Knowledge of Examples Affects Conditional Reasoning with Mathematical Content

David W. Braithwaite1 and Anna N. Rafferty2

1 Florida State University

2 Carleton College

Author note:

Materials, data, and analysis code for all studies are available on [Open Science Framework](https://osf.io/cy59g/files/osfstorage).

Contact information:

David W. Braithwaite

Department of Psychology

Florida State University

1107 W. Call Street

Tallahassee, FL 32304

USA

850-644-3409

[braithwaite@psy.fsu.edu](mailto:braithwaite@psy.fsu.edu)

# Abstract

Knowledge affects human deductive reasoning, but the mechanisms by which this occurs are not fully understood. Example knowledge—the ability to generate and categorize specific examples of general possibilities—is proposed to play a central role, and individual differences in such knowledge are proposed to contribute to differences in deductive reasoning. To test these hypotheses, four studies investigated the role of example knowledge in adults’ conditional reasoning about algebra. Individual differences in domain-specific knowledge predicted conditional reasoning about algebra when controlling for everyday and abstract conditional reasoning (Study 1), individuals spontaneously referred to examples during conditional reasoning about algebra (Study 2), individual differences in example knowledge predicted differences in conditional reasoning about algebra when controlling for general algebra knowledge and everyday conditional reasoning (Study 3), and training designed to increase example knowledge improved conditional reasoning about algebra (Study 4). The findings suggest that deductive reasoning in the domain of math is in part a domain-specific skill, insofar as it depends on domain-specific knowledge, and that one cause of this knowledge-dependence is individuals’ reliance on examples during reasoning.

Keywords:conditional reasoning; mental models; examples; individual differences; mathematical cognition

# Knowledge of Examples Affects Conditional Reasoning with Mathematical Content

When people engage in deductive reasoning involving content about which they have relevant knowledge, that knowledge affects how they reason. For example, people endorse believable inferences more than unbelievable ones (Evans et al., 1983), and endorse inferences with few or no counterexamples more than those with many or salient counterexamples (Cummins et al., 1991). To illustrate, “If water is at 0°C, then it will freeze; water is at 0°C; therefore, it will freeze” is likely to be endorsed because counterexamples (e.g., water under very high pressure) are rare. In contrast, “If a cup is dropped, then it will break; a cup is dropped; therefore, it will break” is less likely to be endorsed because counterexamples (e.g., a metal cup dropped from a low height) are readily available.

Availability of relevant knowledge varies not only among inferences, but also among individuals. Because deductive reasoning depends in part on knowledge, individual differences in knowledge should lead to differences in deductive reasoning. However, this possibility has been underexplored in prior research. Thus, it is not clear whether and to what extent differences in knowledge drive differences in deductive reasoning performance or what types of knowledge play such a role.

The present study focused on a particular type of knowledge that seemed likely to contribute to deductive reasoning: example knowledge, defined as the ability to generate and categorize specific examples of general possibilities. We hypothesized that individual differences in example knowledge affect which possibilities are considered during deductive reasoning, which in turn affects which inferences are accepted. This hypothesis implies that example knowledge should predict and causally affect deductive reasoning performance.

We tested the above hypothesis with respect to a particular type of deductive reasoning—conditional reasoning, in a domain where deductive reasoning is particularly important—math. Below, we review prior research on conditional reasoning, discuss how example knowledge could contribute to such reasoning, and apply these ideas to the domain of math.

## Conditional Reasoning

Deductive reasoning is reasoning that is used to judge whether inferences’ conclusions follow from their premises, and to generate inferences that have this property. By “conditional reasoning” we mean deductive reasoning involving conditionals—that is, statements that express if-then relations. Such reasoning is important because much of our knowledge about the world is naturally represented by conditionals, including in everyday life (“If you snooze, you lose”), science (“If water is at 0°C, then it will freeze”), law (“If you cannot afford an attorney, one will be appointed for you”), and math (“If two numbers are even, then their sum is even”).

Simple inferences with conditionals consist of a conditional premise (“If *p*, then *q*”), a categorical premise that affirms or denies the conditional’s antecedent (*p*) or consequent (*q*), and a conclusion that affirms or denies the conditional’s other part. Such inferences take four forms (Table 1): *modus ponens* (MP), *modus tollens* (MT), affirmation of the consequent (AC), and denial of the antecedent (DA). MP and MT are valid inferences, meaning that their conclusions follow from their premises. The validity of MP follows from the basic meanings of the words “if” and “then.” The validity of MT is less obvious, but can be seen by reasoning thus: “If *p* were true, then *q* would be true, but *q* is not true, so *p* must not be true.” AC and DA are invalid inferences, because their premises are compatible with the possibility that *p* is false and *q* is true, which contradicts those inferences’ conclusions. For instance, the first AC and DA inferences in Table 1 are invalidated by the fact that it may not be raining even though the lawn is wet—or equivalently, that the lawn may be wet even though it is not raining.

*Table 1*. The Four Simple Conditional Inferences, With Examples.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Modus Ponens (MP) | Modus Tollens (MT) | Affirmation of the Consequent (AC) | Denial of the Antecedent (DA) |
|  | **General Form** | | | |
| Conditional premise | If *p*, then *q*. | | | |
| Categorical premise | *p* | Not *q* | *q* | Not *p* |
| Conclusion | *q* | Not *p* | *p* | Not *q* |
|  | **Examples with Everyday Content** | | | |
| Conditional premise | If it is raining, then the lawn is wet. | | | |
| Categorical premise | It is raining | The lawn is not wet | The lawn is wet | It is not raining |
| Conclusion | The lawn is wet | It is not raining | It is raining | The lawn is not wet |
|  | **Examples with Math Content** | | | |
| Conditional premise | If *m* and *n* are both even, then *m*+*n* is even | | | |
| Categorical premise | *m* and *n* are both even | *m+n* is not even | *m+n* is even | *m* and *n* are not both even |
| Conclusion | *m+n* is even | *m* and *n* are not both even | *m* and *n* are both even | *m+n* is not even |

In experiments on conditional inference, MP and MT are usually endorsed by most individuals and AC and DA are endorsed much less often, consistent with the former two being valid and the latter two invalid. However, MP is endorsed more than MT despite both being valid, and AC and DA are often endorsed despite being invalid. Rates of endorsement typically follow the pattern MP > MT > AC ≈ DA. This pattern is robust for inferences involving abstract content (Klauer et al., 2010; Oberauer, 2006). For realistic content, results vary considerably depending on what is known about that content, a point discussed in more detail below.

Various theoretical accounts of conditional reasoning have been proposed (e.g., Braine & O’Brien, 1998; Douven et al., 2018; Evans & Over, 2004; Oaksford & Chater, 2020; Rips, 1994; Skovgaard-Olsen et al., 2016). Here, we focus on the theory of mental models (Johnson-Laird & Byrne, 2002; Khemlani et al., 2018), which inspired the present study. This theory asserts that deductive reasoning is fundamentally about possibilities. In reference to a conditional “If *p*, then *q*,” “possibility” refers to a specification of the truth or falsehood of *p* and *q*. There are four such specifications: (*p* true, *q* true), (*p* true, *q* false), (*p* false, *q* true), and (*p* false, *q* false). The conditional contradicts one of these, namely (*p* true, *q* false). In each of the four simple conditional inferences (Table 1), the categorical premise contradicts one or more of the others, leaving one or two that are consistent with both premises (Table 2).

*Table 2*. Consistency of Each Possibility With Premises of Each Simple Conditional Inference.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Possibility | MP | MT | AC | DA |
| (*p* true, *q* true) | ✓ | 🗶 | ✓ | 🗶 |
| (*p* true, *q* false) | 🗶 | 🗶 | 🗶 | 🗶 |
| (*p* false, *q* true) | 🗶 | 🗶 | ✓ | ✓ |
| (*p* false, *q* false) | 🗶 | ✓ | 🗶 | ✓ |

*Note*. ✓ and 🗶 signify consistent and inconsistent, respectively, with the inference’s premises.

In the theory of mental models, a model is a mental representation of a possibility.[[1]](#footnote-2) The theory claims that, to evaluate a conditional inference, people construct models of possibilities consistent with the conditional premise, then eliminate models that are contradicted by the categorical premise. They endorse the inference if its conclusion is true in all remaining models, and reject it otherwise. MP and MT inferences should be endorsed because their conclusions—that *q* is true or that *p* is false, respectively—are each true in the only models consistent with their premises (Table 2). AC and DA inferences should be rejected because their conclusions—that *p* is true and that *q* is false, respectively—are each false in one of the models consistent with their premises—specifically, (*p* false, *q* true) (Table 2).

People often commit errors in conditional reasoning, such as rejecting MT and endorsing AC and DA (Klauer et al., 2010; Oberauer, 2006). The theory of mental models attributes such errors to several sources (Johnson-Laird & Byrne, 2002), the most relevant of which here is failure to construct relevant models. For instance, when presented the first example of MT in Table 1, a reasoner might fail to construct models in which it is not raining, and consider only the model “it is raining and the lawn is wet.” This model would be eliminated by the categorical premise, “the lawn is not wet.” With no models remaining, a reasoner might answer randomly or reject the inference, despite its logical validity.

## Conditional Reasoning and Example Knowledge

In relation to a given conditional, we use the term “example” to refer to a specific case that instantiates one of the possibilities relevant to the conditional (Table 2). Whereas mental models can represent possibilities in the abstract, examples are necessarily specific. For instance, a tarp covering the lawn during rain is an example instantiating the possibility that it is raining but the lawn is not wet. Further, whereas mental models specify which possibility they represent, examples need not do so. For instance, one can reason about a tarp covering the lawn during rain without knowing that this would keep the lawn dry. Of course, the latter fact is likely obvious to nearly everyone, but relations between examples and possibilities are not always so obvious, a point to which we return in Study 4.

Despite the above differences, evidence suggests a central role for examples in the construction of mental models.Two types of examples have received considerable attention in the psychology of reasoning: *alternatives*, which instantiate (*p* false, *q* true), and *disablers*, which instantiate (*p* true, *q* false). For instance, for “If it is raining, then the lawn is wet,” a sprinkler being on is an alternative, and a tarp covering the lawn during rain is a disabler. The availability of alternatives and disablers affect endorsement of conditional inferences: People endorse MP and MT less for conditionals with many disablers, and endorse AC and DA less for conditionals with many alternatives (Brisson & Markovits, 2020; Cummins et al., 1991; De Neys et al., 2002; de Neys et al., 2003).

Such findings are consistent with the concept of “modulation” in the theory of mental models—a process by which knowledge about the content of an inference facilitates or hinders construction of models (Johnson-Laird & Byrne, 2002). Specifically, knowledge of disablers should elicit models of (*p* true, *q* false). Such models are inconsistent with MP and MT (Table 1), so knowledge of disablers should reduce endorsement of these inferences—for instance, knowledge about tarps should reduce endorsement of “If it’s raining, then the lawn is wet; the lawn is not wet; therefore, it’s not raining.” Analogously, knowledge of alternatives should facilitate construction of models of (*p* false, *q* true). Such models are inconsistent with AC and DA (Table 1), so knowledge of alternatives should reduce endorsement of these inferences—for instance, knowledge about sprinklers should reduce endorsement of “If it’s raining, then the lawn is wet; the lawn is wet; therefore, it’s raining.”

