Conditional Inference in Mathematics

Lara Alcock
l.j.alcock@lboro.ac.uk

Ben Davies
ben.davies@soton.ac.uk

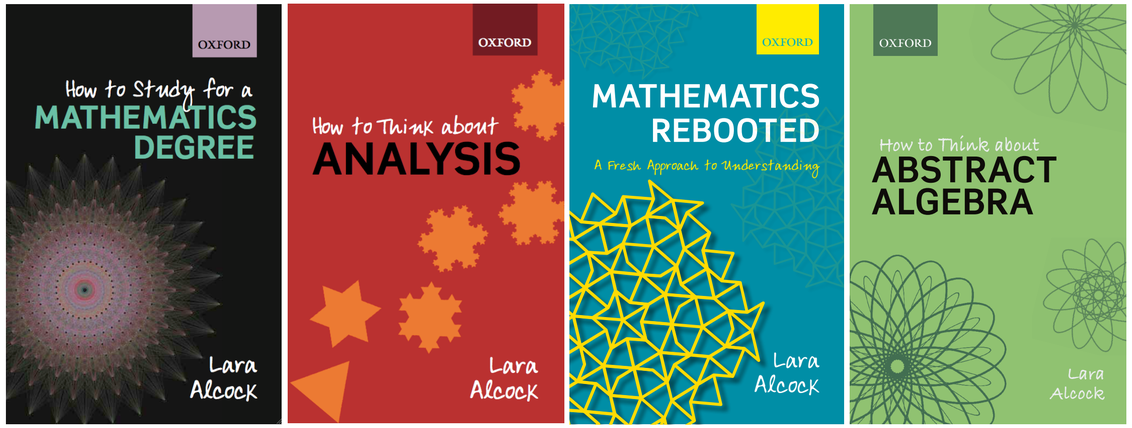
Matthew Inglis
m.j.inglis@lboro.ac.uk

LEVERHULME TRUST

Loughborough University

University of Southampton

Who am I?



- BSc and MSc Mathematics (Warwick).
- PhD Mathematics Education (Warwick).
- Four years at Rutgers (USA) in Mathematics and Education.
- Two years at Essex as a Teaching Fellow in Mathematics.
- 16 years (so far) in the Department of Mathematics Education at Loughborough University.
- Research in undergraduate mathematics education.
- Teaching education but primarily undergraduate mathematics.

Introduction

Introduction

- Logical reasoning in central in mathematics, overtly so in proof.
- Students must validate inferences in proof construction and comprehension.
- Experts do considerable back-and-forth reading (Inglis & Alcock, 2012).
- One important aspect of logical reasoning is *conditional inference*…
- …making inferences from statements of the form ‘If A then B’.
- Everyday conditionals lend themselves to distinct interpretations:
    - Material: ‘If it is a dog, then it is an animal’;
    - Biconditional: ‘If you mow the lawn, I will give you $5’ (Cummins et al., 1991).

Introduction

- Mathematics students must learn to restrict their interpretations.
- There is extensive work on this in cognitive psychology.
- There is only a little work on it in mathematics education.
- Conditional inference tasks exist for abstract and everyday content:
    - ‘If the letter is A then the number is 1.’
    - ‘If John studies hard, then he does well on the test.’
- But not for mathematical content.
- This talk is about developing and using mathematical tasks.

Theoretical Background

Abstract inference

If the letter is Z, then the number is 1.
The letter is Z.
Conclusion: The number is 1.

If the letter is Z, then the number is 1.
The number is 1.
Conclusion: The letter is Z.

If the letter is Z, then the number is 1.
The letter is not Z.
Conclusion: The number is not 1.

If the letter is Z, then the number is 1.
The number is not 1.
Conclusion: The letter is not Z.

Abstract inference

If the letter is Z, then the number is not 1.
The letter is P.
Conclusion: The number is 1.

If the letter is not Z, then the number is 1.
The number is 3.
Conclusion: The letter is Z.

If the letter is not Z, then the number is not 1.
The letter is N.
Conclusion: The number is not 1.

If the letter is not Z, then the number is not 1.
The number is 1.
Conclusion: The letter is Z.

Abstract conditional inference

Inference	Conditional Premise	Categorical Premise	Conclusion	Validity	Percentage Acceptance
modus ponens					
denial of the antecedent					
affirmation of the consequent					
modus tollens					

Evans, Handley, Neilens & Over (2007)

Abstract inference and maths

- Studying mathematics intensively at age 16-18 improves performance in abstract conditional inference (e.g., Attridge, Doritou & Inglis, 2015).
- Performance on abstract conditional inference predicts performance in undergraduate mathematics (e.g., Alcock & Attridge, 2023).
- Mathematics teaches and rewards distinguishing valid/invalid inferences.
- Mathematics is viewed as an abstract subject.
- But it is not abstract in this way, not devoid of meaningful content.
- On the contrary, it is very meaningful.

Conditional inference and maths

- Research with meaningful mathematical content is minimal.
- Students do not use single meaning for conditionals (Dawkins & Norton, 2022).
- Sometimes conflate a conditional and converse (Hoyles & Küchemann, 2002):

  If the product of two whole numbers is odd, then their sum is even.

  If the sum of two whole numbers is even, then their product is odd.
- Others have suggested this is over-stated (for undergraduates)…
- …but that semantic content affects interpretations (Durand-Guerrier, 2003).

