Conditional inference and advanced mathematical study: Further evidence

Matthew Inglis
Mathematics Education Centre, Loughborough University.

Adrian Simpson
School of Education, University of Durham.

Abstract. In this paper we examine the support given for the 'theory of formal discipline' by Inglis and Simpson (2008). This theory, which is widely accepted by mathematicians and curriculum bodies, suggests that the study of advanced mathematics develops general thinking skills and, in particular, conditional reasoning skills. We further examine the idea that the differences between the conditional reasoning behaviour of mathematics and arts undergraduates reported by Inglis and Simpson may be down to different levels of general intelligence in the two groups. The studies reported in this paper call into question this suggestion, but also cast doubt on a straightforward version of the theory of formal discipline itself (at least with respect to university study). The paper concludes by suggesting that either a pre-university formal discipline effect or a filtering effect on 'thinking dispositions' may give a better account for the findings.

Keywords: advanced mathematical thinking, conditional inference, logic, reasoning, theory of formal discipline, thinking dispositions

1. Introduction.

For many years mathematics has held a privileged place in most school curricula: it is typically compulsory for students to study the subject until an older age, and for longer each week, than most other school subjects (Gill, 2003). Historically, one of the primary reasons for this privileged status is the so-called 'theory of formal discipline': the idea that studying mathematics develops thinking skills more generally.

Such views are widely held by modern mathematicians. The distinguished algebraist Shimshon Amitsur was explicit about the link: "through mathematics we also wish to teach logical thinking — no better tool for that has been found so far" (Sfard, 1998, p. 453) and a UK government report claimed that "mathematical training disciplines the mind, develops logical and critical reasoning, and develops analytical and problem-solving skills to a high degree" (Smith, 2004, p. 11). The theory is therefore not only of abstract interest, but also has potentially important policy implications. Stanic (1986) demonstrated that the differing status of mathematics in early twentieth century US curricula was substantially related to changing attitudes towards the theory of formal discipline, and Smith (2004) used the idea to argue that mathematics university students should receive tuition fee rebates, and school mathematics teachers should have targeted salary increases. On the other hand, the privileged place of mathematics in the school curriculum has come under recent attack, including from some sections of the mathematics education community (Bramall and White, 2000). We believe that these policy debates should be driven by empirical evidence.

Inglis and Simpson (2008) compared the inferences drawn from conditional statements by university mathematics students and university students from non-science subjects (used to represent a comparison group of non-mathematical but well-educated people), and found that the mathematics students endorsed fewer invalid inferences than the comparison group, and were less affected by the affirmative premise effect (a reasoning bias which has been shown to interfere with normative logical behaviour; Evans and Handley, 1999). Inglis and Simpson suggested that their findings were consistent with a version of the theory of formal discipline.

Notwithstanding their findings, there are two weaknesses to the study which render the conclusions only suggestive. First, the study did not control for any differences in general intelligence between the two experimental groups: perhaps the difference between the two groups lay in intelligence rather than subject studied (or, indeed, perhaps the students with higher general intelligence are disproportionately filtered into studying mathematics at advanced levels, as suggested by Thorndike (1924)). Since general intelligence measures are related to conditional reasoning measures (Newstead et al., 2004; Stanovich and West, 2000), the differences found by Inglis and Simpson might simply reflect this filtering effect.

Second, the theory of formal discipline implies that it is the study of mathematics which *develops* logical reasoning skills: the cross-sectional design of Inglis and Simpson's study could not address this development directly.

In this paper we report the findings of two studies designed to directly address these issues; thereby providing a stronger test of the hypotheses that underlie the theory of formal discipline.

2. Background.[1]

A typical abstract conditional inference task is shown in Figure 1. Participants are given an abstract conditional sentence, a premise and are asked to decide whether a conclusion either follows or does not follow. Four different conditional statements of the form 'if p then q' are possible by varying the presence of negated components. Researchers have tended to focus on four

[1] In the interests of brevity, this section only briefly revisits some of the theory behind the analysis of conditional inference behaviour, and provides a short summary of earlier findings (for a detailed account see Inglis and Simpson, 2008).

This problem concerns an imaginary letter-number pair. Your task is to decide whether or not the conclusion *necessarily* follows from the rule and the premise.

Rule: If the letter is not T then the number is 6.

Premise: The number is not 6.

Conclusion: The letter is T.

◯ YES (it follows) ◯ NO (no, it does not follow)

Figure 1. A typical conditional inference task (for the rule 'if $\neg p$ then q', the inference MT, and with an explicitly negated premise '$\neg q$').

