 x 21cm booklet, with six stimulus pairs displayed per page (one per row). Each pair consisted of two arrays of 10 to 30 small black squares displayed in 2.2 cm<sup>2</sup> boxes (see Figure 1). Presentation order of the conditions was determined by a Latin square design. For each condition, children were asked to tick the more numerous array in as many stimulus pairs as possible within 30 seconds. Chance performance on this task is 50% correct; scores for children who performed at chance or below on any condition were excluded since such scores reflect guessing only. Where scores are above chance they reflect the efficiency with which a child can make the discrimination required. Although magnitude comparison is typically assessed using computerized tasks, paper-and-pencil based methods have been used successfully in other studies (Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013) and correlations with arithmetic are similar to those reported in meta-analyses of studies typically based on

computer-presented tasks (Chen & Li, 2014; Fazio et al., 2014; Schneider et al., 2017).

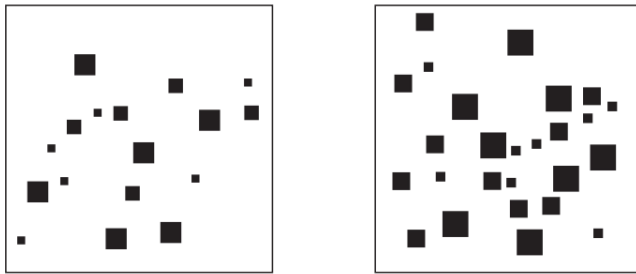


Figure 1. Example of a 2:3 super congruent stimulus pair.

Eight conditions were administered, systematically varying the ratio between stimulus pairs (2:3 or 5:6) and congruency between numerosity and dot surface area (super congruent, congruent, incongruent, super incongruent). For each ratio, ten exemplars were selected with the individual numerosities rounded to the closest whole number (2:3 ratio: 10:15, 11:17, 12:18, 13:20, 14:21, 15:23, 16: 24, 17:26, 18:27, 20:30; 5:6 ratio: 10:12, 11:13, 13:16, 15:18, 16:19, 17:20, 19:23, 21:25, 23:28, 25:30). Each exemplar was repeated four times within a single condition resulting in a total of 40 trials. Stimulus pairs were arranged in a random order and counterbalanced for the location (left or right) of the more numerous array.

To manipulate congruency, the arrays were varied along two continuous dimensions: average square size and total surface area. In the congruent condition, the average square size was matched across arrays meaning that the total surface area was positively correlated with numerosity. The incongruent condition matched the total surface area across arrays, resulting in the average square size being negatively correlated with numerosity. In the super congruent condition, both total surface area and average square size were positively correlated with numerosity, while in the

super incongruent condition, total surface area and average square size were inversely related to numerosity.

***Inhibition*** was assessed using a modified version of the HTKS task (Burrage, Ponitz, McCready, Shah, Sims, & Jewkes, 2008). This consisted of two parts: Part 1 (“touch your head/toes”; 6 trials) and Part 2 (“touch your head/toes” and “shoulders/knees”; 14 trials). For each trial, the researcher told the child to touch a body part, and the child was to inhibit their response and instead complete the opposite action (e.g., touch head when requested to touch toes). Two points were awarded if the child made the correct action, one point for self-corrected trials (i.e. an initial movement made towards to incorrect body part, before self-correcting), and zero points awarded to the incorrect action.

***Arithmetic.*** The addition and subtraction subtests of the Test of Basic Number and Arithmetic Skills (Hulme, Brigstocke, & Moll, 2016) were used to assess arithmetic. All items were presented in a 21 x 29.7cm booklet, with the written sums displayed in two columns of 15 items. For each subtest, children were provided with 60 seconds to answer as many items as possible.

***Non-verbal cognitive ability*** was assessed using a version of Raven’s Coloured Progressive Matrices (Raven, 1995), adapted for group administration. Children were presented with 12 puzzles and asked to indicate which of six potential pieces completed each puzzle. A score of 1 was provided for each correct response.

## **Results**

The means, standard deviations, skewness and reliabilities of variables are shown in Table 1. The number of children varies across the different magnitude discrimination tasks reflecting the fact that not all children performed above chance in all conditions. Correlations between the measures are shown in Table 2.

