

THE ROLE OF MATHEMATICAL CONTEXT IN EVALUATING CONDITIONAL STATEMENTS

Matthew Inglis

University of Warwick

Adrian Simpson

University of Durham

Recently there has been increasing interest in the mathematics education research community about the role of logic in the teaching, learning and production of mathematics. In this paper we investigate how conditional statements are evaluated by successful mathematics students, and argue that the role of context is vital to determine the manner in which this evaluation proceeds. We use two versions of the so-called Labyrinth Task, one in it's original context and one in an overtly mathematical context. We report results that indicates that the manner in which conditional statements are evaluated on these tasks differs depending on the context. These results are supplemented by data from a qualitative task-based interview study.

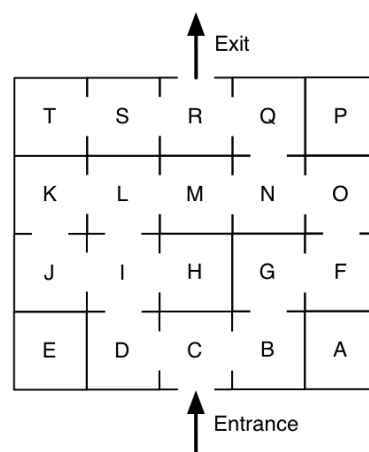
Logical implication is seen as being one of the most important structures in mathematics, and researchers have argued that coming to terms with it is vital for developing an understanding of proof (Durand-Guerrier, 2003; Weber & Alcock, 2005). Our goal in this paper is to describe a psychological framework that explains the processes involved in evaluating conditional statements in everyday language, and to explain how these processes differ in mathematical contexts. To do this we first describe a task used by several researchers and teachers to investigate the role of logic in mathematical reasoning.

THE LABYRINTH TASK

Durand-Guerrier (2003) introduced the so-called Labyrinth Task into the mathematics education literature. In this task participants are presented with a maze, and told that a person *X* managed to pass through it without using the same door twice.

They are then asked to categorise a series of statements as being either *true*, *false*, or that *there is not enough information to tell* (can't tell):

1. *X* crossed *P*.
2. *X* crossed *N*.
3. *X* crossed *M*.
4. If *X* crossed *O*, then *X* crossed *F*.
5. If *X* crossed *K*, then *X* crossed *L*.
6. If *X* crossed *L*, then *X* crossed *K*.



The answers to the first 5 statements appear to be relatively straightforward, but for statement 6 the answer becomes less clear. Durand-Guerrier (2003) argued that the correct answer is “can’t tell” because it is impossible to know whether X passed through K or I before s/he passed through L . When administered to 15-16 year olds apparently this was the answer given by 60% of students, “especially those deemed good at mathematics” (p.9). However, the students’ teachers apparently disagreed:

Surprisingly, some teachers considered this answer to be wrong! (p.8)

The teachers believed that the correct answer was “false”, since according to Durand-Guerrier’s analysis, they had interpreted the statement as “for all X , if X crossed L , then X crossed K ”. Suggesting that “can’t tell” is the “natural” answer for students, Durand-Guerrier worried that the teacher’s interpretation of the statement causes a didactical obstacle for students:

It is necessary to overcome the opinion that every implication met in the classroom is a relation between propositions which is either true or false and that carries necessity. Indeed, *implication between propositions carries no necessity, but is a set of possible cases for truth values*. (Our emphasis, p.29).

The idea that implication is a set of possible truth values may be *logically* correct, but there is a multitude of research that suggests that it is not *psychologically* correct. In the next section we briefly discuss some of this work: Evans & Over’s (2004) theory of conditionals based on the so-called Ramsey Test.

THE RAMSEY TEST – CONDITIONALS IN EVERYDAY LANGUAGE

According to Ramsey (1931), when people judge the truth/falsity of a conditional in natural language they are “hypothetically adding P to their stock of knowledge and arguing on that basis about Q ”, they are in effect “fixing their degrees of belief in Q given P ” (p.247). This idea – that to judge $\mathbf{P}(P \Rightarrow Q)$ a person judges $\mathbf{P}(Q|P)$ rather than $\mathbf{P}(Q \text{ or not-}P)$ – has become known as the Ramsey Test. (Here $\mathbf{P}(X)$ indicates the level of belief that a person has in event X . This is clearly related to, but not necessarily identical to, the probability of event X .)

The notion of the Ramsey Test is a non-trivial model of the manner in which people judge conditional statements. Such a model is at odds with both formal logic and other psychological accounts of conditional statements, specifically Johnson-Laird & Byrne’s (2002) influential mental models framework (for a full discussion of the difference between these theories see Evans, Over & Handley, 2005).