Table I. The four conditional-types and four inference-types used in the study. These gave rise to sixteen inferences, shown here with their premises (Pr), conclusions (Con), inference-type and validity.

	MP		DA		AC		MT	
Conditional	Pr	Con	Pr	Con	Pr	Con	Pr	Con
if p then q	p	q	$\neg p$	$\neg q$	q	p	$\neg q$	$\neg p$
if p then $\neg q$	p	$\neg q$	$\neg p$	q	$\neg q$	p	q	$\neg p$
if $\neg p$ then q	$\neg p$	q	p	$\neg q$	q	$\neg p$	$\neg q$	p
if $\neg p$ then $\neg q$	$\neg p$	$\neg q$	p	q	$\neg q$	$\neg p$	q	p
Inference-type	affirmative		denial		affirmative		denial	
Validity	valid		invalid		invalid		valid	

different inferences, two which are valid—modus ponens (MP) and modus tollens (MT)—and two of which are invalid—denial of the antecedent (DA) and affirmation of the consequent (AC). The premises and conclusions of these four inferences are shown, with their validity and inference-type[2], for the four different conditional statements, in Table I.

Here we concentrate on two main effects which have been found to interfere with normative conditional reasoning behaviour, the negative conclusion effect and the affirmative premise effect (Evans, 2007). The negative conclusion effect refers to the observation that reasoners typically draw more inferences with negative conclusions than inferences with affirmative conclusions. That is to say that the inference 'if A then 3; ¬3; therefore ¬A'[3] is drawn more often than the inference 'if ¬A then 3; ¬3; therefore A' despite both being valid MT inferences (Evans et al., 1995; Evans and Handley, 1999; Schroyens et al., 2000). This is a robust effect on both denial inferences

[2] The *inference-type* of an inference is defined as either 'affirmative' or 'denial' depending on the valence of the conclusion drawn from the non-negated conditional 'if p then q'.

[3] The symbol '¬' here should be read as "not".

(DA and MT), but is only weakly observed (if at all) on AC, and never on MP (Schroyens et al., 2001).

The affirmative premise effect refers to the finding that participants endorse more inferences from affirmative premises than from negative premises. It is primarily observed when those negative premises are represented implicitly. For example, the inference 'if ¬A then 3; A; therefore ¬3' is made more often than the inference 'if A then 3; R; therefore ¬3', even though they are both instances of drawing, invalidly, the DA inference (Evans and Handley, 1999).

Evans and Handley's (1999) two hurdle account of conditional inference brought together these two effects. They suggested that, in order to answer a conditional inference task correctly, the reasoner must, first, avoid the affirmative premise effect, i.e. they must see that the premise is relevant to the conditional statement (i.e. notice that R or ¬A is relevant for the conditional 'if ¬A then 3'). Second, they must avoid the negative conclusion effect, i.e. be able to convert the statement '$\neg\neg p$' into 'p'. It is only when both stages are hurdled successfully that an inference can be made.

Inglis and Simpson (2008) found three main effects when comparing the conditional inferences drawn by a group of mathematics undergraduates and a groups of undergraduates studying for other degrees:

- overall, the mathematics group made fewer incorrect responses than the comparison group;

- both groups exhibited the negative conclusion effect to approximately the same extent;

- there was a significant between-groups difference with respect to the affirmative premise effect: the comparison group showed the standard effect, but the mathematics group showed no effect.

Inglis and Simpson speculated that this difference may be the result of the mathematics undergraduates being better able to 'see through' opaque representations[4] than the comparison group (cf. Zazkis and Gadowsky, 2001; Zazkis and Liljedahl, 2004). An important remaining question is: what is the cause of this between-groups difference? The theory of formal discipline would suggest that it is caused by the study of advanced mathematics. However, as discussed above, it may be that the two groups differed in general intelligence, and that this difference lay behind the findings. The main aim of our first study was to provide further evidence on this issue.

[4] In the example above 'R' is a more opaque representation of p than '¬A', and so it is harder to see its relevance to the conditional 'if ¬A then 3'.

3. The comparative study.

Thus, the primary goals of the first study reported in this paper were (i) to replicate the findings reported by Inglis and Simpson (2008); and (ii) to determine whether these findings were consequences of different levels of general intelligence between the mathematics and comparison groups.