It is clear that performance in the dot comparison task is affected by ratio (performance is better in the 2:3 conditions than the 5:6 conditions) and congruency (performance is better in the congruent conditions). This was confirmed by a 2 (ratio) x 4 (congruency) repeated measures ANOVA which revealed a significant main effects of ratio ( $F(1, 195) = 492.72, p < .001, \eta_p^2 = .72$ ) and congruency ( $F(3, 585) = 129.21, p < .001, \eta_p^2 = .40$ ). The interaction term ratio x congruency was non-significant ( $F(3, 585) = .95, p > .05, \eta_p^2 = .01$ ), although subsequent analysis of the simple main effects revealed that performance in the 2:3 condition was better than in the 5:6 condition in all four congruency conditions ( $ps < .001$ ). This pattern - better performance on arrays with smaller ratios, coupled with rapid performance (average solution time of 1.68 seconds per correct item in the 2:3 condition, and 2.27 seconds in the 5:6 condition) - provides evidence that these numerosity comparison tasks are indeed tapping an ANS system and that children are not performing the tasks by counting (which would result in much slower performance).

We wished to examine the relationship between measures of magnitude judgment, inhibition (HTKS) and arithmetic. It is clear from Table 2 that all the magnitude discrimination tasks show substantial correlations with each other and correlate with addition and subtraction, regardless of congruency. To begin with confirmatory factor analyses were used to assess the factor structure of the magnitude discrimination tasks. A two-factor model in which congruent and incongruent conditions loaded onto separate factors provided a good fit to the data ( $\chi^2(19) = 42.35, p < .0002$ ). Although the two factors were very strongly correlated ( $r = .93$ ), a single factor model had a considerably worse fit (chi-square difference test,  $\chi^2(1) = 11.689, p < .001$ ). This was verified by a lower Bayesian Information Criteria (BIC)

for the two-factor model ( $BIC = 18413.641$ ) compared to the one-factor model ( $BIC = 18419.212$ ).

The theoretical issue at stake is whether incongruent magnitude judgements play a key role in predicting arithmetic skills given the possible influence of inhibition on both abilities (Gilmore et al. 2013). The very high correlation (.93) between the two magnitude judgement factors (congruent and incongruent) suggests they cannot be meaningfully separated – they are collinear and have 86% of variance in common. A simple path model regressing an Arithmetic latent variable on the congruent and incongruent magnitude judgement latent variables accounted for 24% of the variance in Arithmetic ( $p < .001$ ) but neither of the two regression coefficients was significant (their very large standard errors reflecting collinearity between the predictors.)

In order to identify whether incongruent magnitude judgements play a special role in predicting arithmetic (after accounting for the considerable variance in magnitude judgements that is common to both congruent and incongruent judgement tasks) we therefore constructed a bi-factor model. In this model a single magnitude judgment latent variable is defined using all of the magnitude tasks regardless of congruency while a second incongruent magnitude judgment latent variable is defined by the incongruent judgment variables only (see Hulme, Nash, Gooch, Lervag & Snowling, 2015, for the application of a similar model). These two factors were fixed to be uncorrelated so that the incongruent magnitude judgment factor captures the variance in these measures that is separable from a common magnitude judgment factor. This model therefore allows us to assess the influence that incongruent judgments may have on arithmetic that is independent of a general magnitude judgment ability.

The model used (see Figure 1) was estimated with MPlus 8.1 (Muthén & Muthén, 1998-2017) using robust maximum likelihood estimators to allow for deviations from normality in some variables. Missing values were handled using Full Information Maximum Likelihood estimation. The three latent variables of magnitude judgment, incongruent magnitude judgment and inhibition are direct predictors of concurrent arithmetic ability. Inhibition was defined by a single observed variable, with the error variance constrained to 1-reliability of the test. All observed variables were also regressed on non-verbal cognitive ability and age (though these regressions are not shown in the figure); hence the model represents relationships between variables that are independent of the shared variance attributable to non-verbal cognitive ability and chronological age. Both overall magnitude judgment and inhibition predicted arithmetic, however the incongruent magnitude judgment factor played no additional role in explaining arithmetic performance. The pattern of correlations between the latent variables is also informative (see Table 3): although Gilmore et al. (2013) suggested that inhibition was critical for incongruent magnitude discrimination, we found a non-significant correlation between our measure of response inhibition and incongruent magnitude judgment.