Causal conditional inference

If John studies hard, then he does well on the test.
John studies hard.
Conclusion: John does well on the test.
Modus Ponens (MP): VALID

If John studies hard, then he does well on the test.
John does well on the test.
Conclusion: John studied hard.
Affirmation of the Consequent (AC): INVALID

If John studies hard, then he does well on the test.
John does not study hard.
Conclusion: John does not do well on the test.
Denial of the Antecedent (DA): INVALID

If John studies hard, then he does well on the test.
John does not do well on the test.
Conclusion: John did not study hard.
Modus Tollens (MT): VALID

Causal conditional inference

- Cognitive psychology has a long history of studying causal conditional inference (e.g., Evans & Over, 2004; Oaksford & Chater, 2020).
- Task instructions vary in emphasising everyday reasoning or logic.
- Everyday conditionals vary in believability (Evans, Handley, Neilens & Over, 2010):

Ability	Believability	MP	DA	AC	MT
High	High	94	41	41	48
	Low	92	40	41	46
Low	High	89	60	66	58
	Low	75	53	58	51

Theoretical accounts

- Mental models ‘If p then q’ (e.g., Johnson-Laird & Byrne, 1991):

  p q

- Doesn’t account for effects of content.
- And everyday reasoning operates under uncertainty.
- Probabilistic (e.g., Oaksford & Chater, 2007):

  P(If p then q) = P(q|p), conclusions accepted when probability high.

- Dual-strategy (e.g. Markovits et al.,. 2013):
  - Probabilistic and counterexample strategies available.
  - Counterexample use related to time, cognitive capacity, and strategy.

Task design

- Aim to construct a mathematical conditional inference task to parallel those with everyday causal content.
- Four inference types, varying believability.
- Issues in mathematical task design:
    - ‘If *x* is less than 2, then *x* is less than 5’ is a predicate;
    - But often assume universal quantification;
    - And causal tasks are like this too (no specific John or test).
- How to measure believability?

Task design

- Some studies have asked for direct believability ratings (cf. Evans et al., 2010)…
- …assuming universal quantification, mathematical conditionals true or false.
- Some have asked for distinct counterexamples (cf. Cummins et al., 1991)…
- …counterexamples in mathematics are often singular (‘zero’) or in infinite sets (‘the negative numbers’), so this would be artificial or impossible.
- But believability should vary, and not necessarily with truth:
    - ‘If X is a square then X is a parallelogram’: true but not believable? (cf. Grice, 1989)
    - If $x < 3$ then $1/x > 1/3$: false but believable? (cf. Alcock & Attridge, 2023)

Study 1: Believability and Conditional Inference

Ben Davies

Methodology

- Measure believability using *comparative judgement*:
  - Multiple judges make pairwise judgements (Jones & Davies, 2023);
  - Judgements used collectively to generate scores (Bradley & Terry, 1952);
  - Count ‘wins’ then iteratively update to reflect ease of winning.
- Useful for constructs that resist rubric-based assessment:
  - Conceptual understanding (Jones et al., 2019);
  - Problem solving (Jones & Inglis, 2015);
  - Conceptions of proof (Davies et al., 2021);
  - Standards in examination papers over time (Jones et al., 2016).

Generating conditionals

- Asked 8 mathematics education researchers to generate conditionals:
    - Cover a range of mathematical topics;
    - Have plausibly related antecedent and consequent;
    - Not obviously false;
    - Not use additional connectives (‘and’, ‘or’, ‘not’);
    - Vary in believability.
- Also asked to rank for believability (to reinforce variation).

Generating conditionals

- Imaginative range of true and false conditionals.
- Some additional connectives sneaked in via $\leq$, $\neq$.
- Removed some conditionals because phrasing too complex for an inference task (e.g., hard to write AC items).
- Removed some because converse true (confound for DA and AC).
- Good spread of topics.
- Made the collection up to 40 by adapting conditionals from the literature (Alcock, 2013; Alcock & Attridge, 2023; Dawkins & Norton, 2022; Durand-Guerrier, 2003; Houston, 2009; Hoyles & Küchemann, 2002; Selden & Selden, 2003).