3.1. Method.

Participants were first year undergraduate students studying at a highly-ranked UK university. Two groups were recruited: a group of 45 students studying for a degree offered by the Department of Mathematics (studying either Mathematics, or a joint degree with a significant mathematics component); and a group of 33 students studying a range of subjects which did not involve a significant formal mathematics component (e.g. English Literature or Chemistry). Both groups of participants had been highly successful in their school career. Of those who had attended school in England or Wales, 73% had achieved Advanced Level grades of AAA or higher.[5] Most of the participants in the mathematics group had studied both Mathematics and Further Mathematics A–levels, and all but one had gained A grades in both.

Participants were given a booklet containing four sections. The first section consisted of Part 1 of the AH5 intelligence test (Heim, 1968). This test, designed for high achieving adults, contained 36 items in the categories 'directions', 'verbal analogies', 'numerical series', and 'similar relationships'. The AH5 test has been widely used by earlier researchers, including by Newstead et al. (2004) as part of their study of individual differences in reasoning behaviour. Sections 2-4 of the booklet consisted of different reasoning tasks, and were presented to participants in a counterbalanced order. In this paper we report data from one of these sections: that which focussed on abstract conditional inferences.

The conditional inference section of the instrument was identical to that used by Inglis and Simpson (2008): it consisted of 32 reasoning problems of the type shown in Figure 1.[6] The inferences used are shown in Table I; half of the problems used explicitly negated premises (e.g. $\neg 4$ was represented as "not 4") and half used implicitly negated premises (e.g. $\neg 4$ was represented as, for example, "8"). The lexical content of the rules were generated randomly and the order of the problems was randomised for each participant.

[5] The Advanced Level (A–level) is the qualification taken by 18 year old school-leavers in England and Wales. It is marked on a seven point scale from A (highest) to G (lowest), and each candidate typically studies three (or sometimes four) subjects. In 2007 around 3% of 18 year olds across England and Wales achieved AAA or higher.

[6] An explanation of why this task is particularly appropriate for investigating conditional inference behaviour is given by Inglis and Simpson (2008).

Table II. The mean number of each inference correctly categorised by the mathematics and comparison groups.

	Explicit negations				Implicit negations				Overall
	MP	DA	AC	MT	MP	DA	AC	MT	Total
Mathematics	3.97	2.93	2.07	2.41	3.76	3.10	2.38	2.48	23.10
Comparison	4.00	2.38	1.65	2.23	3.58	2.65	2.12	1.96	20.58

The instrument was preceded by the same instructions used by Inglis and Simpson.

The task was administered to participants near the beginning of the university year. No participant had yet attended more than one or two lectures or tutorials. Participants spent twenty minutes on the AH5 test, after which they were told to move on to the remainder of the task and complete it at their own pace without returning to a problem after having moved on to another one.

3.2. Results.

Participants' AH5 scores were calculated, yielding a score out of 36. As expected, participants' AH5 scores were positively correlated with the overall number of inferences they correctly categorised, $r = .485, p < .001$. Furthermore, the mathematics sample had a higher mean AH5 score than the comparison sample, 19.1 vs 14.7, $t(76) = 4.05, p < .001$. To eradicate the influence of general intelligence on the results the samples' scores were balanced by removing the highest scoring 16 participants from the mathematics sample and the lowest scoring 7 participants from the comparison sample from the analysis. This gave a mathematics group of 29 participants (mean AH5 score 16.3; Std Dev 3.5) and a comparison group of 26 participants (mean AH5 score 16.4; Std Dev 2.8).

Table II shows the mean number of correct inferences categorised by both groups, for both explicitly- and implicitly-negated problems. Overall, the mathematics group correctly categorised a higher number of inferences, 23.1, than the comparison group, 20.6, $t(53) = 2.08, p = .043$.

The main analysis followed that conducted by Inglis and Simpson (2008). To interrogate the influence of the negative conclusion effect a *Negative Conclusion Index* (NCI) was calculated, defined as the number of inferences endorsed on arguments with affirmative conclusions subtracted from the number of inferences endorsed on arguments with negative conclusions. As the number of each type of both valid and invalid inferences with negative and affirmative conclusions was equal, we would expect a participant exhibiting no negative conclusion effect to attain an overall NCI of zero, and a participant exhibiting the effect to show a positive NCI.

Participants' NCIs were subjected to an analysis of variance (ANOVA) with two within-participant factors (inference-type and negation-type), and one between-participants factor (group). The inference-type factor was obtained by collapsing the four inferences into two categories: affirmative (MP and AC) and denial (DA and MT). The negation-type factor referred to whether the inference involved explicitly negated premises or implicitly negated premises. Thus the NCIs ranged from -4 to 4.