Overall, the model accounts for 32% of the variance in arithmetic performance and provides a good fit to the data ( $\chi^2(48) = 98.59, p < .001$ , Root Mean Square Error of Approximation (RSMEA) = .048 (90% CI = .034 - .061), Comparative Fit Index (CFI) = .971).

As a further test of whether incongruent magnitude judgements and inhibition are unique predictors of arithmetic (after accounting for congruent magnitude judgements) we estimated a Cholesky-factor model. Cholesky-factoring mimics

hierarchical regression with latent variables (see de Jong, 1999). In this model we used three latent variables as predictors of arithmetic: congruent magnitude judgements, incongruent magnitude judgements and inhibition (HTKS). We included the congruent magnitude factor at the first step, and then added the latent incongruent magnitude factor at the second step and finally inhibition at the third step as predictors. At the first step, the congruent magnitude factor predicted 22.8 % ( $p < .001$ ) of the variance in arithmetic. After controlling for the congruent magnitude factor, the incongruent magnitude factor was not a significant additional predictor (1.4% of the variance in arithmetic,  $p = .262$ ). However, after controlling both the congruent and the incongruent magnitude factors as predictors, inhibition explained an additional 6.5 % ( $p < .001$ ) of the variance in arithmetic. This model fitted the data well,  $\chi^2(53) = 111.226$ , RMSEA = .049 (90% CI .036, .061), CFI = .964. The results from this model are therefore essentially in line with the bifactor model shown in Figure 1 and confirm that incongruent magnitude judgement tasks play no special role (over congruent magnitude tasks) as a predictor of arithmetic. Both models also show that our measure of inhibition was a unique predictor of arithmetic.

### Discussion

This study investigated the possible role of inhibition on the numerosity judgment-arithmetic relationship. We experimentally manipulated the ratio of the numerosity discrimination and the surface area of the dots used in the magnitude judgement task to produce arrays where surface area was congruent or incongruent with numerosity. We took the novel approach of separating the variance associated with performance on incongruent trials from a common magnitude judgment ability. This allowed us to assess the unique role of incongruent judgments in the magnitude-arithmetic relationship. Our results fail to support Gilmore et al.'s (2013) conjecture

that the magnitude judgment-arithmetic relationship is an artifact of the inhibitory control demands of incongruent trials in magnitude judgement tasks. Performance on congruent and incongruent trials in the magnitude judgement task were very highly correlated ( $r = .91$  between an incongruent and congruent numerosity judgement factor). However, the unique variance attributable to performance on the incongruent trials was not a correlate of arithmetic. Instead, general magnitude judgment performance predicted arithmetic over and above both inhibition and the variance specific to incongruent magnitude judgments.

One of the central tenets of the inhibition hypothesis is that the shared influence of inhibitory control underpins the relationship between incongruent numerosity judgments and arithmetic (Gilmore et al., 2013). This suggests that inhibitory control would affect performance on incongruent trials as well as arithmetic. Although our findings lend further support to the growing body of evidence indicating that inhibitory control is a correlate of arithmetic (e.g., Bull & Scerif, 2001; Robinson & Dubé, 2013), we found inhibition to be associated only with *general* magnitude judgments ( $r = .35$ ) rather than specifically incongruent judgments ( $r = -.04$ ). These findings echo those of Keller and Libertus (2015) who also observed a significant correlation between inhibition and overall magnitude performance (congruent and incongruent combined;  $r = -.24$  where inhibition was negatively scored). Taken together, these findings have important implications for the inhibitory control hypothesis put forward by Gilmore and colleagues (2013) as our findings indicate that inhibition is an important element in the general construct of ANS acuity rather than being specifically dependent on features related to congruency in the magnitude judgement task. It may be that irrespective of congruency, children need to inhibit their initial response (based on salient perceptual features) in order to check



which of the two arrays is more numerous. This dovetails with the later competing processes account (Clayton & Gilmore, 2015).