Comparative judgement

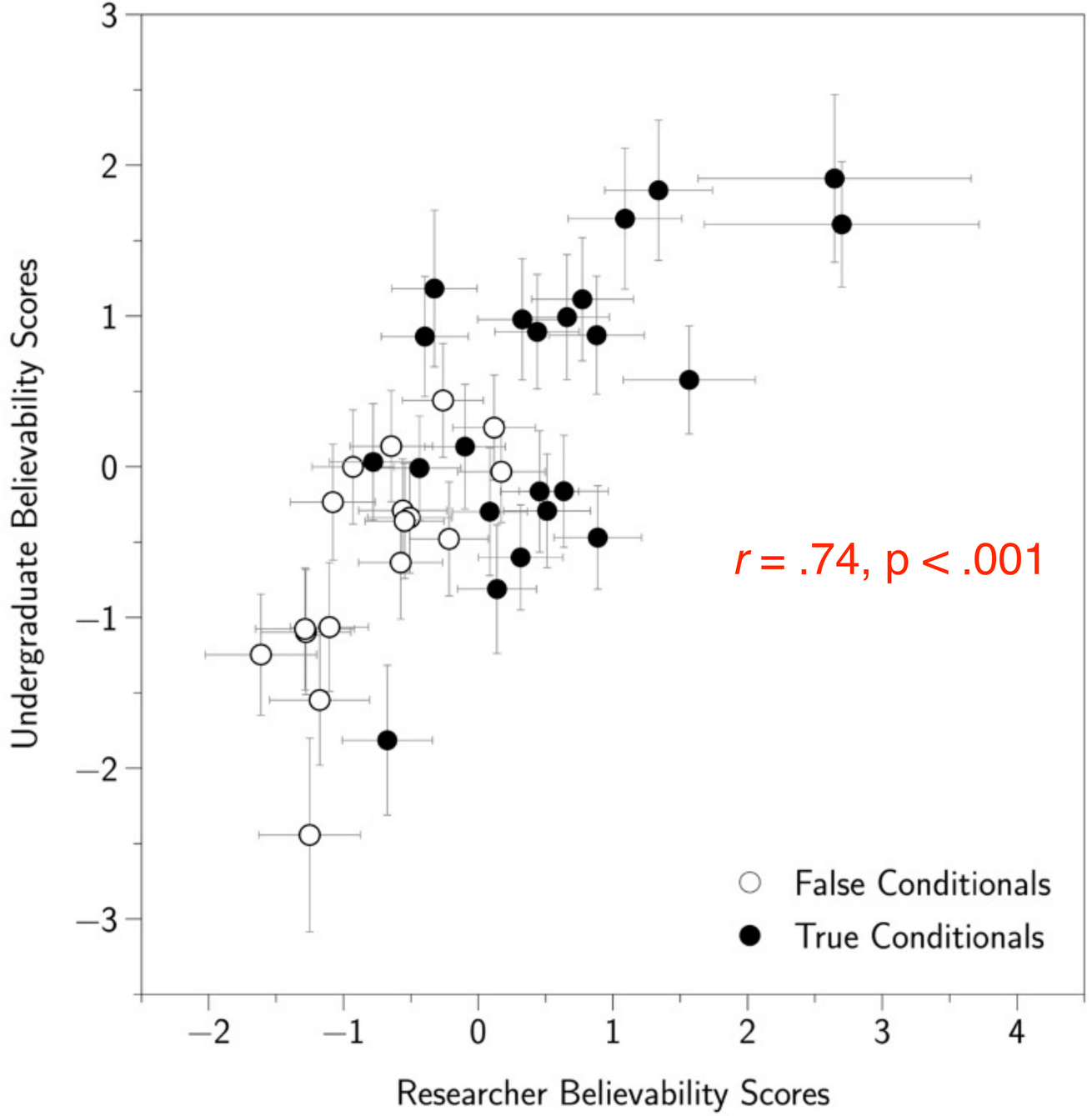
- 40 conditionals, standardised phrasing.
- Each conditional judged 20 times so 400 judgements (Verhavert et al., 2019).
- Comparative judgement with:
    - Same eight researchers, 50 judgements each;
    - 12 mathematics undergraduates, 50 judgements each.
- Assessed reliability (Verhavert et al., 2019):
    - Researchers SRR = .84, IRR = .73 (good);
    - Undergraduates SRR = .83, IRR = .63 (acceptable).

If $x < 2$, then $x < 5$.	1
If n is a multiple of 6, then n is a multiple of 3.	2
If line L is tangent to circle C, then L is perpendicular to a radius of C.	3
If n is a multiple of 4, then n^2 is a multiple of 4.	4
If C is a circle, then its width is the same when measured in any direction.	5
If polygon P is a square, then it is a rhombus.	6
If $a = b$, then $an = bn$.	7
If $x = \sqrt{y}$, then $x^2 = y$.	8
If $x = -4$, then $x^2 + x - 12 = 0$.	9
If X is a circle, then X is an ellipse.	10
If circle C and square S have the same perimeter, then C has bigger area than S.	11
If polygon P is a rhombus, then it has perpendicular diagonals.	12
If n is the product of two consecutive integers, then n is even.	13
If $x < 0$ then $x^3 < x^2$.	14
If $x^2 = y^2$, then $xy = yx$.	15
If fraction x has denominator 7, then it is equivalent to a non-terminating decimal.	16
If quadrilateral Q is cyclic, then it is convex.	17
If x is an integer, then $x^2 > x$.	18
If $x - 12{,}345 = 0.67$, then $x > -12{,}345.67$.	19
If polygon P is a rectangle, then every line through its centre cuts its area in half.	20

If n is the sum of four consecutive numbers, then n is a multiple of 4.	21
If line L is tangent to curve C, then L intersects C at only one point.	22
If $x = 3$, then $2(x - 3) = 5x - 3(x + 2)$.	23
If the product of two whole numbers is odd, then their sum is even.	24
If rectangle R has area 10cm^2, then its perimeter is greater than 10cm.	25
If $a > b$, then $a^2 > b^2$.	26
If $x < 3$, then $1/x > 1/3$.	27
If $a > b$, then $ac > bc$.	28
If equation E is quadratic, then it has exactly two roots.	29
If n is prime, then $n + 1$ is even.	30
If quadrilateral Q has a reflex angle, then it will tesselate.	31
If $\sin x > 0$, then $\cos x < 1$.	32
If n is a multiple of 13, then it has an even number of factors.	33
If x is positive, then $\tan x > \sin x$.	34
If the side lengths of rectangle R are doubled, then its area is doubled.	35
If function f is polynomial, then f has a real root.	36
If the mean of dataset D is greater than 100, then the median is greater than 100.	37
If the mean of dataset D is 7, then the median is 7.	38
If $a = 42$, then $a \times b > 42$.	39
If composite number c ends in a 3, then it is a multiple of 3.	40

Believability and truth

- Researchers:
  - True (N = 23) M = 0.554, SD = 0.907;
  - False (N = 17) M = -0.749, SD = 0.519;
  - $t(38) = -5.31$, $p < .001$.
- Undergraduates:
  - True (N = 23) M = 0.435, SD = 0.952;
  - False (N = 17) M = -0.589, SD = 0.742;
  - $t(38) = -3.68$, $p < .001$.



Discussion

- Believability is a meaningful construct in mathematics.
- Several features good for conditional inference task design:
    - True conditionals spread out, so can avoid truth confound;
    - Agreement across groups, so potentially task widely suitable.
- But maybe the task is artificial?
    - Maybe people *can* consider believability when asked…
    - …but this would not actually affect their reasoning.
- Conditional inference task provides evidence.