There was a significant effect for inference-type, $F(1,53) = 16.4$, $p < .001$, with a mean NCI for denial inferences of 2.35 compared to 0.84 for affirmative inferences. The mean NCIs for each group, for each inference-type are shown on the left of Figure 2. There was no significant effect for negation-type, nor were there significant inference-type $\times$ group, negation-type $\times$ group or inference-type $\times$ negation-type $\times$ group interaction effects, all $Fs < 1.5$.

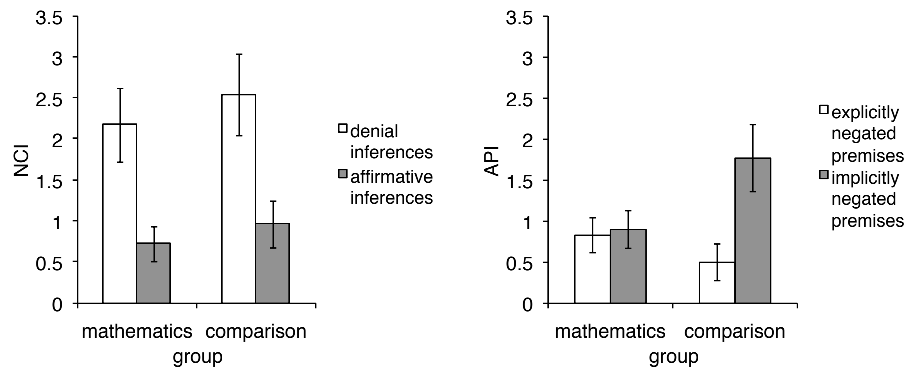

Figure 2. Left: The mean NCIs for each group, by inference-type. Right: The mean APIs for each group, by negation-type. Error bars represent ± 1 SE of the mean.

The affirmative premise effect was analysed by calculating an *Affirmative Premise Index* (API), defined as the number of inferences endorsed on arguments with negative premises subtracted from the number of inferences endorsed on arguments with affirmative premises. As with the NCIs, we would expect a participant exhibiting no affirmative premise effect to attain an overall API of zero, and a participant exhibiting the effect to show a positive API.

Participants' APIs were subjected to an analysis of variance (ANOVA) with two within-participant factors (validity and negation-type), and one between-participants factor (group). The validity factor was obtained by collapsing the four inferences into two categories: valid (MP and MT) and invalid (DA and AC). The negation-type factor referred to whether the inference involved explicitly negated premises or implicitly negated premises. Thus, like the NCIs, the APIs ranged from -4 to 4.

As predicted by the literature there were significant effects for validity, $F(1,53) = 7.92$, $p = .007$, and negation-type, $F(1,53) = 6.61$, $p = .013$. The mean API was 0.62 for valid inferences compared to 1.36 for invalid inferences, and 0.67 for explicit-negations compared to 1.31 for implicit negations. There was no validity $\times$ group interaction, $F < 1.5$, but the negation-type $\times$ group interaction did reach significance, $F(1,53) = 5.31, p = .025$. This interaction is shown on the right of Figure 2. The comparison group showed a large difference in affirmative premise effects between the two negation-types, with a difference between the mean APIs of implicitly and explicitly negated premises of 1.27, $t(25) = 2.85, p = .009$. In contrast the difference for the mathematics group was only a non-significant 0.07. No significant validity $\times$ negation-type or validity $\times$ negation-type $\times$ group interaction effects were found, $Fs < 1$.

3.3. DISCUSSION.

In this study we constructed groups balanced by AH5 scores, and used the same instrument as Inglis and Simpson (2008) to interrogate conditional inference behaviour. The findings of this study essentially replicated the earlier study. First, overall the mathematics group had more correct responses compared to the comparison group. Second, both groups exhibited the negative conclusion effect to approximately the same degree. And third, the comparison group showed the standard affirmative premise effect, but the mathematics group showed no effect.

Most importantly, the study demonstrates that the differences between the mathematics and non-mathematics students detected by Inglis and Simpson (2008) can not be accounted for by suggesting the two groups had differing levels of general intelligence. However, while this finding is consistent with the theory of formal discipline, on its own it does not provide overwhelming support for it. The cross-sectional nature of the study means it does not provide compelling evidence that the study of mathematics *develops* logical reasoning, the potential for a filtering effect (on a factor unrelated to general intelligence) remains.