The research cited above has examined how item-level differences in availability of relevant examples impacts conditional reasoning. However, individual differences in knowledge of relevant examples ought to impact conditional reasoning analogously. For instance, when presented “If it is raining, then the lawn is wet,” some individuals may think of a sprinkler, while others may not think of any alternative cause for the lawn being wet. The former individuals should be more likely to reject AC and DA inferences involving the above conditional, for the reasons described above. Similarly, some individuals and not others may think of disablers for a given conditional. The former group should be more likely to reject MP and MT inferences involving the conditional, for the reasons given above.

Although the above hypothesis is consistent with prior theory and research on conditional reasoning, evidence for it is scant, in part because of the tasks that are typically used to assess individual differences in conditional reasoning. These tasks often involve abstract content for which no one has relevant knowledge (e.g., Evans et al., 2007) or everyday content for which the relevant examples are common knowledge (e.g., Verschueren et al., 2004). In both cases, there is little variation among individuals in knowledge of relevant examples, and therefore little opportunity for such variation to affect conditional reasoning. However, in domains where example knowledge varies more among individuals, such variation should contribute to differences in conditional reasoning. One such domain is math.

## Conditional Reasoning and Example Knowledge in Math

Math is a suitable domain for investigating the role of example knowledge in conditional reasoning for at least two reasons. One reason is that in math, individuals vary substantially with respect to both knowledge of examples (Alcock & Inglis, 2008; Lynch & Lockwood, 2019; Zaslavsky & Peled, 1996) and conditional reasoning performance (Datsogianni et al., 2020; A. J. Stylianides et al., 2004). The presence of natural variation in both of these variables makes it possible to study whether and how they are related at the level of individual differences.

The second reason is that conditional reasoning is particularly important in math. Math students are expected to produce and evaluate deductive arguments (CCSSI, 2010), a norm that is reinforced by practices such as argumentation and formal proof (Knuth, 2002; Ross, 1998; Wu, 1996). Conditionals are pervasive in these contexts (Dawkins & Norton, 2022), where they are used to express definitions (“If a number is a multiple of 2, then it is even”), axioms (“If *a* = *b*, then *b* = *a*”), and theorems (“If *a* and *b* are divisible by *c*, then *a*+*b* is divisible by *c*”). Accordingly, there is substantial research on conditional reasoning in math education (Attridge et al., 2015; Hoyles & Küchemann, 2002; Inglis & Simpson, 2009; A. J. Stylianides et al., 2004). However, this research has generally adopted a domain-specific perspective, leaving it unclear which aspects of conditional reasoning in math (if any) can be explained in terms of domain-general mechanisms. On the other hand, putatively domain-general theories of reasoning such as the theory of mental models have seldom been tested in the domain of math (for exceptions, Datsogianni et al., 2020; G. J. Stylianides & Stylianides, 2008), leaving it unclear how well these theories can explain conditional reasoning in that domain.

To address this knowledge gap, the present study tested the hypothesis outlined in the previous section—a hypothesis inspired by the theory of mental models—using conditionals expressing basic algebra facts, such as “If *m* and *n* are both even, then *m*+*n* is even” (Table 1). Based on the rationale given in the previous section, we expected that reasoning about these conditionals would depend on relevant knowledge—that is, knowledge of algebra. Further, we expected knowledge of examples to play a central role. In this context, examples are ways to instantiate algebraic variables. For instance, with respect to the conditional “If *m* and *n* are both even, then *m*+*n* is even,” *m*=4, *n*=2 is an example of (*p* true, *q* true); *m*=3, *n*=2 is an example of (*p* false, *q* false); and *m*=3, *n*=5 is an example of (*p* false, *q* true)—that is, an alternative. For true conditionals in math, examples of (*p* true, *q* false)—that is, disablers—do not exist.

# Study 1

Study 1 investigated relations between individual differences in conditional reasoning and knowledge in the domain of algebra. We refer to these constructs as “algebra conditional reasoning” and “algebra knowledge.” We distinguish algebra conditional reasoning from conditional reasoning with everyday content (“everyday conditional reasoning”), in which relevant knowledge exists but algebra knowledge specifically is not relevant, and conditional reasoning with imaginary categories (“abstract conditional reasoning”), for which relevant knowledge does not exist. For the reasons outlined in the Introduction, we predicted:

**Prediction 1.1.** Individual differences in algebra knowledge predict algebra conditional reasoning when controlling for everyday and abstract conditional reasoning.

Study 1 represents a re-analysis of data that were collected for other purposes and have not been reported before. The study was preregistered on [Open Science Framework](https://osf.io/n7vrp), but the analyses reported below are not the preregistered ones. Deviations from the preregistration and results of the preregistered analyses are detailed in the Supplementary Materials. Materials, data, and analysis code for this and all subsequent studies are available on [Open Science Framework](https://osf.io/cy59g/files/osfstorage).

## Method

### Participants

Participants were 84 undergraduate students (37 male, 45 female, 1 other) at Florida State University who participated for psychology course credit. No participants correctly answered fewer than 75% of MP items on the conditional inference tasks, so none were excluded.

### Tasks and Materials

**Conditional Inference Task.** On each trial, participants were shown the conditional and categorical premises of a conditional inference, numbered 1 and 2 respectively. They were asked, “If statements 1 and 2 are true, which of the following must be true?” followed by three answer choices. The first two choices were the standard conclusion of the inference and the negation of that conclusion, with the positively phrased choice appearing first (e.g., for both MT and AC items involving a conditional “If *p*, then *q*,” the first two choices would be *p* and Not-*p* in that order); the third choice was “Neither of the above.”After a choice was selected, participants were asked to estimate the probability that the first choice was true, given that statements 1 and 2 were true; results from this question are reported in the Supplementary Materials.[[2]](#footnote-3)

There were four versions of this task, which involved real algebra content, real everyday content, imaginary algebra content, and imaginary everyday content. The real algebra and real everyday versions served as measures of algebra conditional reasoning and everyday conditional reasoning respectively; the two imaginary versions served as measures of abstract conditional reasoning. For each version, stimuli were created using three conditionals, with one trial per inference form (MP, MT, AC, DA) per conditional, for a total of 12 trials. On MP and MT trials, the standard conclusions were counted as correct. On AC and DA trials, “Neither of the above” was counted as correct. Cronbach’s alphas[[3]](#footnote-4) in this sample were .64, .77, .73, and .80 for the real algebra, real everyday, imaginary algebra, and imaginary everyday versions of the task.

The real algebra version involved the following conditionals: “If *m* and *n* are both even, then *m*+*n* is even”; “If *m* is a multiple of 6, then 3*m* is a multiple of 6”; and “If *n* is greater than 3, then *n*2 is greater than 3.” These conditionals are true and their converses are false. Before the trials, participants were told that they would see questions about integers. They were reminded that integers are numbers like 0, 1, 2, -1, -2, and so on, and were told to assume that variables like *m* and *n* represented integers. Next, the task format was described, and participants were instructed to “answer all these questions based only on the statements provided. Don’t answer based on other things you know about math.”

The real everyday version involved causal conditionals about everyday events, such as “If it is raining outside, then the lawn is wet.” Before the trials, participants were told that they would see questions involving everyday events. Next, the task format was described, and participants were instructed to “answer all these questions based only on the statements provided. Don’t answer based on other things you know about the world.”

The imaginary algebra version involved conditionals about fictional numerical categories, such as “If *b* is a sun number, then *b*2 is a moon number.” Before the trials, participants were told that mathematicians had proven some facts about three new kinds of numbers—sun, moon, and star numbers—and that they would be asked to reason about those facts. Other instructions were as in the real algebra version.

The imaginary everyday version involved conditionals about a fictional type of fruit called wumpa fruit, such as “If a wumpa is round, then its seeds are green.” Before the trials, participants were told that scientists had discovered some facts about a new kind of fruit called wumpa fruit, which comes in round, spiky, and tubular varieties, and that they would be asked to reason about those facts. Other instructions were as in the real everyday version.

**Algebra Word Problem Task.** This task served as a measure of general algebra knowledge. It consisted of 16 multiple choice items released from the math portions of the 8th and 12th grade versions of the National Assessment of Educational Progress (NAEP; US Department of Education, Institute of Education Sciences, 2019). The items were algebra word problems that required reasoning—for example, “If *n* represents an even number greater than 2, what is the next larger even number?” Cronbach’s alpha in this sample was .81.

### Procedure

The conditional inference task was presented first, followed by the algebra word problem task. In the conditional inference task, the two algebra versions were presented either before or after the two everyday versions, and the real versions were presented either before or after the imaginary versions for both math and everyday, with the four possible orders being counterbalanced among participants. Within each version, trials were blocked by conditional; order of blocks and of trials within each block were randomized. The tasks were presented on a desktop computer via Qualtrics, and were administered in person. Average session length excluding outliers was 32 minutes.

## Results

Descriptives and correlations among measures are reported in the Appendix (Table A1).

### Prediction 1.1. Individual differences in algebra knowledge predict algebra conditional reasoning when controlling for everyday and abstract conditional reasoning.

Prediction 1.1 was tested using two linear regression models with accuracy on the real algebra version of the conditional inference task as the dependent variable. In model 1, predictors were accuracies on the real everyday version of the conditional inference task and the algebra word problem task. Effects of both predictors were significant (Table 3). In model 2, accuracies on the conditional inference task with imaginary algebra and imaginary everyday content were added as predictors. Algebra word problem accuracy remained a significant predictor (Table 3), consistent with Prediction 1.1. Among the three conditional inference predictors in model 2, only accuracy on the imaginary algebra version of the conditional inference task was significant.

*Table 3*. Linear regressions of conditional inference accuracy with real algebra content (Study 1).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 1 | | Model 2 | |
| Predictor | β (SE) | *p* | β (SE) | *p* |
| Conditional Inference Accuracy – Real Everyday | .46 (.09) | < .001 | .10 (.09) | .28 |
| Conditional Inference Accuracy – Imaginary Algebra |  |  | .52 (.10) | < .001 |
| Conditional Inference Accuracy – Imaginary Everyday |  |  | .09 (.11) | .40 |
| Algebra Word Problem Accuracy | .38 (.09) | < .001 | .26 (.07) | < .001 |
| Fit Statistics | *R*2 = .47, *p* < .001 | | *R*2 = .65, *p* < .001  Δ*R*2 = .18, *p* < .001 | |

## Discussion

The hypothesis that individual differences in conditional reasoning in a domain depend in part on domain-specific knowledge implied that algebra knowledge should predict algebra conditional reasoning, even when controlling for everyday and abstract conditional reasoning. Results were consistent with this prediction. This outcome suggests that participants relied on algebra knowledge to do algebra conditional reasoning, despite having been told not to do so.

Algebra knowledge is a complex construct that includes many specific types of knowledge. Thus, the findings of Study 1 raise the question, what are the “active ingredients” of algebra knowledge that account for its relation to algebra conditional reasoning? A possible answer is: knowledge of examples, which may contribute to conditional reasoning by supporting construction of models representing logical possibilities. This hypothesis was investigated in Studies 2 and 3. In those studies, we use the term “algebra example knowledge” to refer to knowledge of examples that are relevant to conditional statements in algebra, as distinct from other aspects of algebra knowledge.