Method

Matthew Inglis

- Constructed a task with true conditionals:

relatively believable

- If x is less than 2, then x is less than 5.
- If n is a multiple of 6, then n is a multiple of 3.
- If line L is tangent to circle C, then L is perpendicular to a radius of C.
- If $x = \sqrt{y}$, then $x^2 = y$.

relatively unbelievable

- If $x^2 = y^2$, then $xy = yx$.
- If rectangle R has area 10cm^2, then its perimeter is greater than 10cm.
- If quadrilateral Q has a reflex angle, then it will tesselate.
- If $\sin x > 0$, then $\cos x < 1$.

Method

Tuya Sa

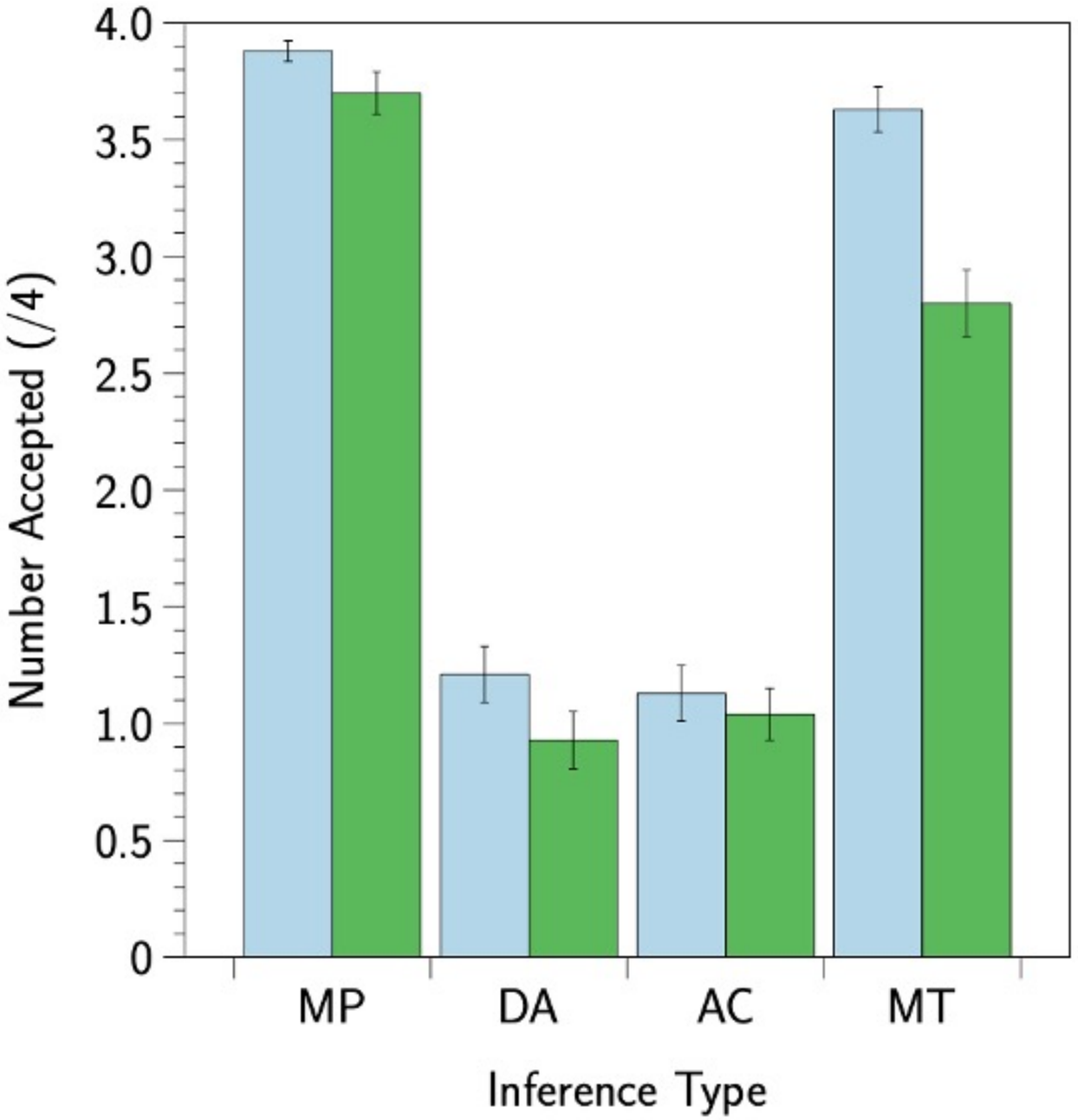
- Each conditional paired with MP, DA, AC, MT inferences.
- So 32 items.
- Instructions emphasised logic (‘Does the conclusion follow necessarily?’).
- Participants in a course called *Mathematical Thinking.*
- Instruction on logic week preceding data collection.
- Up to 20 minutes, with space to write actual finish time.
- 56 participants included; 52 completed all questions.