To illustrate how the Ramsey Test operates, consider the statement “if you’re in Birmingham, then you have a good choice of Indian takeaways”. This statement is judged by hypothetically supposing that you are in Birmingham, and then considering the availability of Indian food, given this supposition and your existing knowledge and beliefs. Note that this process is both psychologically and logically different to the truth evaluation of formal material conditionals. A material conditional $P \Rightarrow Q$ is true whenever P or not- Q is true. Thus if you are not in Birmingham and the statement is evaluated as a material conditional, then it is automatically true. But

evaluated as a suppositional conditional using a Ramsey Test it may be true or false depending on the individual involved's beliefs. There is an increasing body of evidence that supports the notion that the Ramsey Test is an accurate model for how humans judge conditional statements (e.g. Evans & Over, 2004; Hadjichristidis et al., 2001).

How then, does the Ramsey Test apply to the Labyrinth Task? The participant hypothetically adds “ X crossed L ” to their stock of beliefs and then evaluates their degree of belief in “ X crossed K ”. Given the layout of the maze it is clear that

$$P(X \text{ crossed } K \mid X \text{ crossed } L) = 0.5$$

so Evans & Over's (2004) theory of suppositional conditionals would predict that most people might categorise “if “ X crossed L , then X crossed K ” as “can't tell”, as they neither have strong belief nor disbelief in the statement.

However, as we have seen Durand-Guerrier (2003) reports that the mathematics teachers who administered the task believed that the correct answer was “false”. It seems clear that they were evaluating the statement in a somewhat differently to the manner which Evans & Over's (2004) theory predicts. The purpose of this paper is to investigate the role that mathematical context plays in the evaluation of conditional statements with the Ramsey Test.

METHOD

We were interested in discovering exactly how successful mathematicians evaluate the Labyrinth Task, and whether mathematical context plays a part in this. To this end we administered two versions of the task. The first version was identical to the original task reported above, and the second was phrased in an overtly mathematical context:

Your friend X is interested in a sequence of real numbers, (a_n) . X writes down sentences about the sequence. For each of the sentences you must decide whether it is *true*, *false*, or whether there is *not enough information* to tell.

Place each of the following statements into the categories: *true*, *false* or *can't tell*.

1. $a_k=4$ for some $k \in \mathbf{N}$.
2. $a_{46} \in \mathbf{R}$.
3. $a_n \rightarrow \frac{3}{4}$.
4. If $a_{n+1} > a_n + 1$ (for all n), then $a_n \rightarrow \infty$.
5. If $\sum a_n$ converges, then $a_n \rightarrow 0$.
6. If $a_n \rightarrow 0$, then $\sum a_n$ converges.

Parts 4,5 and 6 of this task were designed to be isomorphic to the maze task, but set in a mathematical context. Thus presumably Durand-Guerrier (2003) would argue that the correct answer to part 6 is “can't tell” as there are some sequences (a_n) for which this statement is true (from a formal logic standpoint), but some sequences for

which it is false. For example the sequence $a_n = n^{-2}$ both tends to zero and its associated series converges. Thus, since both P and Q are true, “if P , then Q ” formally is also true. However the sequence $a_n = n^{-1}$ tends to zero but its associated series does not converge, thus “if P , then Q ” is false. So (it could be argued) there is not enough information to tell whether statement 6 is true or false.

The participants were 433 first year mathematics undergraduates from the first author’s institution. All the students in our sample had been highly successful school level mathematicians, and had typically achieved A Level grades of AAB or higher. The cohort was randomly split into two equal groups and each group were given either the original or the mathematical version of the task. The task was administered as part of a biweekly test that formed a minor part of the assessmentⁱ of a first year Foundations of Mathematics course (which included sections on logic and implication). All the students were simultaneously taking a course in Analysis and so should have been familiar with the terms used in the mathematical version of the question.

RESULTS

In this section we report the results of part 6 of each version of the labyrinth task. These figures are shown in Table 1. It can clearly be seen that the range of responses was different between the versions. In the original version participants were fairly evenly split between the ‘false’ (44%) and ‘can’t tell’ (54%) responses, with almost no one selecting ‘true’ (2%). However, in the mathematical version a large minority of participants selected ‘true’ (30%), and few selected ‘can’t tell’ (14%).

Answer	Original	Maths	Total
T	2	30	16
F	44	56	50
C	54	14	34
n	216	214	430

Table 1: The breakdown of responses, as percentages, to statement 6.

T – true, F – false, C – can’t tell.

The difference between the responses (by test version) reported in Table 1 is highly significant, with a large effect size, $\chi^2=107$, $df=2$, $p<0.001$, $\phi=0.498$.