4. The longitudinal study.

To investigate the relationship further, we conducted a longitudinal study which directly sought to address the developmental claim of the theory of formal discipline. We asked the first year mathematics undergraduates from the comparison study to return at the end of their first year of studies to

Table III. The mean number of each inference correctly categorised by the mathematics group in Sessions 1 and 2.

	MP	DA	AC	MT	Total
Session 1	7.82	6.52	5.39	5.06	24.79
Session 2	7.91	6.55	5.61	5.27	25.33

participate in a follow-up. If studying first year undergraduate mathematics did develop logical reasoning skills, we would expect a different profile of scores on the second administration.

4.1. Method.

All the participants from the mathematics sample in the comparison study were invited to take part in a follow-up study at the end of their first year of university mathematics studies. A total of 33 (of the 45) participants agreed to take part. The follow-up instrument was administered in the same way as, and was identical to, the first, except that the AH5 test was replaced by a separate task designed to take approximately the same time.

Between the two studies participants had studied six modules towards their mathematics degree, of which four were compulsory (with the other two coming from a wide range of options). The content of the core modules included some calculus, geometry, algebra, probability, analysis and dynamics, as well as a module on problem solving designed to encourage 'mathematical thinking' (partly based on Mason et al., 1982). The options included discrete mathematics, data analysis, modelling and simulation. Although all the modules integrated proof techniques into their syllabuses, only the discrete mathematics module did so explicitly (with a formal description of proof by induction and the pigeonhole principle). Note that no module contained any explicit teaching of formal logic in the forms of propositional or predicate calculus. Despite this, each module included "the development of abstract reasoning" as a "non-specific" goal in its syllabus.

4.2. Results.

The mean number of each inference correctly categorised by participants in each session is shown in Table III. Although there was a slight improvement in scores between the two sessions (of 1.8%), this difference was small and did not approach significance, $t(32) = 1.16, p > .25$.

Participants' NCIs were calculated for each of the two sessions and analysed using an ANOVA with three within-participant factors (session, negation-type and inference-type; defined in the same way as in the analysis of the comparative study). There was a significant effect for inference-type, $F(1,32) = 19.1$, $p < .001$, with a mean NCI for denial inferences of 1.09 compared to 0.29 for affirmative inferences. The mean NCIs for each session, for each inference-type are shown on the left hand side of Figure 3. There was no significant effect for session or negation-type, nor were there any significant two- or three-way interaction effects, all $Fs < 1$.

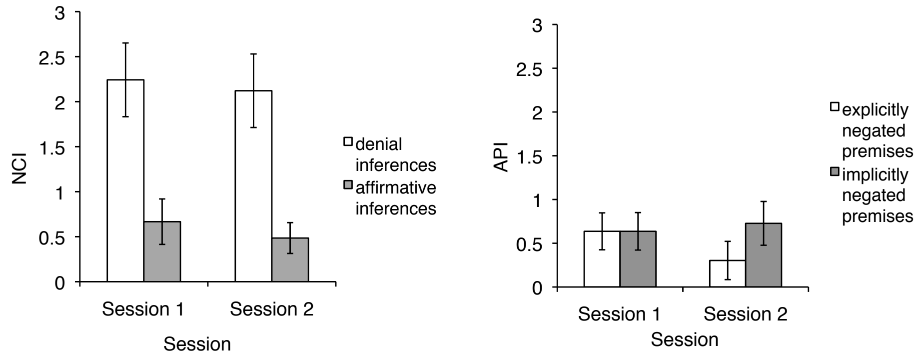

Figure 3. Left: The mean NCIs from each session, by inference-type. Right: The mean APIs from each session, by negation-type. Error bars represent ±1 SE of the mean.

Participants' APIs were calculated for the two sessions and analysed using an ANOVA with three within-participant factors (session, negation-type and validity; defined in the same way as in the analysis of the comparative study). The main effect of validity was borderline significant, $F(1,32) = 3.71$, $p = .063$, but no other main effect, two-way, or three-way interaction effect approached significance, ps $> .2$. The mean APIs for each session, for each negation-type are shown on the right hand side of Figure 3.

4.3. Discussion.

The primary aim of the longitudinal study was to determine whether studying first year undergraduate mathematics leads to a change in abstract conditional inference behaviour. Such a change would provide compelling evidence for the theory of formal discipline. However, no conditional inference index used in this study showed a significant improvement between the first session at the start of the year and the second session at the end. Students were not significantly more accurate at categorising inferences at the end of their first year, nor they did not show a reduced negative conclusion effect. In neither session did the mathematics students show an affirmative premise effect.