Results

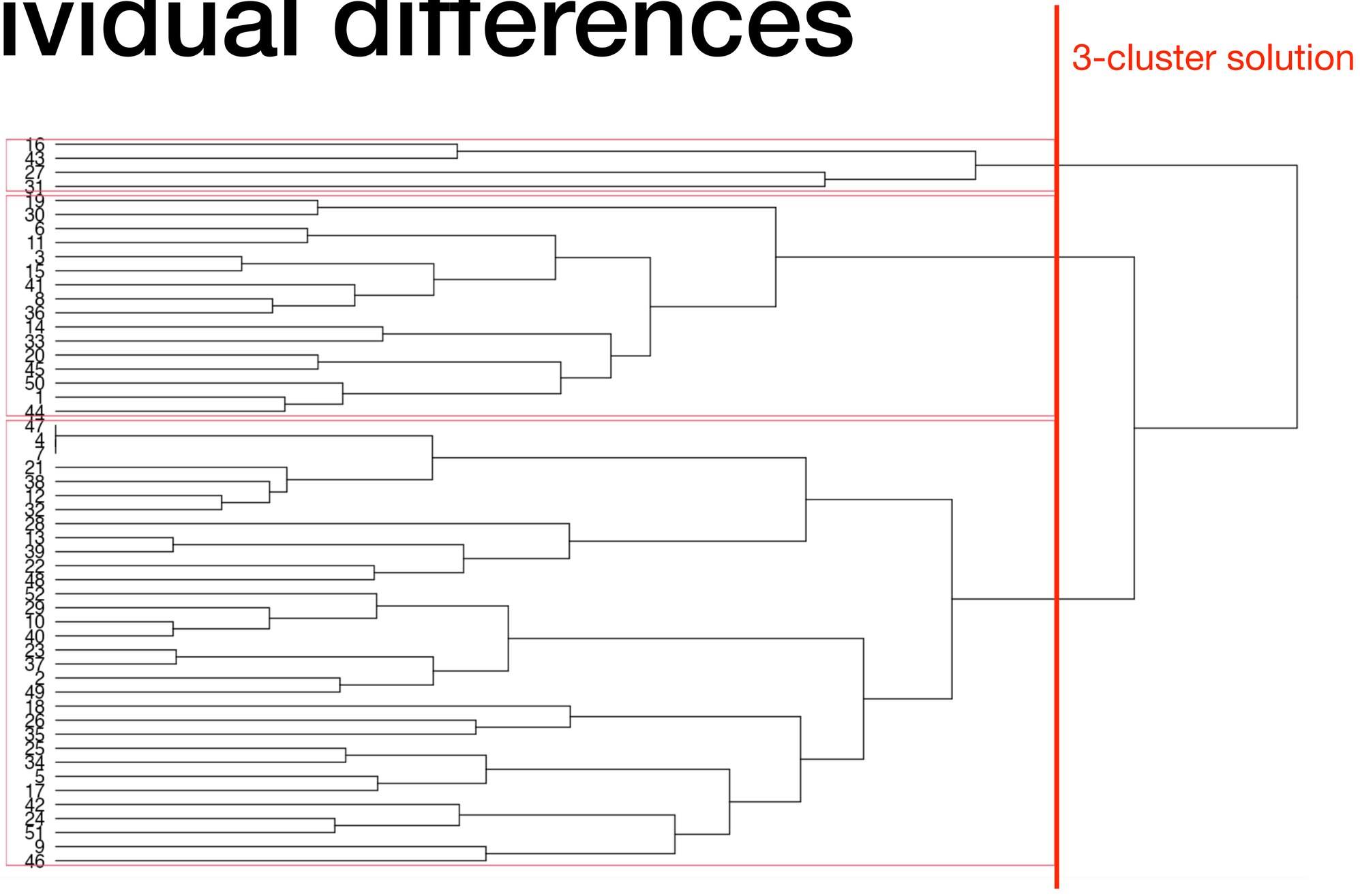
- Time for completers M = 14.3 minutes, SD = 3.13 minutes.
- Normative score M = 25.7 (80%), SD = 3.21 (10%).
- Score and time not related, r = -.146, p = .308.
- Does believability affect inference acceptance?
    - 4 (inference type) x 2 (believability) ANOVA (with Greenhouse-Geisser correction).
- Are there systematic individual differences?
    - Hierarchical cluster analysis (Ward's method with Euclidean squared distance function).

Results: Main

- Main effect of inference type, $p < .001$, $\eta_p^2 = .859$.
- Main effect of believability $p < .001$, $\eta_p^2 = .261$.
- Significant interaction $p < .001$, $\eta_p^2 = .103$.



Individual differences



Results: Clusters

- One cluster had only 4 (low-scoring) participants.
- Cluster 1 had 32 participants; Cluster 2 had 16.

		MP	DA	AC	MT
Cluster 1	Believable	98.5	18.8	21.1	92.3
	Unbelievable	97.8	20.3	29.0	79.0
Cluster 2	Believable	100.0	54.8	40.8	95.3
	Unbelievable	97.0	25.0	21.9	67.3

Discussion

- Inference in mathematics is like everyday inference in being affected by believability.
- This effect is driven by some students being less normative/mathematical in that they are more influenced by content.
- Specifically, these students were more likely to accept invalid inferences from believable conditionals - semantic content has an effect.
- Next questions:
    - Is this effect uniform across mathematical and everyday content?
    - Where does abstract content fit in?

Study 2: Conditional Inference Across Contexts

Task: three contents

- Abstract:
    - If the letter is C, then the number is 7.
    - If the letter is N, then the number is 2.
- Mathematical relatively believable (true):
    - If n is a multiple of 6, then n is a multiple of 3.
    - If x is less than 2, then x is less than 5.
- Mathematical relatively unbelievable (true):
    - If $\sin x$ is greater than 0, then $\cos x$ is less than 1.
    - If quadrilateral Q has a reflex angle, then it will tessellate.