We next investigated the magnitude-arithmetic relationship, and the specific role of inhibition. Our findings conflicted directly with those of Gilmore et al. (2013) in two notable ways: 1) *general* magnitude discrimination, and not *incongruent* discrimination, predicted arithmetic; and 2) even after removing the variance associated with inhibitory control, general magnitude judgment remained a unique predictor of arithmetic. These findings indicate that it is the ability to discriminate between quantities regardless of congruency that is predictive of arithmetic, rather than the shared influence of inhibition. The findings of Halberda, Mazocco, and Feigenson (2008) and Keller and Libertus (2015) also support this interpretation: both controlled for executive function (including inhibition) and found that ANS acuity remained a predictor of mathematics as assessed using the TEMA ( $r_p^2 = 0.17$  and  $r = .36-.45$ , respectively). Our study extends these findings to arithmetic ability rather than the broader definition of mathematics assessed by the TEMA.

Instead of supporting the idea that inhibitory control drives the magnitude understanding-arithmetic relationship (Gilmore et al., 2013), our findings align better with the more common interpretation of the magnitude-arithmetic relationship. That is, the data here demonstrate a hallmark of the ANS: numerosity discrimination was better when the numerical distance between arrays was larger (i.e. the distance effect; Moyer & Landauer, 1967). This finding supports the view that children use the ANS when identifying the larger of two arrays: if they were counting items in the array we would expect them to be quicker in arrays where the numbers are closer. This that in line with the later competing processes account (Clayton & Gilmore, 2015) that argues ANS acuity is initially employed to determine the more numerous array.

## **Limitations and Future Directions**

This study examined the relationship between magnitude judgement tasks, a measure of inhibition, and arithmetic skills in children. Strengths of the study include the very large sample size and the experimental manipulation of the form of magnitude judgement required. A limitation of the study is the use of a single measure of inhibition. Future studies would benefit from using multiple measures of inhibition (see Gilmore et al., 2015; Nigg, 2000) and more broadly should consider the extent to which a “unitary” inhibition construct can be meaningfully identified and, if it can, the role it may play as a predictor of arithmetic. The use of these multiple measures would potentially allow for the extraction of a latent “inhibition” variable that could capture the variance associated with a general inhibition construct in the absence of task specific variance (e.g., the motor skills component of the HTKS task).

As inhibition is a multidimensional construct, it provides many avenues for further investigation. For example, future research could examine the roles of proactive (effortful goal-orientated) versus reactive (stimulus-driven) forms of inhibitory control in the magnitude-arithmetic relationship. This would assist in determining whether the relation is more specific to stimulus driven suppression than effortful goal oriented control (e.g., Braver, 2012). Furthermore, given the differences in tasks and measurement, the extent to which they correlate with the magnitude-arithmetic relation should also be considered. A wide range of tasks are used to assess inhibition, with most tapping into skills that are additional to response suppression (e.g. motor skills in HTKS task).

A further research direction is to consider the developmental trajectory of the influence of inhibition on this magnitude-arithmetic relationship. As noted previously, inhibition and ANS acuity develop across the lifespan (Halberda & Feigenson, 2008; Hughes, Ensor, Wilson, & Graham, 2009; Williams et al., 1999; Xu, Han, Sabbagh, Wang, Ren, & Li, 2013). Thus, the pattern of relationships found at a specific age and developmental stage may not necessarily reflect the same pattern at other ages and stages. The current study demonstrates the particular pattern of relationships in 6-7 year old children who are in the early stages of formal education. These developmental changes may account for the differences in findings between the current study, and those of Gilmore et al. (2013) and Fuhs and McNeil (2013). Although the children in our study fall within the age range of those in Study 1 of Gilmore et al. (2013; 4.7- to 11.9 years), they are younger than the children in their second study (7.8- to 10.5-years) and older than those in Fuhs and McNeil (3.7- to 5.9-years). Examining any possible developmental changes in this relationship is of particular importance given that in order to support children with arithmetic development it is necessary to have a solid understanding of the predictors of arithmetic at the different stages of childhood.

Finally, it is worth considering the current theoretical debates around the measurement of ANS acuity. A growing body of studies have highlighted salient perceptual features that may influence decisions in the numerosity-discrimination task, including average dot size, cumulative surface area, and convex hull (see Dietrich, Huber, & Nuerk, 2015; Gebius & Reynvoet, 2011). While much research has attempted to control for the influence of these features, it may be near impossible to determine ANS acuity in isolation given the natural correlation between numerosity and perceptual features (e.g., Leibovich et al., 2017). Despite this difficulty, a high

correlation has been observed between average dot size, cumulative surface area and convex hull (Gebuis & Reynvoet, 2012). Accordingly, we focused on controlling for average dot size and cumulative surface area within our sample. Future research examining the role of inhibition in magnitude judgement and arithmetic performance, could also control for further perceptual features (e.g., luminance; De Smedt, Noel, Gilmore, & Ansari, 2013).