A secondary finding of Study 1 involved relations among the different versions of the conditional inference task. Interestingly, accuracy on the real algebra version of that task was uniquely predicted by accuracy on only one of the other three versions of the task—the imaginary algebra version. Because participants could not have had prior knowledge about the imaginary algebra categories used in that task, the unique relation between the real and imaginary algebra versions presumably reflects shared reliance on some domain-specific competence other than knowledge. One candidate is the ability to construct and maintain mental models of possibilities involving algebraic variables (e.g., “*m* and *n* are both even and *m*+*n* is even,” “*b* is a sun number and *b*2 is a moon number”). Differences in this ability may have been at least partially independent of algebra knowledge and may have affected performance on both algebra versions of the conditional inference task.

# Study 2

Study 2 tested the hypotheses that individuals use algebra example knowledge to support construction of models when engaging in conditional reasoning about algebra, and that such uses of examples provide a mechanism by which algebra knowledge affects algebra conditional reasoning. Participants were asked to think aloud while performing a conditional inference task with algebra content, and their verbal protocols were coded according to whether examples were mentioned. The above hypotheses suggested three predictions:

**Prediction 2.1.** Individuals refer to examples during conditional reasoning about algebra.

**Prediction 2.2.** Referring to examples during conditional reasoning about algebra predicts algebra conditional reasoning accuracy when controlling for algebra knowledge and general conditional reasoning.

**Prediction 2.3.** Referring to examples during conditional reasoning about algebra partially or fully mediates the relation between algebra knowledge and algebra conditional reasoning accuracy.

Like Study 1, Study 2 represents a re-analysis of data originally collected for other purposes and not previously reported. The study was preregistered on [Open Science Framework](https://osf.io/68vkj/?view_only=9e6b9123d89c4929b19e8394cc2bd4fe), but the analyses reported below are not the preregistered ones. Deviations from the preregistration and results from the preregistered analyses are reported in the Supplementary Materials.

## Method

### Participants

Participants were 99 undergraduate students (22 male, 75 female, 2 other) at Florida State University who participated for psychology course credit. Five additional participants were excluded, three because they correctly answered fewer than 75% of MP items on the conditional inference tasks, and two because their audio recordings were lost due to experimenter error.

### Tasks and Materials

**Conditional Inference Task.** This task was as in Study 1 except for the following changes. (1) Only the real algebra and real everyday versions were presented. (2) Before the first trial involving a given conditional, participants were asked “Do you think the following statement is true?” followed by the conditional, with answer choices “Yes,” “Maybe,” and “No.” This question was intended to encourage participants to process the meaning of the conditional before evaluating inferences involving it. Whichever choice was selected, this instruction appeared: “The following pages will ask you some questions about the above statement. While answering those questions, please assume the above statement is true.” The four trials involving that conditional were then presented. (3) The third answer choice on these trials was changed from “Neither of the above” to “Neither of the above must be true.” (4) Participants were not asked to estimate probabilities of conclusions. Cronbach’s alphas in this sample were .50 and .65 for the tasks with algebra and everyday content, respectively.

**Algebra Word Problem Task.** This task was identical to the one used in Study 1. Cronbach’s alpha in this sample was .71.

### Procedure

Before beginning the main tasks, participants were trained to think aloud using an adapted version of the script developed by Fox et al. (2011). They were instructed to read each item aloud and then say aloud all thoughts they had while working on it. They were told not to explain their thoughts, because explaining (as opposed to thinking aloud) has been shown to improve performance on problem solving tasks (Fox et al., 2011). Participants were given several simple problems unrelated to the subsequent tasks to practice thinking aloud.

Next, participants completed the two versions of the conditional inference task in random order while thinking aloud. They were then told to stop thinking aloud and completed the algebra word problem task. As in Study 1, within each version of the conditional inference task, trials were blocked by conditional, and order of blocks and of trials within each block were randomized. Also as in Study 1, all tasks were presented on a desktop computer via Qualtrics, and were administered in person. Session length excluding outliers averaged 54 minutes.

### Coding

Think aloud protocols from the algebra version of the conditional inference task were coded to indicate whether examples were mentioned (1) or not (0). Code 1 was assigned if the participant referred to specific numbers, or specific features of numbers, that could instantiate either or both of the algebraic variables in the conditional premise. To illustrate what is meant by “specific features of numbers,” for the conditional “If *m* and *n* are both even, then *m*+*n* is even,” “*m* and *n* could both be odd” and “*m* could be even and *n* could be odd” would be coded 1 because these statements describe specific ways in which the conditional’s antecedent could be false. However, numerical features that appeared in the conditional premise were excluded for this code; for instance, “*m* and *n* could both be even” would not be coded 1, because it merely restates the conditional’s antecedent. Protocols were coded independently by two coders, who agreed on 96% of trials. Disagreements were resolved through discussion.

## Results

Descriptives and correlations among measures are reported in the Appendix (Table A2).

### Prediction 2.1. Individuals refer to examples during conditional reasoning about algebra.

Consistent with Prediction 2.1, participants mentioned examples on 21.6% (SD = 22.7%, min = 0.0%, max = 91.7%) of trials on the algebra version of the conditional inference task. The frequency with which they did so varied substantially by inference form (MP: 8.8%, MT: 19.5%, AC: 22.6%, DA: 35.4%; *F*(3, 294) = 31.9, *p* < .001, = .102). Table 4 displays protocols from trials involving the two forms for which examples were mentioned most: AC (protocols 1-4) and DA (protocols 5-8). Four protocols are shown for each form, one for each combination of examples being mentioned or not and the participant selecting the correct answer or not.

*Table 4*. Think-aloud protocols from algebra version of conditional inference task (Study 2).

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Protocol | Examples? | Correct? |
| Form: AC; Premises: “1. If *m* and *n* are both even, then *m*+*n* is even. 2. *m*+*n* is even.” | | | |
| 1 | Neither of the above must be true because ***m* and *n* could be like seven plus three and that could equal ten**. | Yes | Yes |
| 2 | Are not both even. So... That's not true. **Three, two, that’s odd. Six and three, that’s odd.** So both of them have to be even to get an even number. I think. | Yes | No |
| 3 | *m* and— *m* and *n* must both be even because… Actually it doesn't have to be either, so I'm gonna go with neither of the above must be true. | No | Yes |
| 4 | *m* and *n* are both even, because they-- because-- if *m* plus *n* is even, then both of them must be even. | No | No |
| Form: DA; Premises: “1. If *m* and *n* are both even, then *m*+*n* is even. 2. *m* and *n* are not both even.” | | | |
| 5 | Neither of the above must be true cause **it could be an odd and even number**, then it wouldn't be even, or **it could be two odd numbers** and then it would be even. | Yes | Yes |
| 6 | *m* plus *n* is not even since **two plus three we get five**, and that is not an even number. | Yes | No |
| 7 | Well, if they're not both even, any number can occur. | No | Yes |
| 8 | Because I'm told that the if statement is not true then I know that the n— then statement is not true. So *m* plus *n* is not even. | No | No |

*Note*. Statements that justified protocols being coded as mentioning examples are **bolded**, and statements indicating which answer participants selected are underlined, where applicable.

### Prediction 2.2. Referring to examples during conditional reasoning about algebra predicts algebra conditional reasoning accuracy when controlling for algebra knowledge and general conditional reasoning.

Accuracy on the algebra version of the conditional inference task was analyzed with two linear regression models. In model 1, predictors were accuracies on the everyday version of the conditional inference task and the algebra word problem task (i.e., the same predictors as model 1 in Study 1). Both predictors were significant (Table 5). In model 2, frequency of mentioning examples in the math version of the conditional inference task was added as a predictor. This predictor was significant, and the other two remained significant as well (Table 5).

*Table 5*. Linear regressions of conditional inference accuracy with algebra content (Study 2).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 1 | | Model 2 | |
| Predictor | β (SE) | *p* | β (SE) | *p* |
| Conditional Inference Accuracy – Everyday | .84 (.05) | < .001 | .82 (.05) | < .001 |
| Algebra Word Problem Accuracy | .14 (.05) | .005 | .12 (.05) | .014 |
| Frequency of Mentioning Examples |  |  | .11 (.05) | .018 |
| Fit Statistics | *R*2 = .79, *p* < .001 | | *R*2 = .81, *p* < .001  Δ*R*2 = .012, *p* = .018 | |

### Prediction 2.3. Referring to examples during conditional reasoning about algebra partially or fully mediates the relation between algebra knowledge and algebra conditional reasoning accuracy.

The fact that algebra word problem accuracy remained a significant predictor in model 2 precludes full mediation but leaves open the possibility of partial mediation. To test for this possibility, mediation analysis was conducted with algebra word problem accuracy as the independent variable, frequency of mentioning examples in the math version of the conditional inference task as the mediator, and accuracy on the algebra version of the conditional inference task as the dependent variable. Accuracy on the everyday version of the conditional inference task was included as a control variable. Bootstrapping was used to generate 95% confidence intervals (95% CIs) for all effects, with 5000 bootstrap samples, using the bias corrected and accelerated procedure (Carpenter & Bithell, 2000). All effects were standardized.

The total effect of algebra word problem accuracy on conditional inference with math was .14 (95% CI = [.03, .23]), the mediated effect via frequency of mentioning examples accuracy was .019 (95% CI = [-.0013, .058]), and the direct effect when controlling for frequency of mentioning examples was .12 (95% CI = [.017, .21]). Because the 95% CI of the mediated effect included zero, Prediction 2.3 was not supported.

## Discussion

Study 2 replicated Study 1’s central finding: Algebra knowledge predicted algebra conditional reasoning when controlling for everyday conditional reasoning. Further, participants spontaneously referred to examples when reasoning about conditionals involving algebra, despite having been instructed not to rely on prior knowledge of math during this task. Further, doing so more often was associated with more accurate conditional reasoning about algebra when controlling for both everyday conditional reasoning and algebra knowledge, consistent with the hypothesis that algebra example knowledge contributes to algebra conditional reasoning.

Use of examples during conditional reasoning about algebra was assessed using think-aloud protocols, an approach that has both strengths and limitations. A strength is that think-aloud protocols permitted uses of examples to be observed rather than inferred from indirect evidence. A limitation is that the participants may have had example knowledge that they used during reasoning, but did not overtly display while thinking aloud. This possibility could explain why frequency of mentioning examples did not mediate the effect of algebra knowledge on algebra conditional reasoning, contrary to our prediction. For these reasons, in Study 3, we assessed algebra example knowledge with a separate task.

# Study 3

The example generation task was created to assess algebra example knowledge. Each trial consisted of four prompts describing the four possibilities relevant to a particular algebra conditional, which itself was not shown (Table 6). The task was to generate an example that would make each statement true or indicate that no such example exists.