However, we were concerned that some of the differences between the test versions could be attributed to poor subject knowledge in the mathematics version. For example, it is hard to see how any structural property of conditional statements could lead participants to judge part 6 of the mathematical version to be true. To try to mitigate this distorting effect we removed all participants from our analysis who answered part 5 incorrectly. That is to say that we were only interested in the responses (to part 6) of participants who were sufficiently aware of the properties of sequences and series to answer part 5 of the mathematical version correctly

(although, for consistency, we also removed the 7 participants who incorrectly answered part 5 of the original version). After removing these participants from the analysis, the percentage of ‘true’ answers to the mathematical version was reduced from 30% to 9%. These data are shown in Table 2.

Answer	Original	Maths	Total
T	2	9	5
F	44	73	56
C	54	18	39
<i>n</i>	209	146	355

Table 2: The breakdown of responses, as percentages, of those participants who answered part 5 correctly, to statement 6.

The figures in Table 2 are clear. The mathematical version of the task elicited many more “false” responses than did the original version. This difference is highly significant, with a moderate effect size, $\chi^2=48.9$, $df=2$, $p<0.001$, $\phi=0.371$.

DISCUSSION

The results from this study are interesting for several reasons. Recall that Durand-Guerrier (2003) reported that approximately 60% of 15-16 year old students responded with “can’t tell” to statement 6 of the original labyrinth task, and noted that this response tended to be given by students of high mathematical abilities. Our results cast doubt upon this interpretation. Our sample of extremely able 18-19 year old students were fairly evenly split on this item. If there was some correlation between mathematical ability and answering “can’t tell”, we would expect a substantially higher percentage of our participants to have answered in this way.

It is also clear that the context in which the question is set has a significant influence upon responses. Conditional sentences in mathematical contexts appear, across the sample, to be treated differently to conditional sentences in non-mathematical contexts. Paradoxically, the mathematical context appears to bias highly able students towards what Durand-Guerrier (2003) believed was the mathematically incorrect answer.

THE RAMSEY TEST – CONDITIONALS IN MATHEMATICS

How then can we account for these results? It seems that a mathematical context fundamentally alters the manner in which conditional statements are evaluated. But how? To see how mathematicians evaluated statement 6 in the labyrinth task we conducted 11 task-based interviews with a range of university level mathematics students. Participants were asked to solve the original task whilst speaking out loud.

In the following extract we report how one student, Rachel, responded to the original version of the task. Rachel is a postgraduate student, and had been a teaching

assistant on the Foundations of Mathematics module in which formal logic is taught to first year undergraduates.

Rachel: This one [statement 6] is wrong.

Interviewer: Why?

Rachel: Well the statement is saying that if he crossed L then he definitely crossed K , which is not true. Because you could have gone $I-L-M$ and then leave the maze and then you wouldn't have gone through K , I mean it would have been a possibility to go through both, but it's not a necessity, which makes the statement wrong.

Here Rachel is clearly not using the Ramsey Test to evaluate statement 6. Instead she interprets the conditional statement as demanding that X *necessarily* has to have gone through K if s/he went through L . The interviewer asks what would happen if more knowledge about the route became available:

Interviewer: OK. How would you react if I told you what the route was? [Describes a route that does go through L and K]. How would that affect [statement 6]? So if the person did go through K and L ?

Rachel: Well it's still wrong. Because this is just a conditional thing saying that if this happens then something else happens and this, you know, this has got to be true for all routes that cross L not just the particular one chosen. You know, as I said, you can go through K and L and still leave the maze without going through any door twice. So it's a possibility, so it's not wrong in the sense that it can never ever happen, but it's an implication that you can't make.

So Rachel clearly believed, contrary to Durand-Guerrier's logical analysis, that implication *does* carry necessity, for her it is *not* merely a set of truth values. When the interviewer points out the truth table for " $P \Rightarrow Q$ " and argues that an analysis along these lines suggests that the sentence could be either true or false depending on the particular route, Rachel remains unconvinced:

Rachel: I don't believe your argument, I'm sorry.

Interviewer: So where's the flaw in my argument?

Rachel: Umm, I don't know... umm, I don't know, that's the problem I have at the moment [...]

Interviewer: So if you were teaching some first years and [the labyrinth task] was a question on their exam, what answer would you hope that they'd give to number 6?

Rachel: That's a tricky question. If they'd just done logic and they'd drawn a truth table and said, you know, you've got both in the last column so you can't tell which it is, I suspect I would feel obliged to give them full marks.

Whilst Rachel, a successful mathematician, clearly understands the argument based on truth values, she remains unconvinced by it. Nevertheless she grudgingly accepts it may be 'correct' in some unnatural formal sense.

Although the transcript indicates that Rachel has all the information required to perform a Ramsey Test successfully and deduce that $P(Q|P)=0.5$, she resists. Instead Rachel seems to be demanding that, for a conditional statement to be true in mathematics, $P(Q|P)$ must be equal to 1. That is to say that for the statement “ $P \Rightarrow Q$ ” to be evaluated as “true”, once P has been added hypothetically to her stock of knowledge, she is demanding to be able to conclude Q with absolute certainty. Thus the Ramsey Test appears to operate differently in mathematical contexts for mathematicians than in general day-to-day life.