5. General discussion.

5.1. Summary of main findings.

The theory of formal discipline suggests that studying mathematics at advanced levels develops skills of logical thinking and, in particular, conditional inference. In an earlier study, Inglis and Simpson (2008) demonstrated that university mathematics students do seem to have a different profile of responses to a series of conditional inference tasks compared to arts undergraduates. However, it was unclear whether these differences were the result of differing levels of intelligence between the groups, or whether the mathematics students had—as suggested by the theory of formal discipline—been influenced by their study of advanced mathematics.

In this paper we have extended the findings of Inglis and Simpson (2008) in two significant ways. First, we demonstrated that the differences they found *cannot* be accounted for by differing levels of general intelligence between the groups. In the comparison study reported in this paper the groups were balanced according to their scores on a widely used intelligence test, and all the main findings of the earlier study were replicated. Second, we demonstrated that the profile of responses to conditional inference tasks of mathematics undergraduates did not change over the course of their first year of studies. An essentially identical range of responses were given by the students in the first week of their first year studies to those in the last week of their first year studies, even though they had all studied a full programme of proof-based mathematics modules in the intervening period; as well as a problem solving module designed to develop 'mathematical thinking'.

There would appear to be two reasonable hypotheses that can account for these data. One possibility is that the differences identified between the mathematics and comparison group in this study *are* developmental, but they are the result of differences in study patterns which occur in pre-university education, not at university. A second possibility is that the differences between the groups are not the direct result of instruction, and nor are they the result of differences in intelligence, but rather are caused by between-group differences on a separate unrelated factor. The most obvious candidate for such a factor is an individual's thinking disposition (sometimes referred to as an individual's cognitive style) which could act as a filter into (or out of) mathematical study.

5.2. Pre-university instruction

One obvious way of accounting for the data reported in this paper would be to suggest that studying pre-university mathematics develops logical reasoning skills, but studying university mathematics does not. This would imply that there is a limit to such development and that the limit is achieved between the end of compulsory mathematics and the start of university mathematics. In the UK, where the large majority of our participants had attended school, mathematics is a compulsory subject until the age of 16. At this point those students who continue in education typically opt to study three or four subjects (for the A Level qualification) until the age of 18. One of these subjects can be mathematics and sometimes a second ("further mathematics") is taken. Perhaps it is this period of education which, following the theory of formal discipline, caused the between-groups differences detected in the comparison study.

To investigate this possibility further we conducted an informal analysis of the comparison group who participated in the comparison study. Of course, many of the students in the comparison group—who were studying subjects at university with no significant mathematics component—had nevertheless studied A–level mathematics prior to university (12 of the 33 fell into this category). Those who had studied A–evel mathematics had a mean number of inferences correctly classified of 21.1, compared to a figure of 20.1 for those who had not studied A-level mathematics (with relatively high standard deviations of 5.0 and 3.8 respectively). This difference did not approach significance, $t(31) = 0.591, p > .5$. Although the small sample involved in this comparison requires caution in interpreting this result, it certainly does not provide any support to the hypothesis that the differences detected were the result of pre-university but post-compulsory mathematical study.

5.3. Differences in 'thinking dispositions'

In general, an individual's ability to successfully perform a task will depend on a number of factors: their background knowledge, their concentration, and so on. However, sometimes a task is simply beyond the cognitive capacity of an individual to perform: for example, the solution may require an algorithm too complex to complete (such reasons for task failure are distinct from one-off mistakes or performance errors). Individual differences in general intelligence scores can be taken to reflect individual differences in cognitive capacity (e.g. Baron, 1988); and it has been shown that an individual's cognitive capacity, measured in this way, tends to correlate with their performance on a variety of reasoning tasks (e.g. Newstead et al., 2004; Stanovich and West, 2000). However, in such studies cognitive capacity does not account for all the variance in individual performances.

Stanovich (1999) pointed out that if an individual was operating close to the limit of their cognitive capacity when tackling a reasoning task we should expect a near perfect correlation between reasoning performance and general intelligence. In fact the correlations found by researchers fall a long way short of this ($r = .485$ in the current study). Stanovich attributed this, at least in part, to individual differences in how individuals choose to allocate their cognitive resources: i.e. their thinking disposition (also variously called cognitive style, or learning style). Various measures have been proposed to measure differences in thinking disposition, leading to a diffuse and poorly integrated literature (Coffield et al., 2004). Nevertheless, there are some important findings relevant to the current discussion.