*Table 6*. Prompts and possible correct responses for one trial of the example generation task, corresponding to the conditional “If *m* and *n* are both even, then *m*+*n* is even.”

|  |  |  |
| --- | --- | --- |
| Possibility | Prompt | Correct Response |
| (*p* true, *q* true) | *m* and *n* are both even and *m*+*n* is even | *m* = 4, *n* = 2 |
| (*p* true, *q* false) | *m* and *n* are both even and *m*+*n* is **not** even | none |
| (*p* false, *q* true) | *m* and *n* are **not** both even and *m*+*n* is even | *m* = 3, *n* = 1 |
| (*p* false, *q* false) | *m* and *n* are **not** both even and *m*+*n* is **not** even | *m* = 3, *n* = 2 |

Besides the example generation task, participants also completed a conditional inference task involving the same algebra conditionals on which the example generation task was based, an everyday conditional inference task, and an algebra word problem task, which again served as our measure of general algebra knowledge. We predicted:

**Prediction 3.1.** Algebra conditional reasoning is correlated with algebra example knowledge, as well as with everyday conditional reasoning and algebra knowledge.

The predicted correlation with algebra example knowledge is analogous to the correlation observed in Study 2 with frequency of mentioning examples, whereas the other predicted correlations were observed in both previous studies.

Based on the hypothesis that reliance on knowledge of examples constitutes a mechanism by which algebra knowledge affects algebra conditional reasoning, we further predicted:

**Prediction 3.2.** Algebra example knowledge predicts algebra conditional reasoning when controlling for everyday conditional reasoning and algebra knowledge.

**Prediction 3.3.** Algebra example knowledge partially or fully mediates the relation between algebra knowledge and algebra conditional reasoning.

A final prediction was derived from the assumption that example knowledge affects conditional reasoning by facilitating construction of models that are consistent with inferences’ premises. This assumption implies that relations between example knowledge and conditional reasoning should be specific to the premises of a given inference. To illustrate, consider this AC inference: “If *m* and *n* are both even, then *m*+*n* is even; *m*+*n* is even; therefore, *m* and *n* are both even.” Specificity to the conditional premise means that evaluation of the inference should be better predicted by knowledge of examples for the possibilities related to that conditional—that is, those in Table 6—than by overall algebra example knowledge. Specificity to the categorical premise means that knowledge of examples for the two possibilities consistent with the categorical premise—in this case, the (*p* true, *q* true) and (*p* false, *q* true) possibilities in Table 6—should be an even better predictor. We therefore predicted:

**Prediction 3.4.** Knowledge of specifically relevant algebra examples predicts algebra conditional reasoning better than less specific measures of algebra example knowledge.

The other less specific measures mentioned in Prediction 3.4 are detailed under “Results.”

Study 3 was preregistered on [Open Science Framework](https://osf.io/fgmpw/?view_only=af56a01f69fb4773a7837750e3fa4426). Except where stated otherwise, the preregistration was followed and reported analyses were preregistered.

## Method

### Participants

Participants were 87 undergraduate students (37 male, 48 female, 2 other), including 43 students at Florida State University who participated for psychology course credit, and 44 residents of the US or UK recruited via Prolific and paid $15 for participation. Participants from these two sources did not differ significantly on any measure in our analyses (*p*s > .56). One additional participant was excluded due to answering correctly fewer than 75% of MP items in the conditional inference tasks.

### Tasks and Materials

Stimuli for the example generation task and the algebra conditional inference task were created based on the conditionals numbered 1 to 6 in Table 7. To ensure a consistent response format, all of these conditionals involved two variables, *m* and *n*. The conditionals all were true, and their converses were false. The conditionals differed with respect to which numerical features and arithmetic operations they involved, characteristics that played no role in Study 3 but did play a role in Study 4, in which the same conditionals were used.

*Table 7*. Algebra Conditionals Used in Studies 3 and 4.

|  |  |  |  |
| --- | --- | --- | --- |
| Number | Conditional | Feature | Operation |
| 1 | If *m* and *n* are both positive, then *m*+*n* is positive. | Sign | Addition |
| 2 | If *m* and *n* are both positive, then *m*×*n* is positive. | Sign | Multiplication |
| 3 | If *m* and *n* are both greater than 2, then *m*+*n* is greater than 4. | Magnitude | Addition |
| 4 | If *m* and *n* are both greater than 2, then *m*×*n* is greater than 4. | Magnitude | Multiplication |
| 5 | If *m* and *n* are both even, then *m*+*n* is even. | Parity | Addition |
| 6 | If *m* and *n* are both even, then *m*×*n* is even. | Parity | Multiplication |
| 7 | If *m* and *n* are both multiples of 3, then *m*+*n* is a multiple of 3. | Multiples | Addition |
| 8 | If *m* and *n* are both multiples of 3, then *m*×*n* is a multiple of 9. | Multiples | Multiplication |

*Note*. Conditionals 1 to 6 were used in Study 3, and conditionals 1 to 8 were used in Study 4.

**Example Generation Task.** This task included six trials, one for each of conditionals 1-6 in Table 7. On each page, participants were shown prompts (Table 6) describing the four possibilities relevant to one of those conditionals. These prompts were shown in a random order. After each prompt appeared two text entry fields for *m* and *n* respectively. Participants were instructed to enter values for *m* and *n* that make each statement true, or to enter “none” if they could not think of any such values. The fact that the conditionals were true ensured that examples of (*p* true, *q* false) did not exist, and the fact that their converses were false ensured that examples of (*p* false, *q* true) did exist. Thus, correct responding entailed entering “none” for (*p* true, *q* false) prompts and valid examples for the other three types of prompts. Cronbach’s alpha in this sample (treating each prompt as a separate item) was .56.

**Conditional Inference Task.** As in Study 2, there were two versions of the conditional inference task—algebra and everyday. The algebra version included 24 trials, one per inference form for each of conditionals 1-6 in Table 7. The everyday version included 12 trials, one per inference form for each of the three real everyday conditionals used in Studies 1 and 2. The format of trials was as in Study 2 except that the third answer choice was “Neither of the above” (as in Study 1) rather than “Neither of the above must be true.” Cronbach’s alphas in this sample were .82 and .66 for the algebra and everyday versions respectively.

**Algebra Word Problem Task.** This task consisted of 8 items selected from among the 16 items used in Studies 1 and 2. The task was shortened to allow more time for the other tasks included in this study. Cronbach’s alpha for the shortened task in this sample was .68.

### Procedure

The example generation task and algebra version of the conditional inference task were presented first. Trials of these tasks were presented in blocks, with each block containing the example generation trials for a given algebra conditional followed by the conditional inference trials for the same conditional. Order of blocks and of trials within each block were randomized. The algebra word problem task was presented next. Last, the everyday version of the conditional inference task was presented, with trials again blocked by conditional with block order and trial order within blocks both randomized. Sessions were conducted on Zoom, with experiment materials presented via Qualtrics. Average session length excluding outliers was 31 minutes.

## Results

Descriptives and correlations among measures are reported in the Appendix (Table A3).

### Prediction 3.1. Algebra conditional reasoning is correlated with algebra example knowledge, as well as with everyday conditional reasoning and algebra knowledge.

As predicted, accuracy on the algebra conditional inference task was correlated with accuracies on the example generation task (*r* = .55, *p* < .001), everyday conditional inference task (*r* = .47, *p* < .001), and algebra word problem task (*r* = .37, *p* < .001).

### Prediction 3.2. Algebra example knowledge predicts algebra conditional reasoning when controlling for everyday conditional reasoning and algebra knowledge.

An exploratory analysis (Table 8, model 1) found that accuracies on the everyday conditional inference task and algebra word problem task each independently predicted accuracy on the algebra conditional inference task, replicating findings of Studies 1 and 2. In our preregistered analysis (Table 8, model 2), when controlling for the above measures, accuracy on the example generation task still predicted accuracy on the algebra conditional inference task.

*Table 8*. Linear regressions of conditional inference accuracy with algebra content (Study 3).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model 1 | | Model 2 | |
| Predictor | β (SE) | *p* | β (SE) | *p* |
| Example Generation Accuracy |  |  | .50 (.09) | < .001 |
| Conditional Inference Accuracy – Everyday | .40 (.09) | < .001 | .43 (.08) | < .001 |
| Algebra Word Problem Accuracy | .29 (.09) | .003 | .06 (.09) | .51 |
| Fit Statistics | *R*2 = .30, *p* < .001 | | *R*2 = .49, *p* < .001  Δ*R*2 = .19, *p* < .001 | |

### Prediction 3.3. Algebra example knowledge partially or fully mediates the relation between algebra knowledge and algebra conditional reasoning.

Mediation analysis was conducted using the procedure described in Study 2. Algebra word problem accuracy was the independent variable, example generation accuracy was the mediator, and accuracy on the algebra conditional inference task was the dependent variable. The total effect of algebra word problem accuracy on algebra conditional inference accuracy was .37 (95% CI = [.19, .55]), the mediated effect via example generation accuracy was .21 (95% CI = [.11, .37]), and the direct effect when controlling for example generation accuracy was .16 (95% CI = [-.037, .34]). Thus, the relation between algebra knowledge and algebra conditional reasoning was fully mediated by algebra example knowledge. In an exploratory analysis, accuracy on the everyday conditional inference task was added as a control variable, and the results were qualitatively unchanged (total effect: .29, 95% CI = [.11, .46]; mediated effect: .23, 95% CI = [.12, .38]; direct effect: .06, 95% CI = [-.14, .23]).

### Prediction 3.4. Knowledge of specifically relevant algebra examples predicts algebra conditional reasoning better than less specific measures of algebra example knowledge.

Prediction 3.4 was tested by comparing four Bayesian mixed logistic regression models with correct or incorrect response on each trial of the algebra conditional inference task as the dependent variable, inference form (with MP as the reference category) as a fixed effect, and participant as a random effect. Each model also included, as a fixed effect, one predictor relating to algebra example knowledge. In model 1, this predictor was overall accuracy on the example generation task. In model 2, it was accuracy on all four prompts of the trial of that task that was based on the same conditional as the current conditional inference trial. In model 3, it was accuracy[[4]](#footnote-5) on the two of those four prompts that involved possibilities consistent with the categorical premise of the current conditional inference trial—that is, for MP trials, (*p* true, *q* true) and (*p* true, *q* false); for MT trials, (*p* true, *q* false) and (*p* false, *q* false); for AC trials, (*p* true, *q* true) and (*p* false, *q* true); and for DA trials, (*p* false, *q* true) and (*p* false, *q* false). Thus, model 1 estimated non-item-specific relations between algebra conditional reasoning and algebra example knowledge, model 2 estimated relations that were specific to the conditional premise only, and model 3 estimated relations that were specific to both premises.

In model 4, the algebra example knowledge predictor was accuracy on one of the four prompts of the relevant example generation trial—(*p* true, *q* false) for MP and MT trials, and (*p* false, *q* true) for AC and DA trials. These prompts correspond to disablers and alternatives, respectively, whose importance has been emphasized previously (Brisson & Markovits, 2020; Cummins et al., 1991; de Neys et al., 2003). However, we expected that knowledge of examples for the other two possibilities—(*p* true, *q* true) and (*p* false, *q* false)—would also be important, because models of these possibilities are needed for correct reasoning. Thus, we predicted that model 3 would fit better than model 4.

The models were fit using *brms* (Bürkner, 2017), a front end for Stan (Carpenter et al., 2017), with default priors. Models were fit via MCMC sampling, with four chains, 40,000 samples per chain, and a burn-in of 4,000 samples per chain. Models were assessed using the Leave-One-Out Information Criterion (LOOIC) and Bayes Factors (BFs), which were estimated using bridge sampling. Convergence properties for the chains did not show any issues, including all values < 1.01 and for LOO-CV, all Pareto *k* < 0.7.