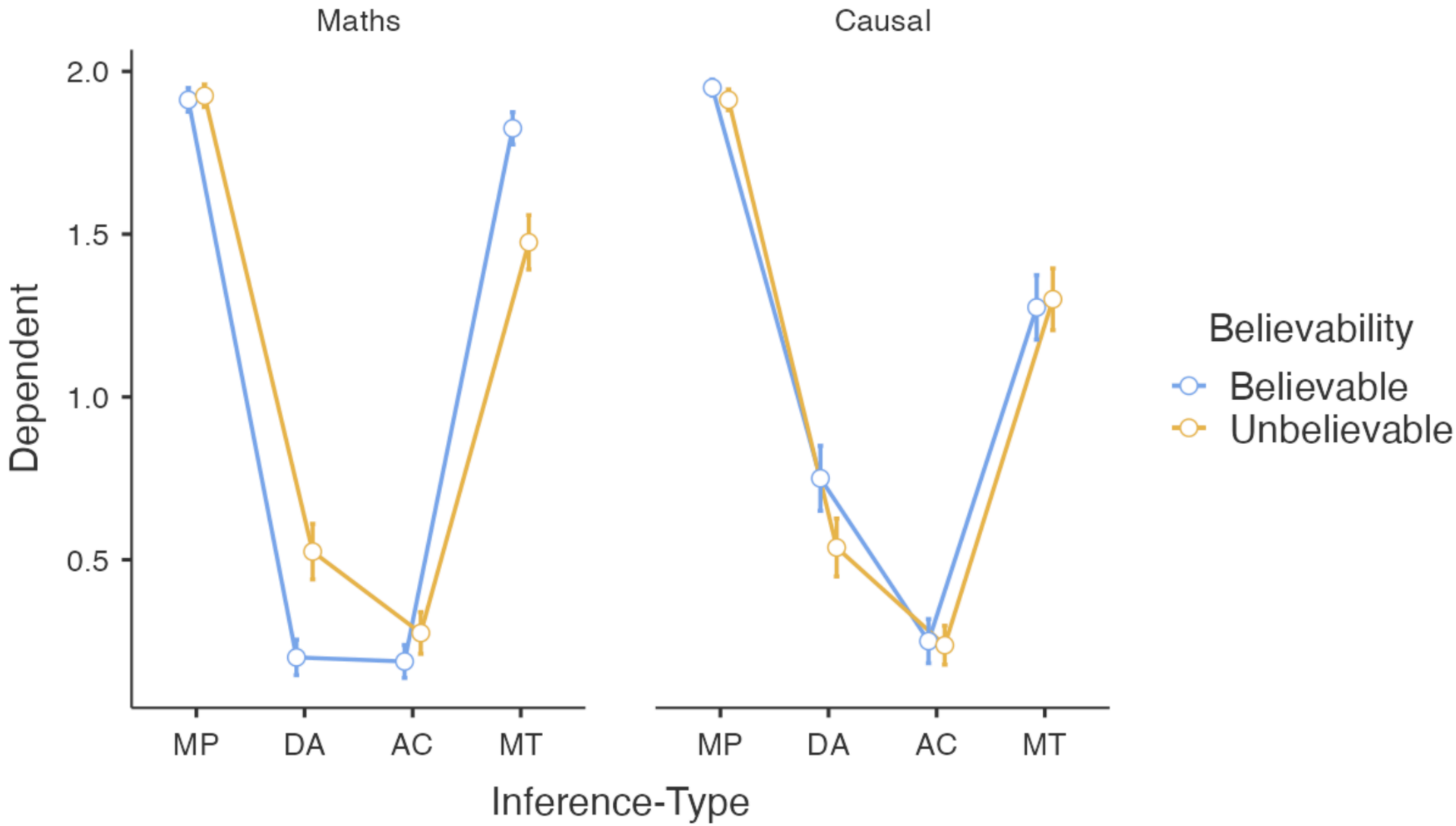
Task: three contents

- Everyday causal relatively believable (Evans, Handley, Neilens & Over, 2010):
    - [82] If car ownership increases, then traffic congestion will get worse.
    - [79] If nurses' salaries are improved, then recruitment of nurses will increase.
- Everyday causal relatively unbelievable:
    - [54] If divorce is made more difficult, then the number of marriages will decrease.
    - [45] If foreign investment is encouraged, then the UK car industry will revive.
- Three contents counterbalanced, items randomised, 80 participants.

Content and believability

4 (inference type) x 2 (content) x 2 (believability) ANOVA.