In summary, our results provide critical evidence against the idea that inhibition drives the relationship between magnitude discrimination and arithmetic (Gilmore et al., 2013). In line with earlier research examining the role of magnitude understanding in mathematics (e.g., Halberda & Feigenson, 2008), we found that the efficiency of the ANS (as indexed by efficiency on a numerosity discrimination task) accounts for individual differences in concurrent arithmetic ability irrespective of inhibitory control.

## References

- Braver, T. S. (2012). The variable nature of cognitive control: A dual-mechanisms framework. *Trends in Cognitive Sciences*, 16(2), 106–113. doi: 10.1016/j.tics.2011.12.010
- Brigstocke, S., Moll, K. & Hulme, C. (2016) *Test of Basic Arithmetic and Number Skills (TOBANS)*. Oxford: Oxford University Press.
- Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *Journal of Experimental Child Psychology*, 114, 375-388. doi: 10.1016/j.jecp.2012.09.015
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 41, 189-201. doi: 10.1037/0012-1649.41.6.189
- Bugden, S., & Ansari, D. (2015). Probing the nature of deficits in the “Approximate Number System” in children with persistent Developmental Dyscalculia. *Developmental Science*, 1-17. doi: 10.1111/desc.12324
- Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives*, 8(1), 36-41. doi: 10.1111/cdep.12059
- Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children’s mathematics ability: Inhibition, switching and working memory. *Developmental Neuropsychology*, 19(3), 273-293. doi: 10.1207/S15326942DN1903\_3
- Burrage, M.S., Ponitz, C.C., McCreedy, E.A., Shah, P., Sims, B.C., & Jewkes, A.M. (2008). Age- and schooling-related effects on executive functions in young children: A natural experiment. *Child Neuropsychology*, 14, 510–524.

- Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance. *Acta Psychologica*, 148, 163-172. doi: 10.1016/j.actpsy.2014.01.016
- Clayton, S. & Gilmore, C. K. (2015). Inhibition in dot comparison tasks. *ZDM: Mathematics Education*, 47(5), 759-770. doi: 10.1007/s11858-014-0655-2
- Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods*, 19, 300-315. doi: 10.1037/a0033805.
- Davis, H., & Anderson, M. (2001). Individual differences and development - One dimension or two? In M. Anderson (Ed.), *The Development of Intelligence* (pp. 161-191). East Sussex: Psychology Press.
- de Smedt, B., Noel, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, 2(2), 48-55. doi: 10.1016/j.tine.2013.06.001
- de Jong, P. F. (1999) Hierarchical regression analysis in structural equation modeling, *Structural Equation Modeling: A Multidisciplinary Journal*, 6:2, 198-211, DOI: 10.1080/10705519909540128
- Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical Cognition*, 1, 83-120.
- Dietrich, J. F., Huber, S., & Nuerk, H. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS) – a research review. *Frontiers in Psychology*, 6, 295. doi: 10.15496/publikation-8580

- Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, 123, 53-72. doi: 10.1016/j.jecp.2014.01.013
- Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science*, 16(1), 136-148. doi: 10.1111/desc.12013
- Gebius, T & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behaviour Research Methods*, 43, 981-986. doi: 10.3758/s13428-011-0097-5
- Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, 141(4), 642-648. doi: 10.1037/a0026218
- Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS One*, 8(6), 1-9. doi: 10.1371/journal.pone.0067374
- Ginsburg, H. P. & Baroody, A. J. (2003). Test of Early Mathematics Ability (3rd Ed.). California: Western Psychological Services.
- Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, 44(5), 1457-1465. doi: 10.1037/a0012682
- Halberda, J., Marzocco, M.M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 2(455), 665-668. doi: 10.1038/nature07246.