Consistent with Prediction 3.4, model 3 had the lowest LOOIC, indicating the best fit (Table 9). BFs favored model 3 over model 1 (BF = 780,994), model 2 (BF = 13,507), and model 4 (BF = 3,175). BFs over 100 constitute extremely strong evidence (Andraszewicz et al., 2015).

*Table 9*. Fit statistics for Bayesian mixed logistic regression models of correct or incorrect response on each trial of the algebra conditional inference task (Study 3).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Prompts/Trials of the Example Generation Task Used to Calculate Example Knowledge Predictor | LOOIC | LOOIC difference from best model | SE of difference from best model |
| Model 1 | All prompts, all trials | 1561.1 | 49.4 | 17.0 |
| Model 2 | All 4 prompts of the trial based on the current conditional premise | 1533.2 | 21.6 | 13.4 |
| Model 3 | The 2 relevant prompts of the trial based on the current conditional premise | 1511.7 | -- | -- |
| Model 4 | One relevant prompt, either (*p* true, *q* false) or (*p* false, *q* true), of the trial based on the current conditional premise | 1523.1 | 11.4 | 11.0 |

*Note*. See main text for details regarding the example knowledge predictors.

## Discussion

Knowledge of examples relating to algebra conditionals predicted accurate reasoning about those conditionals, paralleling an analogous finding in Study 2. The result in Study 2 was based on overt references to examples while thinking aloud during conditional reasoning, whereas the analogous result in Study 3 was based on a separate measure of algebra example knowledge. The fact that diverse methods yielded similar results adds credence to the underlying hypothesis that example knowledge contributes to conditional reasoning in algebra.

In Study 3, algebra example knowledge fully mediated the relation between algebra knowledge and algebra conditional reasoning, consistent with example knowledge being a pathway by which knowledge impacts conditional reasoning in algebra. Our failure to find an analogous result in Study 2 may reflect the think-aloud-based measure used in that study being a weak proxy for example knowledge, perhaps because participants had example knowledge that they used while reasoning, but did not overtly display while thinking aloud.

A relation at the level of individual differences between algebra example knowledge and algebra conditional reasoning could be driven by other abilities that are related to both. However, two aspects of our findings argue against this possibility. First, the above relation remained when controlling for two potential confounding variables, general algebra knowledge and everyday conditional reasoning, paralleling an analogous finding in Study 2. Second, the relation was stronger for knowledge of examples that were specifically relevant to a particular conditional inference item than for general knowledge of algebra examples. This item-specificity seems difficult to explain if the underlying relation reflects general abilities that are related to both algebra example knowledge and algebra conditional reasoning. Nonetheless, definitively demonstrating a causal relation between algebra example knowledge and algebra conditional reasoning requires experimentally manipulating the former. We did so in Study 4.

# Study 4

The example categorization task was created to improve knowledge of examples relevant to particular algebra conditionals. As in the example generation task, each trial of this task presented prompts describing the four possibilities relevant to a particular conditional, which itself was not shown (Figure 1, right side). Unlike the example generation task, candidate examples were also presented (Figure 1, left side)—two for each of the three possibilities other than (*p* true, *q* false). The task was to classify these examples into the categories defined by the four prompts. Incorrect classifications, if any, were highlighted and had to be corrected before participants could proceed.

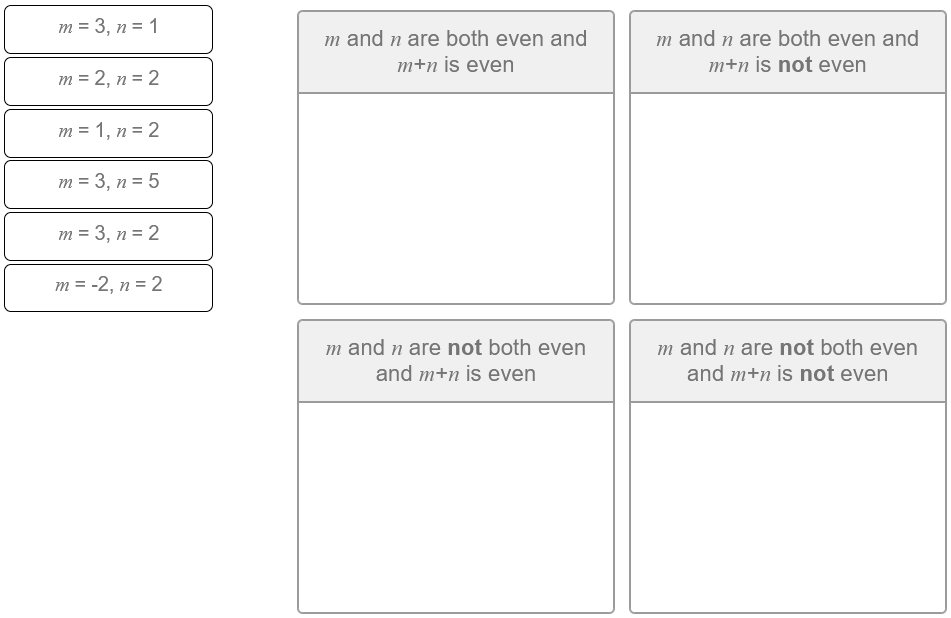


Figure 1. One Trial of the Example Categorization Task (Study 4). Participants would drag each item on the left into one of the four boxes on the right. After all items were so placed, incorrectly placed items were highlighted red. Incorrect placements had to be corrected before continuing.

The example categorization task was designed to address two limitations of participants’ algebra example knowledge that were observed in Study 2: inability to generate examples for one or more of the relevant possibilities, and incorrect categorization of examples. In Study 3, 62% of incorrect responses on the example generation task involved responding “none” to a prompt for which valid examples existed (i.e., a prompt other than (*p* true, *q* false)), suggesting the first limitation above; the other 38% involved responses that were not in fact examples of the possibilities for which they were advanced, suggesting the second limitation above. The example categorization task addressed these limitations by ensuring exposure to two examples for each of the possibilities that exist and by requiring incorrect categorizations to be corrected.

If the relation between algebra example knowledge and algebra conditional reasoning observed in Study 3 reflected a causal effect of the former on the latter, then training via the example categorization task should improve algebra conditional reasoning. Thus, we predicted:

**Prediction 4.1.** Training in categorizing examples that are relevant to an algebra conditional improves subsequent conditional reasoning involving that conditional.

The hypothesized mechanism underlying this prediction was that example categorization training improves algebra example knowledge, which in turn improves algebra conditional reasoning. This mechanism implied a further prediction, namely:

**Prediction 4.2.** The effect of example categorization training described in Prediction 4.1 is partially or fully mediated by algebra example knowledge.

The item-specificity observed in Study 3 of relations between example knowledge and conditional reasoning in algebra suggested that example knowledge could be meaningfully manipulated between items, within participants. Accordingly, in Study 4, participants completed the example categorization training for half of the algebra conditionals on which our stimuli were based, but completed the conditional inference task for all of the conditionals. Prediction 4.1 was tested by comparing accuracies on conditional inference trials involving trained versus untrained conditionals. Further, participants completed the example generation task for all conditionals, which yielded a measure of algebra example knowledge that enabled us to test Prediction 4.2.

Study 4 was preregistered on [Open Science Framework](https://osf.io/dxspg/?view_only=4d23399243254662a3a257a30e5ebbc8). Except where stated otherwise, the preregistration was followed and reported analyses were preregistered.

## Method

### Participants

Participants were 103 undergraduate students (20 male, 82 female, 1 other) at Florida State University who participated for psychology course credit. Fifty participants were assigned to the sign and magnitude group, and 53 participants to the parity and multiples group, as described under “Procedure.” Two more participants were excluded due to answering correctly fewer than 75% of MP items in the conditional inference task.

### Tasks and Materials

Stimuli for the example categorization task, example generation task, and conditional inference task were created based on the 8 algebra conditionals in Table 7. These included the six algebra conditionals used in Study 3 (conditionals 1-6) and two new ones (conditionals 7-8). The two new conditionals, like the original six, are true and have false converses. The eight conditionals were divided into two groups of four each, one involving sign and magnitude (conditionals 1-4) and one involving parity and multiples (conditionals 5-8). Half of the conditionals in each group involved addition and the other half involved multiplication. The everyday version of the conditional inference task was not included in Study 4.

**Example Categorization Task.** On each page, participants were shown a display like the one in Figure 1. The items on the left were shown in random order. The task was to drag each of these items into the appropriate box on the right. Once this was done, incorrectly placed examples—if any—were highlighted red and the message “Some of the items weren’t placed under the correct statement. Please fix that before proceeding” appeared. Participants could proceed once all items were placed correctly.

**Example Generation Task.** This task was as in Study 3, except that there were eight trials—one for each conditional in Table 7. Cronbach’s alpha in this sample was .75.

**Conditional Inference Task.** This task was the same as the algebra version of the conditional inference task in Study 3, except that there were 32 trials—four for each of the eight conditionals. Cronbach’s alpha in this sample was .82.

### Procedure

The three tasks were presented in eight blocks, one for each conditional. For each participant, four conditionals— either those relating to sign and magnitude or those relating to parity and multiples (Table 7)—were randomly selected to be trained, and the other four were designated untrained. Blocks involving trained conditionals included the corresponding example categorization, example generation, and conditional inference trials, in that order. Blocks involving untrained conditionals included only the corresponding example generation and conditional inference trials, in that order. Order of blocks, and of trials within tasks within blocks, were randomized. The tasks were presented on a desktop computer and were administered in person. Session length excluding outliers averaged 25 minutes.

## Results

Descriptives and correlations among measures are reported in the Appendix (Table A4).

### Prediction 4.1. Training in categorizing examples that are relevant to an algebra conditional improves subsequent conditional reasoning involving that conditional.

Conditional inference accuracy was submitted to *ANOVA* with block type (trained or untrained) and form as within-participants factors and training condition (sign and magnitude or parity and multiples) as a between-participants factor. The analysis found effects of block type (*F*(1, 101) = 44.1, *p* < .001, = .022), form (*F*(3, 303) = 110.5, *p* < .001, = .35), and a block type\*form interaction (*F*(3, 303) = 8.7, *p* < .001, = .011; Figure 2). There were no effects involving training condition (*p*s > .15). Due to the interaction, as preregistered, post-hoc contrasts were used to compare trained to untrained blocks separately for each form. Block type did not affect accuracy for MP items (trained: mean = 97.1%, *SD* = 8.8%; untrained: mean = 97.5%, *SD* = 7.4%; *p* = .86), but for all other forms, accuracy was higher on trained than untrained blocks (MT: mean = 86.7%, *SD* = 21.2% vs. mean = 80.3%, *SD* = 21.8%, *p* = .008; AC: mean = 53.6%, *SD* = 36.8% vs. mean = 42.5%, *SD* = 37.0%, *p* < .001; DA: mean = 65.0%, *SD* = 33.1% vs. mean = 49.3%, *SD* = 33.3%, *p* < .001), as predicted.