This idea of a modified Ramsey Test has strong connections with Weber & Alcock’s (2005) notion of a *warranted conditional*. Drawing on Toulmin’s (1958) work on informal logic and argumentation, they define a conditional “ $P \Rightarrow Q$ ” to be *warranted* if the consequent Q necessarily follows, by some valid mathematical procedure, from the antecedent P . They suggest that a conditional statement is invalid in mathematics unless it is warranted. Recall our example from the mathematical labyrinth task:

If $a_n \rightarrow 0$, then $\sum a_n$ converges. (*)

In Weber & Alcock’s terms, this statement is unwarranted as the consequent does not necessarily follow from the antecedent. However the statement is true for certain sequences (a_n) . In the language of Ramsey (1931) and Evans & Over (2004). Weber & Alcock are saying that when evaluating this statement a person hypothetically adds the belief that (a_n) tends to zero to their stock of knowledge, and evaluates their degree of belief in the series converging. If their degree of belief is not 100%, or thereabouts, the conditional is rejected as unwarranted and false.

Evaluating the Ramsey Test in mathematical contexts may be a non-trivial matter, and in some circumstances it may rely more upon general knowledge of the subject matter than it does on the actual argument contained in the proof. Indeed it has even been argued that acceptable mathematical proofs routinely contain ‘gaps’ that break the chain of implications justified by ‘valid’ mathematical warrants (Fallis, 2003).

WHICH RAMSEY TEST? THE ROLE OF CONTEXT

Our results clearly indicate that the majority of first year undergraduates evaluated statement (*) as being false, suggesting that they conducted a modified version of the Ramsey Test. However, for the original version of the task roughly half the sample answered “false” and half answered “can’t tell”. So if our analysis is correct then there was no clear agreement whether to use the modified version of the Ramsey Test or the standard version. We argue that this is because the context was less clearly mathematical in this version. The labyrinth task is not overtly a mathematical question, despite appearing in a mathematics test. However, the mathematical labyrinth task is visibly mathematical: it refers to subject matter from real analysis.

We believeⁱⁱ that mathematicians judge everyday conditionals – such as “if you’re in Birmingham, then you have a good choice of Indian takeaways” – in the same manner as the rest of the population. Namely, according to Evans and Overs’s (2004) theory, they conduct a standard Ramsey Test to fix their degree of belief in Q given

P. But when in the mathematics classroom, the lecture theatre or the office, they seem to behave differently: they use a modified version of the Ramsey Test, which demands that $P(Q|P)$ is equal to 1.

We, therefore, believe that Durand-Guerrier (2003) should not have been surprised that the teachers she spoke to considered “can’t tell” to be the incorrect answer to part 6 of the labyrinth task. The teachers were evaluating what they believed to be a mathematical statement in a manner appropriate for a mathematical context.

References

- Durand-Guerrier, V. (2003). Which notion of implication is the right one? From logical considerations to a didactic perspective. *Educational Studies in Mathematics*, 53, 5-34.
- Evans, J. St. B. T., and Over, D. E. (2004). *If*. Oxford: OUP.
- Evans, J. St. B. T., Over, D. E., and Handley, S. J. (2005). Suppositions, extensionality, and conditionals: A critique of the mental model theory of Johnson-Laird and Byrne (2002). *Psychological Review*, 112(4), 1040-1052.
- Hadjichristidis, C., Stevenson, R. J., Over, D. E., Sloman, S. A., Evans, J. St. B. T., and Feeney, A. (2001). On the evaluation of ‘if p then q’ conditionals. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. Edinburgh, Scotland.
- Fallis, D. (2003). Intentional gaps in mathematical proofs. *Synthese*, 134, 45-69.
- Johnson-Laird, P. N., and Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics and inference. *Psychological Review*, 109, 646-678.
- Ramsey, F. P. (1931/1990). General propositions and causality. D. H. Mellor (Ed.), *Philosophical papers* (p. 145-163). Cambridge: CUP.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: CUP.
- Weber, K., and Alcock, L. (2005). Using warranted implications to understand and validate proofs. *For the Learning of Mathematics*, 25(1), 34-38.

ⁱ The assessment was set up in such a way so that the students’ overall mark would be improved if they performed well on the experimental question, but that if they scored below their average mark for the rest of the test, the experimental section would be ignored. Thus all the participants had an incentive to take the experimental questions seriously, but would not be disadvantaged by a poor performance on this section.

ⁱⁱ Note, however, that we have no empirical evidence to back this belief up. More work is needed on individual differences in contextual reasoning behaviour.