- Howard, S. J., Johnson, J., & Pascual-Leone, J. (2014). Clarifying inhibitory control: Diversity and development of attentional inhibition. *Cognitive Development*, 31, 1-21. doi: 10.1016/j.cogdev.2014.03.001
- Hughes, C., Ensor, R., Wilson, A., & Graham, A. (2009). Tracking executive function across the transition to school: A latent variable approach. *Developmental Neuropsychology*, 35(1), 20-36. doi: 10.1080/87565640903325691
- Hulme, C., Nash, H. Gooch, D., Lervag, A., Snowling, M.J. (2015) The foundations of literacy development in children at risk of dyslexia. *Psychological Science*, 26, 1877–1886. DOI: 10.1177/0956797615603702
- Jonkman, L. M., & Lansbergen, M., & Stauder, J. E. A. (2003). Developmental differences in behavioral and event-related brain responses associated with response preparation and inhibition in a go/nogo task. *Psychophysiology*, 40(5). doi:10.1111/1469-8986.00075
- Keller, L., & Libertus, M. (2015). Inhibitory control may not explain the link between approximation and math abilities in kindergarteners from middle class families. *Frontiers in Psychology*, 6, 685. doi: 10.3389/fpsyg.2015.00685
- Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: A development neuropsychological assessment*. San Antonio, TX: The Psychological Corporation.
- Kroesbergen, E.H., Van Luit, J.E.H., Van Lieshout, E.C.D.M., Van Loosbroek, E., & Van de Rijt, B.A.M. (2009). Individual differences in early numeracy. *Journal of Psychoeducational Assessment*, 27(3), 226-236. doi: 10.1177/0734282908330586



- Lee, H. W., Lo, Y. H., Li K. H., Sung W. S., & Juan C. H. (2015). The relationship between the development of response inhibition and intelligence in preschool children. *Frontiers in Psychology*, 6, 802. doi: 10.3389/fpsyg.2015.00802.
- Leibovich, T., Katzin, N., Harel, M. & Henik, A. (2017). From “sense of number” to “sense of magnitude”: The role of continuous magnitudes in numerical cognition. *Behavioural and Brain Sciences*, 1-62. doi: 10.1017/S0140525X16000960
- Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology*, 116, 829-838. doi: 10.1016/j.jecp.2013.08.003
- McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). Predictors of early growth in academic achievement: The head-toes-knees-shoulders. *Frontiers in Psychology*, 5, 599. doi: 10.3389/fpsyg.2014.00599
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, 215, 1519–1520. doi: 10.1038/2151519a0
- Muthén, L.K., & Muthén, B.O. (1998-2017). Mplus User’s Guide (6<sup>th</sup> Ed.). Los Angeles, CA: Muthén, L.K., & Muthén.
- Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, 126(2), 220. doi: 10.1037/0033-2909.126.2.220
- Nosworthy, N., Budgen, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical

magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE*, 8(7), e67918. doi: 10.1371/journal.pone.0067918

Raven, J. C. (1995). *Manual for the Coloured Progressive Matrices (Revised)*. NFER-Nelson: Windsor.

Robinson, K. M. & Dubé, A. K. (2013). Children's additive concepts: Promoting understanding and the role of inhibition. *Learning and Individual Differences*, 23, 101-107. doi: 10.1016/j.lindif.2012.07.016.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & de Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, 20, 1-16. doi: 10.1111/desc.12372

St Clair-Thompson, H., & Gathercole, S. E. (2006). Executive functions and achievement in school: Sifting, updating, inhibition, and working memory. *The Quarterly Journal of experimental Psychology*, 59, 745-759. doi: 10.1080/17470210500162854

Szucs, D., Devine, A., Soltez, F., Nobes, A., & Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex*, 49, 2674-2688. doi: 10.1016/j.cortex.2013.06.007

Tillman, C. M., Thorell, L. B., Brocki, K. C., & Bohlin, G. (2008). Motor response inhibition and execution in the stop-signal task: development and relation to ADHD behaviors. *Child Neuropsychology*, 14, 42-59. doi: 10.1080/09297040701249020

- Wiebe, S. A., Sheffield, T. D., & Espy, K. A. (2012). Separating the fish from the sharks: a longitudinal study of preschool response inhibition. *Child Development, 83*, 1245–1261. doi: 10.1111/j.1467-8624.2012.01765
- Williams, B. R., Ponesse, J. S., Schacher, R. J., Logan, G. D., & Tannock, R. (1999). Development of inhibitory control across the life span. *Developmental Psychology, 35*(1), 205-213. doi: 10.1037/0012-1649.35.1.205
- Xu, F., Han, Y., Sabbagh, M.A., Wang, T., Ren, X., & Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PLoS ONE 8*(10): e77770. doi: org/10.1371/journal.pone.0077770
- Xu, F. & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1-B11. doi: 10.1016/S0010-0277(99)00066-9

Table 1.