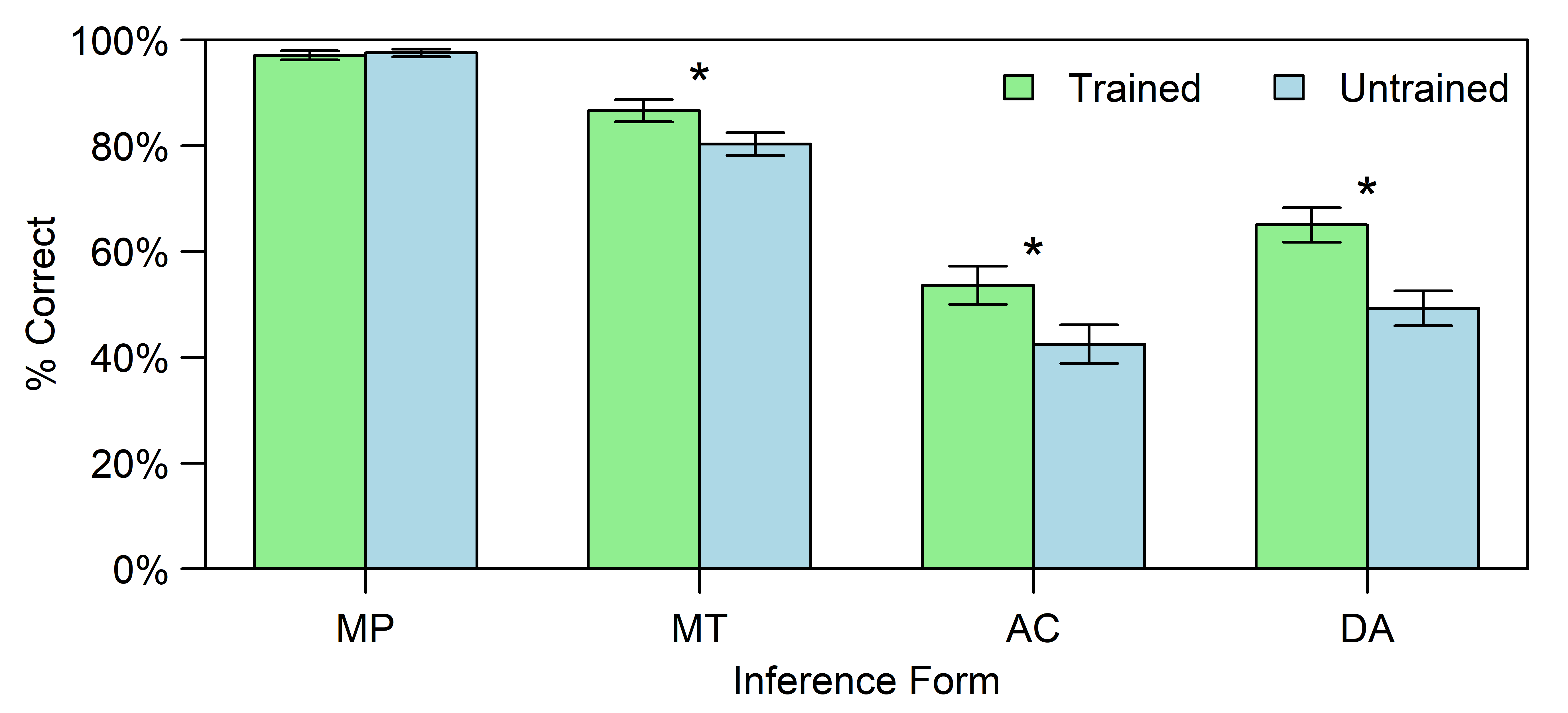


Figure 1. Accuracy by block type and inference form (Study 4). Error bars show standard errors.

### Prediction 4.2. The effect of example categorization training described in Prediction 4.1 is partially or fully mediated by algebra example knowledge.

Prediction 4.2 was tested using mediation analysis, with block type as the independent variable, example generation accuracy as the mediator, and conditional inference accuracy as the dependent variable. To account for block type being a repeated measure, the mediation analysis employed linear mixed models with participant as a random effect. As in Studies 2 and 3, all effects were standardized, and 95% CIs for estimated effects were generated using the bias corrected and accelerated procedure with 5000 bootstrap samples. The total effect of block type on conditional inference accuracy was .082 (95% CI = [.057, .10]), the mediated effect via example generation accuracy was .035 (95% CI = [.022, .053]), and the direct effect when controlling for example generation accuracy was .047 (95% CI = [.021, .072]). Thus, the relation between block type and conditional inference accuracy was partially mediated by example generation accuracy.

# Discussion

Consistent with the hypothesis that example knowledge causally affects conditional reasoning in algebra, example categorization training increased reasoning accuracy with trained conditionals relative to untrained conditionals, and this effect was partially mediated by example generation accuracy. The fact that the mediation was partial rather than full suggests that knowledge of examples may not be the only pathway by which the training affected subsequent conditional reasoning. A possible explanation is suggested by the fact that the prompts of the example categorization task included phrases corresponding to the antecedents (e.g., “*m* and *n* are both even”) and consequents (e.g., “*m*+*n* is even”) of the conditionals that appeared in the conditional inference task. Thus, although the conditionals themselves did not appear in the example categorization task, that task may nonetheless have improved participants’ ability to encode and understand those conditionals, leading to part of the improvement in conditional reasoning accuracy that was observed.

# General Discussion

Human deductive reasoning about meaningful content is not governed solely by syntactic rules, but is also affected by reasoners’ knowledge of the content (e.g., Cheng & Holyoak, 1985; Cummins et al., 1991; Thompson, 1994). However, theories of reasoning have not reached consensus regarding what types of knowledge impact deductive reasoning. We hypothesized that example knowledge, defined as the ability to generate and categorize specific examples of general possibilities, plays a central role in this context. Consistent with this hypothesis, individuals spontaneously generated examples during conditional reasoning (Study 2). Further, individual differences in conditional reasoning accuracy were correlated with frequency of generating examples during conditional reasoning (Study 2) and with performance on a separate assessment of example knowledge (Study 3); these relations remained when controlling for potential confounding variables. Finally, training designed to increase knowledge of examples relevant to particular conditionals improved reasoning about inferences involving those conditionals, compared to inferences involving other untrained conditionals (Study 4). Below, we relate these findings to major theoretical perspectives in the psychology of reasoning.

## Mental models perspective on the present findings

Prior research about effects of example availability on conditional reasoning have been interpreted within the general framework of the theory of mental models (Johnson-Laird & Byrne, 2002) as suggesting a role for semantic memory (Markovits, 2010). According to this account, retrieval of examples from semantic memory can affect which models are constructed during reasoning and thereby affect which inferences are accepted. Specifically, retrieval of disablers can elicit models of (*p* true, *q* false), which contradict the conclusions of MP and MT inferences, and retrieval of alternatives can elicit models of (*p* false, *q* true), which contradict the conclusions of AC and DA inferences. Thus, disablers should decrease endorsements of MP and MT, and alternatives should decrease endorsements of AC and DA.

The present findings are broadly consistent with the above theoretical account, but suggest at least two refinements to it. First, the findings suggest that conditional reasoning depends on knowledge of examples for all four possibilities relevant to a conditional—that is, not only disablers and alternatives, but also examples of (*p* true, *q* true) and (*p* false, *q* false). Evidence for this conclusion comes from Study 3, in which individual differences in conditional reasoning were better predicted by a measure of example knowledge based on all four relevant possibilities than by one based on knowledge of disablers and alternatives only. This result suggests that construction of models representing (*p* true, *q* true) and (*p* false, *q* false) is not always automatic, but may instead depend on generation of relevant examples, as for (*p* true, *q* false) and (*p* false, *q* true).

A second refinement suggested by the present findings pertains to the component processes involved in the use of example knowledge during conditional reasoning. The theoretical model described by Markovits (2010) emphasizes the role of memory retrieval processes, such that retrieval of an example enables construction of a model for the corresponding possibility, whereas failure to construct such a model would reflect failure to retrieve a corresponding example. Indeed, in Study 3 of the present study, 62% of errors on the example generation task involved failure to generate any example for a possibility for which valid examples existed. However, the remaining 38% of errors on that task involved miscategorization of retrieved examples in relation to the four possibilities relevant to a conditional. This result indicates that categorization of examples can be challenging, and suggests that construction of a model for a given possibility may depend not only on retrieval, but also on accurate categorization, of an appropriate example.

## New paradigm perspective on the present findings

The general idea that deductive reasoning about meaningful content depends in part on relevant knowledge is compatible not only with the theory of mental models, but also with theories in the “new paradigm” (Oaksford & Chater, 2020). These theories construe reasoning as concerned primarily with degrees of belief, which can be represented as probabilities, rather than with binary judgments about possibilities. The new paradigm “focuses on knowledge-rich inference based on content and background knowledge” (Oaksford & Chater, 2020, p. 306) and therefore broadly predicts that differences in knowledge should lead to differences in inference, as found in the present study.

However, how might new paradigm theories account for the more specific finding that knowledge of examples affects conditional reasoning? One possibility involves the idea of sampling, which has been proposed (Oaksford & Chater, 2020; Vance & Oaksford, 2021) as a possible basis for algorithmic implementations of the central ideas of new paradigm theories. For example, in a model of syllogistic reasoning described by Hattori (2016), the premises of a syllogism elicit a probability prototype model (PPM), which represents a probability distribution for the possibilities relevant to the premises. The PPM guides subsequent generation of a sample mental model (SMM): a small set of elements, each of which represents an individual case consistent with one of the aforementioned possibilities. The SMM serves as the basis for generating or evaluating conclusions. Within this framework, it is tempting to assume that elements of the SMM correspond to examples. If so, then knowledge of examples could affect the sampling process that generates the SSM and thereby affect reasoning outcomes.

However, at a conceptual level, SMM elements in Hattori’s (2016) model differ from examples in at least two ways. First, SMM elements are generated specifically to represent particular possibilities relevant to the premises of an inference. In contrast, examples can be generated without knowing in advance which possibility they will instantiate; this can be determined post-hoc, a process we have termed “categorization.” Second, the process of sampling elements to generate the SMM is constrained by the probability distribution of the PPM. In contrast, example generation can be constrained by other goals, such as to determine whether examples of a particular possibility exist regardless of whether it is probable.

The above characteristics of examples are evident in the behavior of professional mathematicians when they deliberately seek counterexamples for conjectures (Alcock & Inglis, 2008; Lakatos, 1976; Lynch et al., 2022). That is, examples are sometimes generated without knowing in advance which possibilities they instantiate, and are sometimes generated deliberately to instantiate low-probability possibilities. It is less certain whether these phenomena appear in the reasoning of individuals who are not expert in math when reasoning about math or other topics. Answering this question could help to assess how well sampling-based models of reasoning could account for the present findings.

## Implications about the nature of individual differences in deductive reasoning

Individuals differ considerably in their ability to generate and evaluate deductive arguments. For instance, in the present study, accuracies on the conditional inference task with real math content ranged from 17% to 100% in Studies 1 to 3 (mean = 59%, SD = 17%). Similarly, accuracies on the conditional inference task with real everyday content ranged from 33% to 100% (mean = 60%, SD = 17%). Understanding the nature of these individual differences is an important goal, considering the importance of deductive reasoning as a component of human reasoning skill and as an educational outcome.