First, Stanovich and West (1998) found that an individual's score on a composite thinking dispositions inventory (designed to measure open-mindedness, willingness to decontextualise and the tendency to consider alternative opinions) was positively correlated with performance on a variety of reasoning tasks (including some focused on conditional inference). Second, there is some evidence that an individual's thinking disposition is related to their performance in mathematics courses. Riding and Agrell (1997) found a relationship between an individual's profile on Riding's (1991) 'Cognitive Styles Analysis' instrument and their performance on mathematics examinations at age 14-16; suggesting that access to post-compulsory mathematical study (which is based, in part, upon school level examination performance) could be correlated with thinking dispositions. Similarly, Drysdale et al. (2001) found that an individual's classification on the Gregorc (1979) Style Delineator predicted student performance on a first year undergraduate course on linear methods (which included sections on matrices, vectors, and determinants). Perhaps surprisingly, students classified as being "ordered, perfection-oriented, practical and thorough" significantly outperformed other groups, including those in a category described as "logical, analytical, rational and evaluative" (cf. Moutsios-Rentzos and Simpson, 2005).

These findings suggest that it is plausible that the differences found between the groups in the comparison study are related to differences in thinking disposition rather than differences in general intelligence. If, for example, those in the mathematics group were more likely to engage in and enjoy effortful analytic activity (i.e. be more willing to allocate their cognitive capacity to solving abstract conditional inference tasks; cf. Cacioppo et al., 1996), this could plausibly cause the between-groups differences in the presence of the affirmative premise effect (recall that avoiding this effect requires the effortful realisation that, for example, the premise R is relevant to the conditional 'if $\neg$A, then 3'). In addition, one would presume (if Evans and Handley's (1999) account were correct) that the negative conclusion effect would be more influenced by cognitive capacity than thinking disposition (avoiding it requires the cognitively demanding task of holding in mind a series of deductive steps, and then converting the statement '$\neg\neg p$' into 'p'), and thus—as we found—one might not expect any between-group differences.

However, it is worth emphasising that this account, that the between-groups differences found in the comparison study are the result of group differences in thinking disposition, while being inconsistent with a straightforward version of the theory of formal discipline, need not necessarily be inconsistent with the theory entirely. Although an individual's thinking disposition may influence their success in mathematical study, the reverse may also be the case: i.e., that studying mathematics influences a persons' thinking disposition, which in turn influences their logical reasoning skills. Indeed Kolb (1984) suggested that "people choose fields that are consistent with their learning styles and are further shaped to fit the learning norms of their field once they are in it" (p. 88); i.e. that thinking dispositions and study choices can reinforce each other.

6. Concluding remarks.

Inglis and Simpson (2008) found that mathematics and other undergraduates behaved differently when asked to judge the validity of abstract conditional inferences. In this paper we have demonstrated that this difference cannot be accounted for by differing levels of general intelligence between the groups. However, we have also shown that mathematics undergraduates do not show any development on any conditional inference measures over the course of their first year university studies.

At first glance these results appear contradictory. The first is consistent with the theory of formal discipline, the second appears not to be. However, we have outlined two possible hypotheses which account for these findings. First, that any developmental gains due to non-compulsory mathematical study take place in pre-university courses, rather than at university. Second, that the between-group differences found in these studies are the result of group differences in intelligence-independent thinking dispositions. This latter hypothesis is potentially consistent with both the theory of formal discipline (where initial minor differences in thinking dispositions are reinforced through mathematical study) and its negation (where initial large differences in thinking dispositions cause some to filter into studying mathematics, and some not to). If empirical mathematics education research is to usefully contribute to the policy debates discussed earlier, then further studies will need to distinguish between these two hypotheses.

References

Baron, J.: 1988, *Thinking and deciding*. Cambridge: Cambridge University Press.

Bramall, S. and J. White (eds.): 2000, *Why learn maths?* London: Institute of Education, University of London.

Cacioppo, J. T., R. E. Petty, J. A. Feinstein, and W. B. G. Jarvis: 1996, 'Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition'. *Psychological Bulletin* **119**, 197–253.

Coffield, F., D. Moseley, E. Hall, and K. Ecclestone: 2004, *Learning styles and pedagogy in post-16 learning: A systematic and critical review*. London: Learning & Skills Research Centre.