*Descriptive Statistics for Measures of Magnitude Discrimination, Arithmetic and Inhibition*

Task (max)	<i>N</i>	Mean ( <i>SD</i> )	Range	Skewness	Reliability
Dot comparison					
2:3 Congruent (40)	430	19.72 (6.75)	2-40	.17	-
2:3 Incongruent (40)	412	16.82 (5.49)	3-36	.25	-
2:3 Super congruent (40)	436	21.86 (7.22)	1-40	.02	-
2:3 Super incongruent (40)	362	15.86 (5.96)	2-40	.55	-
5:6 Congruent (40)	419	14.82 (5.13)	3-31	.21	-
5:6 Incongruent (40)	365	12.67 (4.33)	2-26	.38	-
5:6 Super congruent (40)	420	16.67 (5.53)	3-40	.29	-
5:6 Super incongruent (40)	283	11.55 (3.94)	2-25	.44	-
Inhibition					
HTKS (40)	459	34.67 (5.51)	5-40	-2.27	.76
Arithmetic					
Addition (60)	474	12.32 (7.42)	0-58	1.89	.92 <sup>1</sup>
Subtraction (60)	469	7.32 (4.33)	0-33	1.34	.88 <sup>1</sup>
Non-verbal cognitive ability					
Ravens CPM (12)	488	8.12 (1.50)	2-11	-.78	.48

*Note.* Unless otherwise specified, all reliabilities are Cronbach's alpha. <sup>1</sup>Test-retest reported in the manual.

Table 2.

*Correlations between Magnitude Discrimination, Inhibition and Arithmetic Tasks*

	1	2	3	4	5	6	7	8	9	10	11	12
1. 2:3 Congruent	-	.60**	.64**	.52**	.50**	.43**	.48**	.41**	.23**	.25**	.36**	.01
2. 2:3 Incongruent	.57**	-	.56**	.56**	.54**	.51**	.48**	.46**	.26**	.35**	.36**	.01
3. 2:3 Super congruent	.64**	.52**	-	.45**	.55**	.53**	.53**	.42**	.10	.36**	.34**	-.04
4. 2:3 Super incongruent	.47**	.56**	.49**	-	.52**	.33**	.34**	.36**	.26**	.33**	.33**	.05
5. 5:6 Congruent	.53**	.52**	.58**	.51**	-	.52**	.45**	.39**	.09	.30**	.28**	.01
6. 5:6 Incongruent	.50**	.52**	.54**	.41**	.54**	-	.40**	.47**	.07	.28**	.26**	.18*
7. 5:6 Super congruent	.52**	.49**	.60**	.40**	.55**	.47**	-	.34**	.16*	.21**	.26**	.03
8. 5:6 Super incongruent	.43**	.44**	.41**	.40**	.39**	.51**	.38**	-	.08	.17**	.23**	.05
9. Inhibition (HTKS)	.30**	.22**	.20**	.14**	.18**	.15**	.24**	.06	-	.26**	.25**	.01
10. Addition	.31**	.37**	.37**	.38**	.34**	.28**	.27**	.18**	.29**	-	.78**	.13*
11. Subtraction	.31**	.28**	.31**	.33**	.29**	.19**	.22**	.18**	.31**	.74**	-	.17*
12 Non-verbal IQ (RCPM)	.14**	.06	.13**	.12*	.14**	.11*	.09	-.00	.14**	.23**	.25**	-
13. Age	.17**	.20**	.18**	.13*	.15**	.13*	.24**	.04	.05	.10*	.03	.17**

Note. \* $p < .05$ ; \*\*  $p < .01$ ; simple correlations below the diagonal; partial correlations controlling for age above the diagonal; HTKS = heads, toes, knees, shoulders; RCPM = Raven's Coloured Progressive Matrices

Table 3.