Prior research on individual differences in deductive reasoning has often focused on domain-general aspects of these differences. For example, Janveau-Brennan and Markovits (1999) observed that children’s ability to generate examples relevant to certain conditionals predicted differences in their endorsements of inferences involving other conditionals; because the conditionals used in the two tasks involved different content, Janveau-Brennan and Markovits (1999) attributed the above predictive relation to general cognitive processes, specifically memory retrieval (see also De Neys et al., 2002; Markovits & Quinn, 2002). Along similar lines, several studies have found associations between individual differences in conditional reasoning and general cognitive capacities such as verbal ability (Klaczynski & Daniel, 2005) and intelligence (Evans et al., 2008, 2007).

More recently, researchers have proposed a dual strategy framework for conditional reasoning (Brisson & Markovits, 2020; de Chantal et al., 2020; Markovits et al., 2017). In this framework, conditional reasoning performance depends in part on individuals’ tendencies to employ either probabilistic reasoning strategies, which are consistent with new paradigm theories of reasoning, versus example-based strategies, which are consistent with the theory of mental models. When such differences in strategy use are assessed via an abstract conditional reasoning task, they predict performance on a range of other reasoning tasks even when controlling for differences in general cognitive capacities (Thompson & Markovits, 2021). Thus, these strategic differences are distinct from differences in general cognitive capacities. However, they are still conceptualized by the framework as domain-general in nature.

In contrast, in the present study, individual differences in conditional reasoning accuracy were linked to domain-specific knowledge, a (to our knowledge) novel finding that suggests such differences are at least in part domain-specific in nature. Could these findings instead be explained in terms of domain-general differences, such as cognitive capacities or reasoning strategies? Such an explanation could be viable for Studies 1 and 2. For example, individuals who adopt example-based reasoning strategies would refer to examples more often during reasoning, and might also reason more accurately, which could explain the link found in Study 2 between conditional reasoning accuracy and frequency of reference to examples during conditional reasoning. However, the findings of Studies 3 and 4 are difficult to explain in terms of domain-general differences. The reason is that such differences should have had similar effects on different items of the conditional inference task. In contrast, the effects of example knowledge (Study 3) and example categorization training (Study 4) on conditional reasoning accuracy were item-specific, that is, stronger for knowledge that was relevant rather than irrelevant to a given conditional inference.

If domain-specific knowledge contributes to individual differences in conditional reasoning, do domain-general factors also play a role? The answer is likely yes, as evidenced by the fact that in Studies 1-3, individual differences in everyday and abstract conditional reasoning predicted conditional reasoning about algebra even when controlling for algebra knowledge, including example knowledge. Further, interactions between domain-general and domain-specific differences seem possible. For instance, example knowledge may exert effects on reasoning primarily among individuals who adopt example-based reasoning strategies, and have little effect on reasoning performance among individuals who adopt probabilistic reasoning strategies. If so, effects of example knowledge and reasoning strategies on reasoning performance would be super-additive. Alternatively, individuals who do not normally adopt example-based reasoning strategies might do so nevertheless when reasoning about content for which they have strong relevant knowledge. In this case, effects of example knowledge and reasoning strategies might be compensatory.

## Deductive reasoning in math and other domains

The present findings extend prior research showing effects of examples on conditional reasoning (Brisson & Markovits, 2020; Cummins et al., 1991; De Neys et al., 2002; de Neys et al., 2003; Johnson-Laird & Hasson, 2003) from everyday content to mathematical content. This extension is not trivial, for at least two reasons. First, logic is emphasized in math education, which could increase the influence of logical form on peoples’ evaluations of conditional inferences. If logical form and content knowledge compete in such evaluations, as proposed in the dual-source model of conditional reasoning (Klauer et al., 2010), then increased influence of logical form would imply decreased influence of content knowledge, including example knowledge. Second, conditionals are interpreted probabilistically more often when they involve everyday rather than mathematical content, whereas mathematical conditionals are more often interpreted deterministically (Braithwaite, 2025). If examples are perceived as more relevant under probabilistic than deterministic interpretations of conditionals, then examples might play a greater role in everyday than mathematical conditional reasoning. Nonetheless, the present findings suggest a substantial role for examples in conditional reasoning in math.

We have focused on similarities between mathematical and everyday conditional reasoning, but substantial differences may exist. One possible difference involves the level of abstraction at which reasoning occurs. On the algebra conditional inference task in Study 2, utterances were coded as mentioning examples if they described specific ways in which part of a conditional (e.g., “*m* and *n* are both even”) could be true or false. Utterances met this criterion by referring either to particular numbers, as in “*m* and *n* could be like seven plus three,” or to abstract properties, as in “it could be two odd numbers” (Table 4). Although Studies 3 and 4 focused on the former more specific type of example, the latter more abstract type may play a critical role in math. Consistent with this possibility, Dawkins and Norton (2022) recently argued that an abstract understanding of logic in math emerges from reflection on reasoning at the level of abstract properties (which they call “inferring”) and on connecting such properties to specific examples of them (which they call “populating”). Future research should investigate the cognitive mechanisms that support reasoning at the level of abstract properties and explore whether and to what extent these mechanisms are unique to math.

## Limitations

Participants in this study were adults. Developmental research suggests that children and adolescents are less likely than adults to consider all of the possibilities relevant to a conditional statement (Barrouillet & Lecas, 1999). Such developmental differences might affect the relations between example knowledge and conditional reasoning that were studied here. Further, this study focused on students enrolled in university but not necessarily focused on math. This population was selected because there has been considerable prior research on how examples are used during reasoning by math teachers (e.g., Zaslavsky & Peled, 1996) and mathematicians (e.g., Alcock & Inglis, 2008), but less research on this topic among adults who are not particularly focused on math but may still need math in their studies or careers. Nevertheless, it cannot be assumed that the conclusions generalize to populations with more (or less) expertise in math. Similarly, the findings may not generalize to cultural contexts or countries other than the ones studied here—that is, the US and UK. In particular, the math curricula of some countries, such as China, include explicit instruction in formal rules of logical inference. Such instruction, and other national or cultural differences, could affect whether and to what extent individuals rely on examples during deductive reasoning.

## Conclusion

In math, as in other domains, conditional reasoning is not purely syntactic but instead depends in part on relevant prior knowledge. Knowledge of examples, in particular, plays an important role. Such knowledge varies substantially among individuals, and this variation contributes to individual differences in conditional reasoning performance. Understanding in greater detail the processes underlying the use of examples for reasoning, and how experiences such as mathematical training affect these processes, are important goals for future research.

# References

Alcock, L., & Inglis, M. (2008). Doctoral students’ use of examples in evaluating and proving conjectures. *Educational Studies in Mathematics*, *69*(2), 111–129. https://doi.org/10.1007/s10649-008-9149-x

Andraszewicz, S., Scheibehenne, B., Rieskamp, J., Grasman, R., Verhagen, J., & Wagenmakers, E.-J. (2015). An introduction to Bayesian hypothesis testing for management research. *Journal of Management*, *41*(2), 521–543. https://doi.org/10.1177/0149206314560412

Attridge, N., Doritou, M., & Inglis, M. (2015). The development of reasoning skills during compulsory 16 to 18 mathematics education. *Research in Mathematics Education*, *17*(1), 20–37. https://doi.org/10.1080/14794802.2014.999014

Barrouillet, P., & Lecas, J.-F. (1999). Mental models in conditional reasoning and working memory. *Thinking & Reasoning*, *4*, 289–302. https://doi.org/10.1080/135467899393940

Braine, M. D., & O’Brien, D. P. (1998). *Mental logic*. Psychology Press.

Braithwaite, D. W. (2025). Domain effects on interpretations of general conditionals: The case of mathematics. *Thinking & Reasoning*, *31*(2), 214–236. https://doi.org/10.1080/13546783.2024.2406197

Brisson, J., & Markovits, H. (2020). Reasoning strategies and semantic memory effects in deductive reasoning. *Memory and Cognition*, *48*(6), 920–930. https://doi.org/10.3758/s13421-020-01027-3

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, *17*(4), 391–416. https://doi.org/10.1016/0010-0285(85)90014-3

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*(3), 274–282. https://doi.org/10.3758/BF03211151/METRICS

Datsogianni, A., Sodian, B., Markovits, H., & Ufer, S. (2020). Reasoning with conditionals about everyday and mathematical concepts in primary school. *Frontiers in Psychology*, *11*, 2915. https://doi.org/10.3389/fpsyg.2020.531640

Dawkins, P. C., & Norton, A. (2022). Identifying Mental Actions for Abstracting the Logic of Conditional Statements. *The Journal of Mathematical Behavior*, *66*(100954). https://doi.org/10.1016/j.jmathb.2022.100954

de Chantal, P. L., Newman, I. R., Thompson, V., & Markovits, H. (2020). Who resists belief-biased inferences? The role of individual differences in reasoning strategies, working memory, and attentional focus. *Memory and Cognition*, *48*(4), 655–671. https://doi.org/10.3758/S13421-019-00998-2/TABLES/10

De Neys, W., Schaeken, W., & D’Ydewalle, G. (2002). Causal conditional reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory and Cognition*, *30*(6), 908–920. https://doi.org/10.3758/BF03195776

de Neys, W., Schaeken, W., & d’Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: Every counterexample counts. *Memory & Cognition*, *31*(4), 581–595.

Douven, I., Elqayam, S., Singmann, H., & van Wijnbergen-Huitink, J. (2018). Conditionals and inferential connections: A hypothetical inferential theory. *Cognitive Psychology*, *101*, 50–81. https://doi.org/10.1016/J.COGPSYCH.2017.09.002

Evans, J. S. B. T., Handley, S. J., Neilens, H., & Over, D. (2008). Understanding causal conditionals: A study of individual differences. *Quarterly Journal of Experimental Psychology*, *61*(9), 1291–1297. https://doi.org/10.1080/17470210802027961/ASSET/IMAGES/LARGE/10.1080\_17470210802027961-FIG1.JPEG

Evans, J. St. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, *11*(3), 295–306. https://doi.org/10.3758/BF03196976/METRICS

Evans, J. St. B. T., Handley, S. J., Neilens, H., & Over, D. E. (2007). Thinking about conditionals: A study of individual differences. *Memory & Cognition*, *35*(7), 1772–1784.

Evans, J. St. B. T., & Over, D. (2004). *If*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198525134.001.0001

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*(2), 316–344. https://doi.org/10.1037/a0021663

Gowers, T. (2023). What Makes Mathematicians Believe Unproved Mathematical Statements? *Annals of Mathematics and Philosophy*, *1*(1), 57–110.