Drysdale, M. T. B., J. L. Ross, and R. A. Schulz: 2001, 'Cognitive learning styles and academic performance in 19 first-year university courses: Successful students versus students at risk'. *Journal of Education for Students Placed at Risk* **6**, 271–289.

Evans, J. St. B. T.: 2007, *Hypothetical Thinking: Dual Processes in Reasoning and Judgement*. Hove, UK: Psychology Press.

Evans, J. St. B. T., J. Clibbens, and B. Rood: 1995, 'Bias in conditional inference: Implications for mental models and mental logic'. *Quarterly Journal of Experimental Psychology* **48A**(3), 644–670.

Evans, J. St. B. T. and S. J. Handley: 1999, 'The Role of Negation in Conditional Inference'. *Quarterly Journal of Experimental Psychology* **52A**, 739–769.

Gill, P.: 2003, 'Mathematics'. In: J. White (ed.): *Rethinking the School Curriculum*. London: Routledge, pp. 104–116.

Gregorc, A. F.: 1979, 'Learning/teaching styles: potent forces behind them'. *Educational Leadership* **36**, 234–237.

Heim, A. W.: 1968, *AH5 group test of intelligence*. London: National Foundation for Educational Research.

Inglis, M. and A. Simpson: 2008, 'Conditional inference and advanced mathematical study'. *Educational Studies in Mathematics* **67**, 187–204.

Kolb, D. A.: 1984, *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.

Mason, J., L. Burton, and K. Stacey: 1982, *Thinking Mathematically*. London: Addison-Wesley.

Moutsios-Rentzos, A. and A. Simpson: 2005, 'The transition to postgraduate study in mathematics: A thinking styles perspective'. In: H. L. Chick and J. L. Vincent (eds.): *Proceedings of the 29th International Conference on the Psychology of Mathematics Education*, Vol. 3. Melbourne, Australia, pp. 329–336, IGPME.

Newstead, S. E., S. J. Handley, C. Harley, H. Wright, and D. Farrelly: 2004, 'Individual differences in deductive reasoning'. *Quarterly Journal of Experimental Psychology* **57A**, 33–60.

Riding, R.: 1991, *Cognitive styles analysis*. Birmingham: Learning and Training Technology.

Riding, R. and T. Agrell: 1997, 'The effect of cognitive style and cognitive skills on school subject performance'. *Educational Studies* **23**, 311–323.

Schroyens, W., W. Schaeken, and G. d'Ydewalle: 2001, 'The processing on negations in conditional reasoning: A meta-analytical case study in mental model and/or mental logic theory'. *Thinking and Reasoning* **7**, 121–172.

Schroyens, W., W. Schaeken, W. Fias, and G. d'Ydewalle: 2000, 'Heuristic and Analytic Processes in Propositional Reasoning With Negatives'. *Journal of Experimental Psychology: Learning, Memory and Cognition* **26**, 1713–1734.

Sfard, A.: 1998, 'A mathematician's view of research in mathematics education: An interview with Shimson A. Amitsur'. In: A. Sierpinska and J. Kilpatrick (eds.): *Mathematics education as a research domain: A search for identity*, Vol. 2. Dordrecht: Kluwer, pp. 445–458.

Smith, A.: 2004, *Making Mathematics Count: The report of Professor Adrian Smith's Inquiry into Post-14 Mathematics Education*. London: The Stationery Office.

Stanic, G. M. A.: 1986, 'The Growing Crisis in Mathematics Education in the Early Twentieth Century'. *Journal for Research in Mathematlcs Education* **17**, 190–205.

Stanovich, K. E.: 1999, *Who Is Rational? Studies of Individual Differences in Reasoning*. Mahwah, NJ: Lawrence Erlbaum.

Stanovich, K. E. and R. F. West: 1998, ‘Individual differences in rational thought’. *Journal of Experimental Psychology: General* **127**, 161–188.

Stanovich, K. E. and R. F. West: 2000, ‘Individual differences in reasoning: Implications for the rationality debate?’. *Behavioural and Brain Sciences* **23**, 645–726.

Thorndike, E. L.: 1924, ‘Mental discipline in high school studies’. *Journal of Educational Psychology* **15**, 1–22.

Zazkis, R. and K. Gadowsky: 2001, ‘Attending to transparent features of opaque representations of natural numbers’. In: A. Cuoco (ed.): *The roles of representation in school mathematics*. Reston, VA: NCTM, pp. 146–165.

Zazkis, R. and P. Liljedahl: 2004, ‘Understanding primes: The role of representation’. *Journal for Research in Mathematics Education* **35**, 164–186.