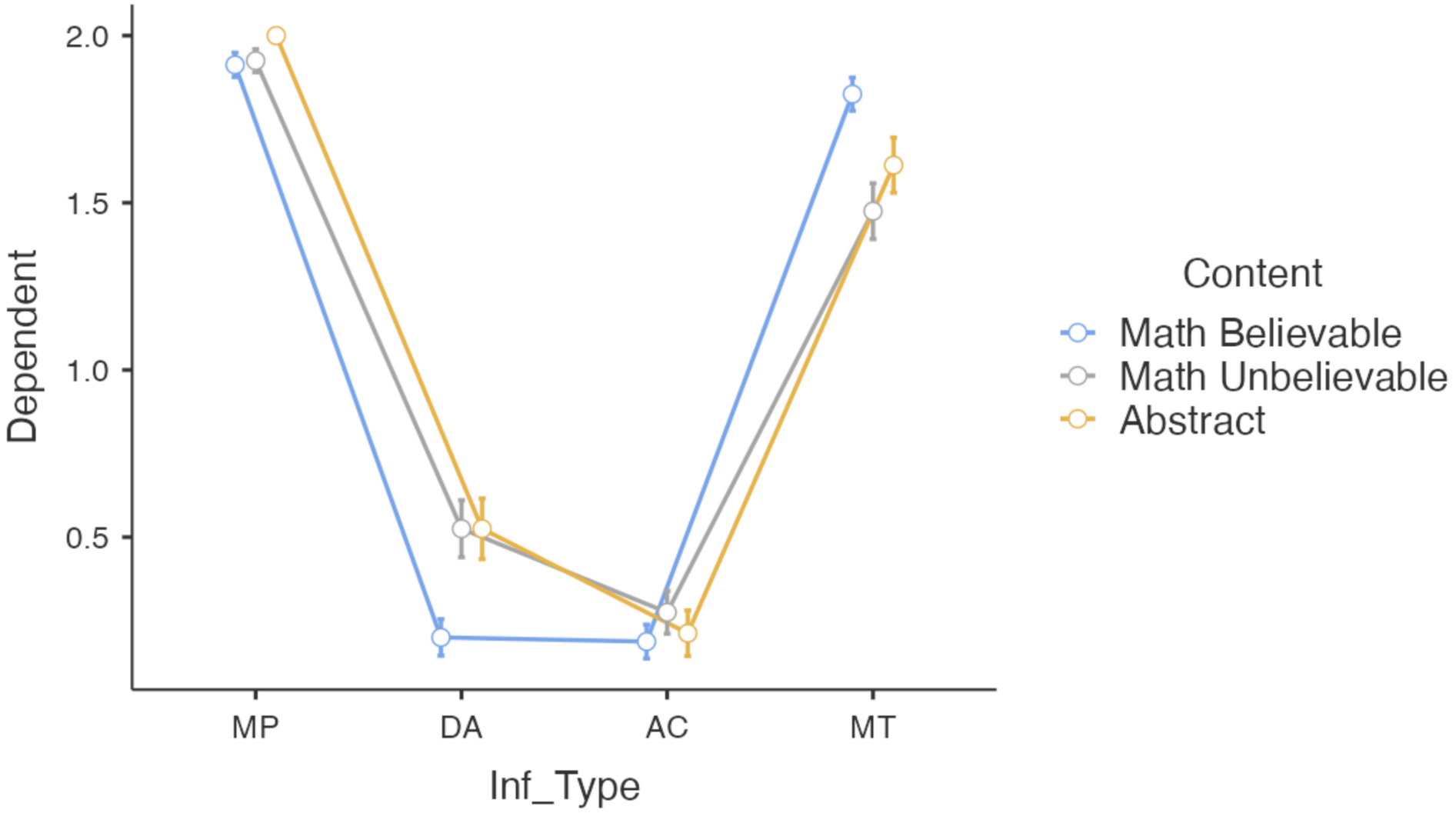
3-way interaction ($p < .001$).



Mathematics and abstract

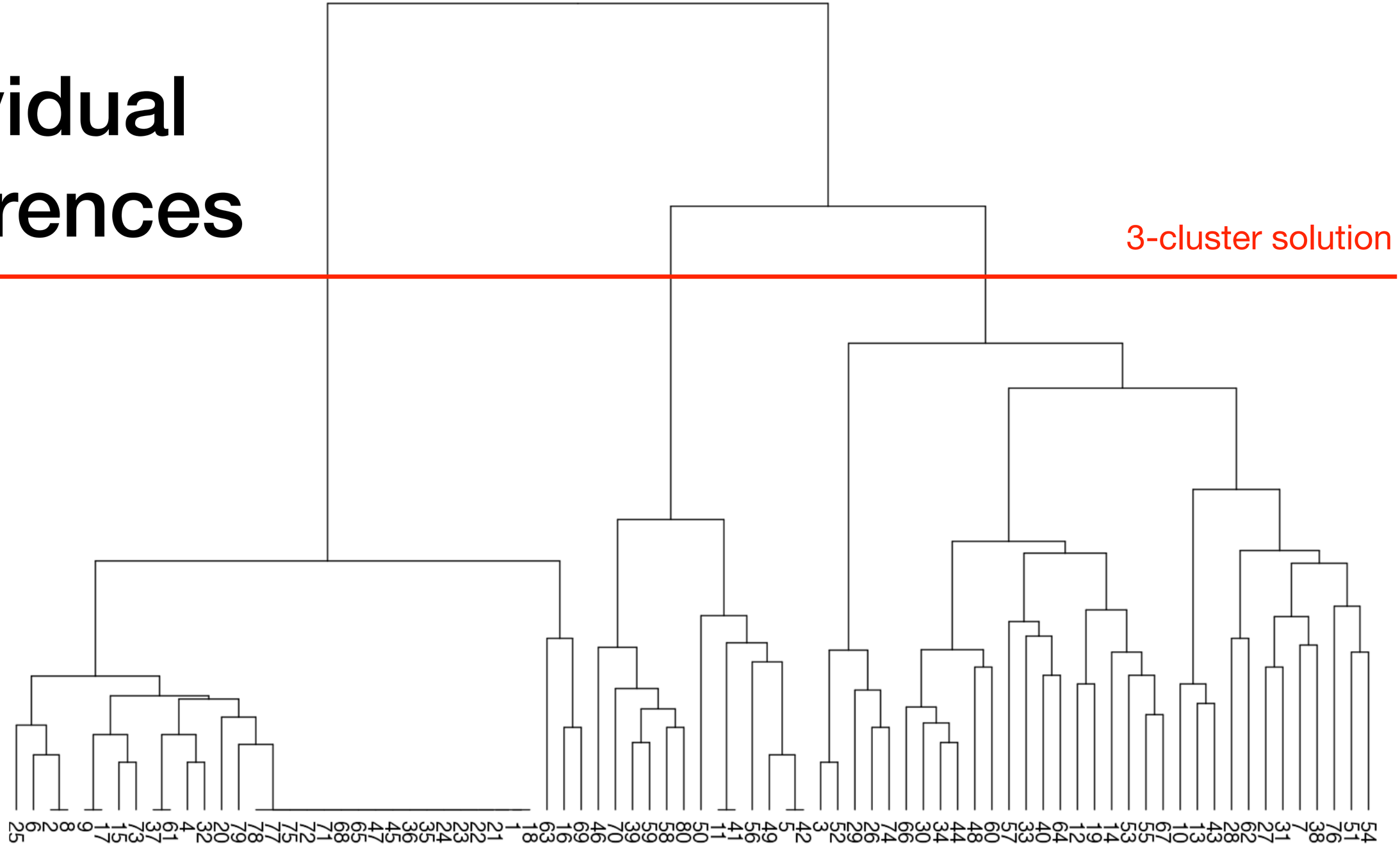
4 (inference type) x 3 (content) ANOVA.