Hattori, M. (2016). Probabilistic representation in syllogistic reasoning: A theory to integrate mental models and heuristics. *Cognition*, *157*, 296–320. https://doi.org/10.1016/j.cognition.2016.09.009

Hoyles, C., & Küchemann, D. (2002). Students’ understandings of logical implication. *Educational Studies in Mathematics*, *51*, 193–223. https://doi.org/10.1023/A:1023629608614

Inglis, M., & Simpson, A. (2009). Conditional inference and advanced mathematical study: further evidence. *Educ Stud Math*, *72*, 185–198. https://doi.org/10.1007/s10649-009-9187-z

Janveau-Brennan, G., & Markovits, H. (1999). The development of reasoning with causal conditionals. *Developmental Psychology*, *35*(4), 904–911. https://doi.org/10.1037/0012-1649.35.4.904

Johnson-Laird, P. N., & Byrne, R. M. J. (2002). Conditionals: A theory of meaning, pragmatics, and inference. *Psychological Review*, *109*(4), 646–678. https://doi.org/10.1037/0033-295X.109.4.646

Johnson-Laird, P. N., & Hasson, U. (2003). Counterexamples in sentential reasoning. *Memory and Cognition*, *31*(7), 1105–1113. https://doi.org/10.3758/BF03196131

Johnson-Laird, P. N., Khemlani, S. S., & Goodwin, G. P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, *19*(4), 201–214. https://doi.org/10.1016/j.tics.2015.02.006

Khemlani, S. S., Byrne, R. M. J., & Johnson‐Laird, P. N. (2018). Facts and possibilities: A model‐based theory of sentential reasoning. *Cognitive Science*, *42*(6), 1887–1924. https://doi.org/10.1111/cogs.12634

Klaczynski, P. A., & Daniel, D. B. (2005). Individual differences in conditional reasoning: A dual-process account. *Thinking &amp; Reasoning*, *11*(4), 305–325. https://doi.org/10.1080/13546780442000196

Klauer, K. C., Beller, S., & Hütter, M. (2010). Conditional Reasoning in Context: A Dual-Source Model of Probabilistic Inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 298–323. https://doi.org/10.1037/a0018705

Knuth, E. J. (2002). Proof as a tool for learning mathematics. *The Mathematics Teacher*, *95*(7), 486–490. https://doi.org/doi.org/10.5951/MT.95.7.0486

Lakatos, I. (1976). *Proofs and refutations: The logic of mathematical discovery*. Cambridge University Press.

Lynch, A. G., & Lockwood, E. (2019). A comparison between mathematicians’ and students’ use of examples for conjecturing and proving. *The Journal of Mathematical Behavior*, *53*, 323–338. https://doi.org/10.1016/J.JMATHB.2017.07.004

Lynch, A. G., Lockwood, E., & Ellis, A. B. (2022). Comprehensive example generation: mathematicians’ uses of examples when developing conjectures. *Research in Mathematics Education*. https://doi.org/10.1080/14794802.2022.2156586

Markovits, H. (2010). Semantic memory retrieval, mental models, and the development of conditional inferences in children. In *Cognition and conditionals: Probability and logic in human thinking* (pp. 177–195). Oxford University Press.

Markovits, H., Brisson, J., & de Chantal, P. L. (2017). Logical reasoning versus information processing in the dual-strategy model of reasoning. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(1), 72–80. https://doi.org/10.1037/XLM0000291

Markovits, H., & Quinn, S. (2002). Efficiency of retrieval correlates with “logical” reasoning from causal conditional premises. *Memory and Cognition*, *30*(5), 696–706. https://doi.org/10.3758/BF03196426/METRICS

Morsanyi, K., McCormack, T., & O’Mahony, E. (2018). The link between deductive reasoning and mathematics. *Thinking & Reasoning*, *24*(2), 234–257. https://doi.org/10.1080/13546783.2017.1384760

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards*. Authors.

Oaksford, M., & Chater, N. (2020). New paradigms in the psychology of reasoning. *Annual Review of Psychology*, *71*(1), 305–330. https://doi.org/10.1146/annurev-psych-010419-051132

Oberauer, K. (2006). Reasoning with conditionals: A test of formal models of four theories. *Cognitive Psychology*, *53*(3), 238–283. https://doi.org/10.1016/J.COGPSYCH.2006.04.001

Rips, L. (1994). *The psychology of proof: Deductive reasoning in human thinking*. The MIT Press. https://doi.org/10.7551/mitpress/5680.001.0001

Ross, K. A. (1998). Doing and proving: The place of algorithms and proofs in school mathematics. *The American Mathematical Monthly*, *105*(3), 252–255. https://doi.org/10.2307/2589080

Skovgaard-Olsen, N., Singmann, H., & Klauer, K. C. (2016). The relevance effect and conditionals. *Cognition*, *150*, 26–36. https://doi.org/10.1016/j.cognition.2015.12.017

Stylianides, A. J., Stylianides, G. J., & Philippou, G. N. (2004). Undergraduate students’ understanding of the contraposition equivalence rule in symbolic and verbal contexts. *Educational Studies in Mathematics*, *55*, 133–162. https://doi.org/10.1023/B:EDUC.0000017671.47700.0b

Stylianides, G. J., & Stylianides, A. J. (2008). Proof in School Mathematics: Insights from Psychological Research into Students’ Ability for Deductive Reasoning. *Mathematical Thinking and Learning*, *10*(2), 103–133. https://doi.org/10.1080/10986060701854425

Thompson, V. A. (1994). Interpretational factors in conditional reasoning. *Memory & Cognition*, *22*(6), 742–758. https://doi.org/10.3758/BF03209259

Thompson, V. A., & Markovits, H. (2021). Reasoning strategy vs cognitive capacity as predictors of individual differences in reasoning performance. *Cognition*, *217*, 104866. https://doi.org/10.1016/J.COGNITION.2021.104866

US Department of Education, Institute of Education Sciences, N. C. for E. S. (2019). *National Assessment of Educational Progress Mathematics Assessment*.

Vance, J., & Oaksford, M. (2021). Explaining the implicit negations effect in conditional inference: Experience, probabilities, and contrast sets. *Journal of Experimental Psychology: General*, *150*(2), 354–384. https://doi.org/10.1037/xge0000954.supp

Verschueren, N., Schaeken, W., De Neys, W., & D’ydewalle, G. (2004). The difference between generating counterexamples and using them during reasoning. *Quarterly Journal of Experimental Psychology*, *57A*(7), 1285–1308. https://doi.org/10.1080/02724980343000774

Wu, H. H. (1996). The role of Euclidean geometry in high school. *Journal of Mathematical Behavior*, *15*(3), 221–237. https://doi.org/10.1016/S0732-3123(96)90002-4

Zaslavsky, O., & Peled, I. (1996). Inhibiting Factors in Generating Examples by Mathematics Teachers and Student Teachers: The Case of Binary Operation. *Journal for Research in Mathematics Education*, *27*(1), 67–78. https://doi.org/10.5951/JRESEMATHEDUC.27.1.0067

# Appendix

Table A1. Descriptive Data and Correlations Among Measures (Study 1).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Correlations | | | |
|  | Mean (SD) | (2) | (3) | (4) | (5) |
| (1) Conditional Inference Accuracy – Real Math | .54 (.18) | .58 \*\* | .75 \*\* | .65 \*\* | .53 \*\* |
| (2) Conditional Inference Accuracy – Real Everyday | .58 (.17) |  | .64 \*\* | .67 \*\* | .33 \*\* |
| (3) Conditional Inference Accuracy – Imaginary Math | .53 (.18) |  |  | .73 \*\* | .38 \*\* |
| (4) Conditional Inference Accuracy – Imaginary Everyday | .57 (.18) |  |  |  | .41 \*\* |
| (5) Algebra Word Problem Accuracy | .69 (.22) |  |  |  |  |

*Note*. \* denotes *p* < .05 and \*\* denotes *p* < .01.

Table A2. Descriptive Data and Correlations Among Measures (Study 2).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Correlations | | |
|  | Mean (SD) | (2) | (3) | (4) |
| (1) Conditional Inference Accuracy – Math | .61 (.14) | .88 \*\* | .37 \*\* | .32 \*\* |
| (2) Conditional Inference Accuracy – Everyday | .61 (.17) |  | .28 \*\* | .21 \* |
| (3) Algebra Word Problem Accuracy | .74 (.18) |  |  | .22 \* |
| (4) Conditional Inference Example Mentions – Math | .22 (.23) |  |  |  |

*Note*. \* denotes *p* < .05 and \*\* denotes *p* < .01.

Table A3. Descriptive Data and Correlations Among Measures (Study 3).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | Correlations | | |
|  | Mean (SD) | (2) | (3) | (4) |
| (1) Conditional Inference Accuracy – Math | .62 (.17) | .55 \*\* | .47 \*\* | .37 \*\* |
| (2) Example Generation Accuracy | .80 (.11) |  | .05 | .45 \*\* |
| (3) Conditional Inference Accuracy – Everyday | .59 (.17) |  |  | .21 \* |
| (4) Algebra Word Problem Accuracy | .60 (.26) |  |  |  |

*Note*. \* denotes *p* < .05 and \*\* denotes *p* < .01.

Table A4. Descriptive Data and Correlations Among Measures (Study 4).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | Correlations | | | | |
|  | Mean (SD) | (1a) | (1b) | (2) | (2a) | (2b) |
| (1) Conditional Inference Accuracy | .72 (.15) | .90 \*\* | .90 \*\* | .61 \*\* | .58 \*\* | .49 \*\* |
| (1a) Conditional Inference Accuracy – Sign and Magnitude | .71 (.17) |  | .63 \*\* | .51 \*\* | .61 \*\* | .30 \*\* |
| (1b) Conditional Inference Accuracy – Parity and Multiples | .73 (.17) |  |  | .59 \*\* | .43 \*\* | .59 \*\* |
| (2) Example Generation Accuracy | .88 (.10) |  |  |  | .84 \*\* | .88 \*\* |
| (2a) Example Generation Accuracy – Sign and Magnitude | .89 (.11) |  |  |  |  | .49 \*\* |
| (2b) Example Generation Accuracy – Parity and Multiples | .87 (.13) |  |  |  |  |  |

*Note*. \* denotes *p* < .05 and \*\* denotes *p* < .01.

1. Our use of the term “model” to signify a representation that specifies the truth or falsehood of *p* and *q* corresponds to what Johnson-Laird et al. (2015, p. 2) call a “fully explicit model.” Johnson-Laird et al. (2015) reserve the term “mental model” for models that represent only what is true, not what is false. [↑](#footnote-ref-2)
2. One reason why we did not analyze these probability estimates in the main text is that it is unclear how such estimates could be made for statements like “*m* and *n* are both even” or “*m*+*n* is even,” especially for AC and DA items. More generally, it is not clear whether or how probabilities can be assigned to mathematical statements (see Gowers, 2023 for an interesting discussion of this issue). Clarifying this issue could increase the utility of probabilistic theories of reasoning (e.g., Oaksford & Chater, 2020) for explaining deductive reasoning in math. [↑](#footnote-ref-3)
3. Here and in the following studies, Cronbach’s alphas were often relatively low for one or more versions of the conditional inference task. A likely cause is that accuracies on MP/MT items were often uncorrelated or negatively correlated with accuracies on AC/DA items, which is often found (e.g., Morsanyi et al., 2018). We nevertheless used overall accuracies on the conditional inference tasks in our analyses, because competent conditional reasoning requires both accepting valid inferences (MP and MT) and rejecting invalid ones (AC and DA). [↑](#footnote-ref-4)
4. Our preregistration specified that the algebra example knowledge predictor in models 3 and 4 would be based on whether participants correctly indicated whether examples existed for each prompt, regardless of whether their examples were accurate. Thus, any response other than “none” would be counted as correct for all prompts except the (*p* true, *q* false) one. We elected instead to use accuracy for the analyses reported in the main text because accuracy is likely to be more familiar to most readers. Analyses using the preregistered measure yielded the same pattern of results, and are reported in the Supplementary Materials. [↑](#footnote-ref-5)