*Correlations between latent variables*

	1.	2.	3.	4.
1. Magnitude judgment	-			
2. Incongruent magnitude judgment	.00	-		
3. Inhibition	.31**	-.03	-	
4. Arithmetic	.46**	.20	.39**	-

Note. \* $p < .05$ ; \*\*  $p < .01$

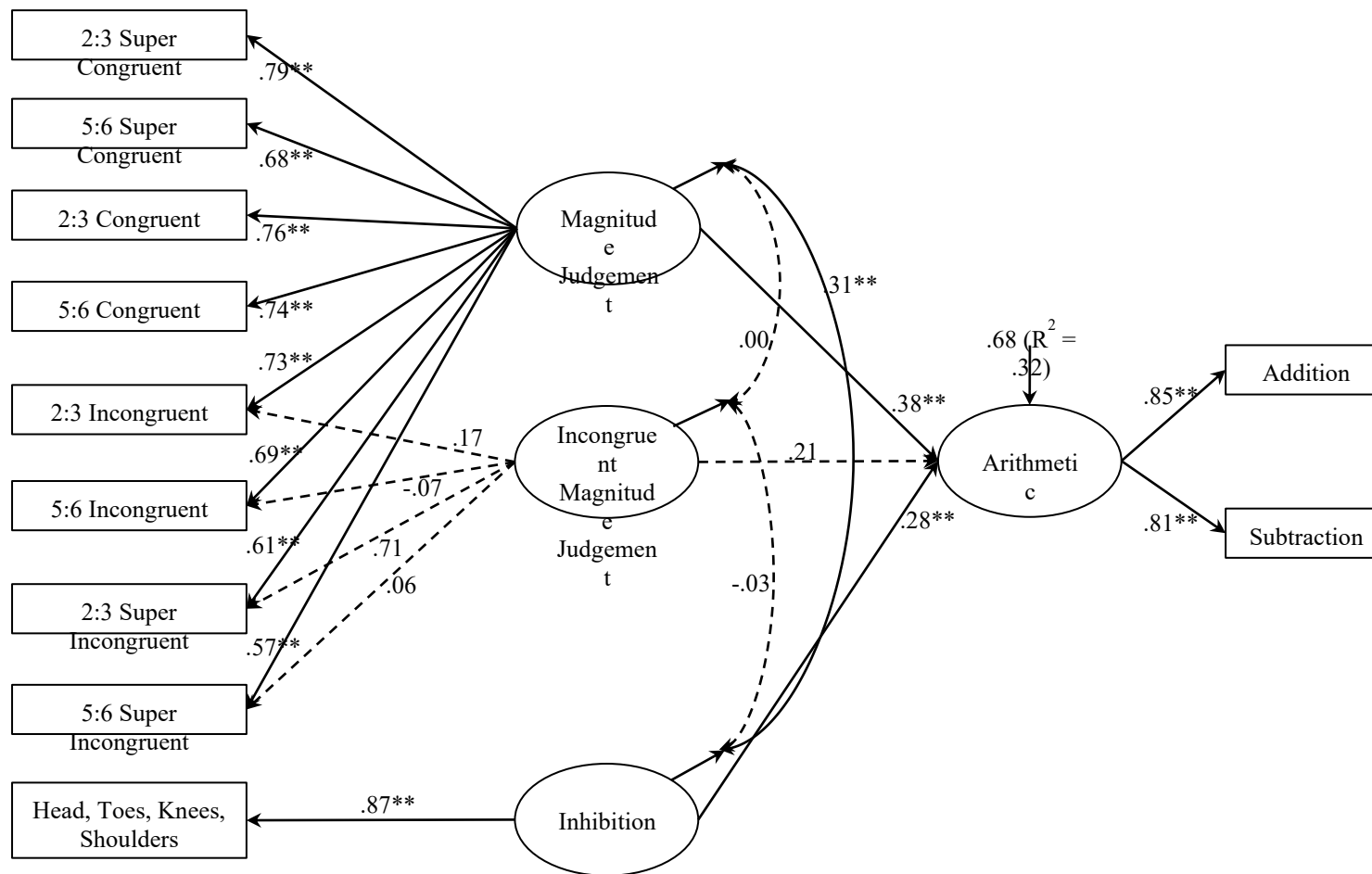


Figure 1. Latent variable path model predicting arithmetic performance in children. Solid paths represent significant regression paths; dotted lines represent nonsignificant paths that are retained in the model. Two headed arrows represent correlations. \*  $p < .05$ , \*\*  $p < .01$