Inference type x content interaction ($p < .001$).



Individual differences

3-cluster solution



Results: Clusters

	N = 34				N = 33				N = 13			
	MP	**DA**	**AC**	**MT**	**MP**	**DA**	**AC**	**MT**	**MP**	**DA**	**AC**	**MT**
Math B	100	6	0	100	93	17	18	88	93	4	12	77
Math U	100	7	4	90	91	53	25	71	100	8	8	38
Caus B	99	4	1	100	96	85	24	50	100	4	12	4
Caus U	99	1	0	99	93	62	27	52	96	4	4	12
Abs	100	4	3	94	100	56	20	79	100	8	8	50

impressively normative

less normative AC and MT; very high DA acceptance

pretty normative DA, AC; (much) lower MT acceptance

Discussion

- Perhaps we have done moderately well in teaching students that a conditional ‘If A then B’ is not equivalent to its converse ‘If B then A’.
- But also, a conditional is not equivalent to its inverse ‘If not-A then not-B’:

  If the product of two whole numbers is odd, then their sum is even.

  If the product of two whole numbers is not odd, then their sum is not even.
- Becomes more salient in undergraduate mathematics:

  If $f : \mathbb{R} \to \mathbb{R}$ is differentiable at *a*, then it is continuous at *a*.

  If $f : \mathbb{R} \to \mathbb{R}$ is not differentiable at *a*, then it is not continuous at *a*.
- Extremely important for understanding theorem meanings.

Study 3:
Believability or Easiness?

Possible confound?

- For everyday tasks, easiness not considered.
- For mathematical tasks, a confound might exist and be unavoidable.
- Mathematical relatively believable (true):
  - If *n* is a multiple of 6, then *n* is a multiple of 3.
  - If *x* is less than 2, then *x* is less than 5.

easier

- Mathematical relatively unbelievable (true):
  - If sin *x* is greater than 0, then cos *x* is less than 1.
  - If quadrilateral Q has a reflex angle, then it will tessellate.

harder

Task design

- 20 true mathematical conditionals from Study 1.
  - Avoided those with antecedent about a single object.
  - Adjusted those whose consequent could not (reasonably) be false.
- 20 true-ish everyday causal conditionals with varying believability:
  - 8 on societal situations (removed ‘will’) (Evans et al., 2010).

    ‘If nurses’ salaries improve, then recruitment of nurses increases.’
  - 12 on everyday physical causes (Cummins, 1995).

    ‘If Joe cut his finger, then it bled.’

Comparative judgement

- Obtained believability and easiness measures for both sets together.

Prompt	Participants	Judgements	Approx judgements per item	SRR	IRR
Believability	105	1575	79	.88	.76
Easiness	110	1650	83	.84	.70

- Believability and easiness significantly correlated:
  - Mathematical $r = .858$, $p < .001$;
  - Causal $r = .520$, $p = .019$.
- This is a problem for using these as predictors of inference acceptance.

Predicting inference acceptance

- Selected 10 of each so that correlations lower:
  - Mathematical $r = .435$, $p = .208$;
  - Causal $r = .141$, $p = .697$.
- Combined with MP, DA, AC, MT inferences.
- Participants had 5 random mathematical and 5 random causal items.
- Content type counterbalanced.
- 175 participants.
- ~22 responses per item.

Predicting inference acceptance

- By-items regression predicting proportion accepted from:
  - Inference type (MP as baseline, DA, AC, MT as predictors);
  - Content (mathematical, causal);
  - Believability;
  - Easiness;
  - Content x believability and content x easiness interaction terms.

Predicting inference acceptance

Predictor	Estimate	SE	β	95% CI	p
Intercept	0.903	0.035			< .001
DA	-0.665	0.040	-0.812	[-0.910, -0.714]	< .001
AC	-0.775	0.040	-0.946	[-1.044, -0.849]	< .001
MT	-0.146	0.040	-0.178	[-0.276, -0.081]	< .001
Content	0.035	0.041	0.049	[-0.066, -0.164]	.397
Believability	0.094	0.031	0.189	[0.065, 0.314]	.003
Easiness	0.062	0.047	0.116	[-0.059, 0.291]	.190
Content x Believability	-0.082	0.046	-0.144	[-0.305, 0.017]	.078
Content x Easiness	-0.053	0.054	-0.095	[-0.290, 0.100]	.336

Discussion

Summary

- Mathematical conditional inference is affected by believability, with substantive individual differences.
- Some students respond close to normatively; others reject fewer invalid inferences, with a believability effect.
- Particular concerns about DA inferences / inverses.
- Students respond similarly across mathematical, abstract and everyday causal content, except more normatively for believable mathematical conditionals.
- Believability and easiness are correlated, but the effect is attributable to believability, not easiness.

Theoretical implications

- What is ‘believability’?
- Probabilistic construct P(If p then q) = P(q|p).
- Dual-process account: some more inclined to seek counterexamples.
- In mathematics, could just use the inference form.
- For experts, triple-process account?
  - Probabilistic;
  - Counterexample-based;
  - Syntactic.

Request

- Among other things, I'd like to investigate conditional inference in more expert mathematicians.
- Might use eye-tracking.
- Want to collect conditionals that:
  - Have content that most students would see in the first year of a mathematics degree (but not until then);
  - Include some true biconditionals, some not;
  - Include some well-known theorems, some not.
- Want to measure believability and easiness using comparative judgement.

Thank you.

LEVERHULME TRUST