Principles Used to Evaluate Mathematical Explanations

Samuel G. B. Johnson[1], Angie M. Johnston[1], Marissa L. Koven[2], & Frank C. Keil[1]
(sgbjohnson@gmail.com, angie.johnston@yale.edu, mlkoven@gmail.com, frank.keil@yale.edu)
[1]Department of Psychology, Yale University, New Haven, CT 06520 USA
[2]Department of Psychology, Emory University, Atlanta, GA 30322 USA

Abstract

Mathematics is critical for making sense of the world. Yet, little is known about how people evaluate mathematical explanations. Here, we use an explanatory reasoning task to investigate the intuitive structure of mathematics. We show that people evaluate arithmetic explanations by building mental proofs over the conceptual structure of intuitive arithmetic, evaluating those proofs using criteria similar to those of professional mathematicians. Specifically, we find that people prefer explanations consistent with the conceptual order of the operations ("9÷3=3 because 3×3=9" rather than "3×3=9 because 9÷3=3"), and corresponding to simpler proofs ("9÷3=3 because 3×3=9" rather than "9÷3=3 because 3+3+3=9"). Implications for mathematics cognition and education are discussed.

Keywords: Mathematics cognition; philosophy of mathematics; explanation; reasoning; concepts and categories.

Introduction

People track statistical regularities and use these regularities to make sense of the world. Some statistical learning abilities emerge early: Infants use statistics to extract complex visual features (Fiser & Aslin, 2002) and form categories (Gómez & Lakusta, 2004). Statistical generalizations are also critical for sense-making in higher cognition. For example, adults and children prefer simpler causal explanations in part because they have higher prior probabilities (Bonawitz & Lombrozo, 2012; Johnson, Valenti, & Keil, 2017; Lombrozo, 2007).

Yet, we also seem to track other truths that do not rely on statistical regularities—*Platonic*, logically necessary regularities such as mathematical truths. From early on, people use mathematical truths to make sense of the world: Even young infants know that if two puppets venture behind a screen, and one comes out, then only one puppet remains behind (Wynn, 1992). Without an understanding of mathematics (i.e., 2–1=1), this event—and many others—would be inexplicable. Mathematical explanation grows even more essential in adulthood, as consumers must account for their spending, programmers must understand the logic of their code, and CEOs must explain their bottom line. For this reason, educators increasingly emphasize the explanatory function of mathematics (Schoenfeld, 1992). For example, the Common Core Standards (2010) state that "one hallmark of mathematical understanding is the ability to justify…*why* a particular mathematical statement is true or where a mathematical rule comes from" (p. 4).

But to what extent, and by what mechanisms, can people track such mathematical regularities? Here, we claim that people use a sophisticated set of mechanisms to evaluate mathematical explanations. We argue that people (1) are sensitive to the conceptual structure of arithmetic, (2) construct mental proofs over this structure, and (3) evaluate those proofs using principles that mirror the history, philosophy, and practice of mathematics.

Just as there are intricate connections among concepts in physics and biology, so are mathematical concepts richly structured (Whitehead & Russell, 1910). For example, geometric facts are grounded in facts about analysis (Bolzano, 1817), and arithmetic facts in set theory (Frege, 1974/1884). More basically, subtraction can be viewed as grounded in addition, multiplication in addition, division in multiplication, and so on (Figure 1; see also Dedekind, 1995/1888; Tao, 2016). Although these concepts need not be viewed asymmetrically, these asymmetries may be psychologically natural. For example, people may follow the principle that more fundamental operations begin with small things and assemble larger things, rather than vice versa. This would make addition more fundamental than subtraction (which breaks larger things into smaller pieces).

We explored the intuitive conceptual structure of mathematics using a simple method—asking people to evaluate mathematical explanations. Consider the explanation "9÷3=3 because 3+3+3=9." In one sense, this is a terrible explanation because it is tautological—both facts are necessarily true and logically equivalent. However, we propose that people are willing to evaluate explanations of this sort, and do so as if constructing a *mental proof* of the explanatory target (here, "9÷3=3") from the putative explanation ("3+3+3=9"), over the conceptual structure in Figure 1. For example, to evaluate "9÷3=3 because 3+3+3=9," one would first derive a multiplication fact ("3×3=9") from the addition fact, and then derive the division fact from that intermediate multiplication fact. We test two principles that people might use for evaluating implicit mental proofs.

First, people may be sensitive to the asymmetric nature of mathematical explanation (Bolzano, 1817; Kitcher, 1975). For example, consider the explanation "4–2=2 because 2+2=4." Although tautological, if this explanation respects a perceived *conceptual order*, it may be seen as superior to an explanation that does not, such as "2+2=4 because 4–2=2." That is, a statement may be explained in terms of a logically equivalent statement, if that explanation helps to highlight the more conceptually primitive facts grounding it.

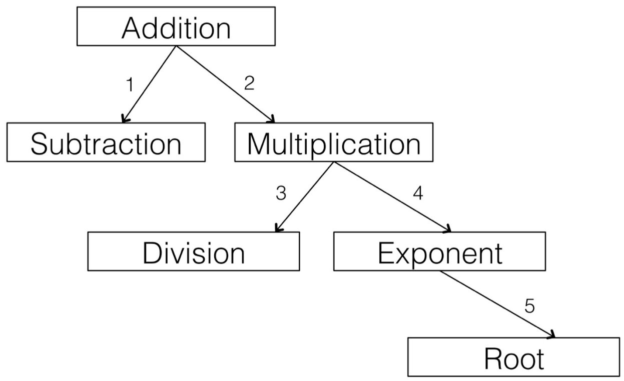

Figure 1: Proposed intuitive structure of arithmetic.

Note. Numbers correspond to the proof rules in Table 1, with forward proof rules flowing in the direction of the arrows and reverse proof rules flowing against the direction of the arrows.

Rule	Input	Output
	Addition/Subtraction Conversion	
1F	$X + Y = Z$	$Z - X = Y$
1R	$Z - X = Y$	$X + Y = Z$
	Addition/Multiplication Conversion	
2F	$\Sigma_Y X = Z$	$X \times Y = Z$
2R	$X \times Y = Z$	$\Sigma_Y X = Z$
	Multiplication/Division Conversion	
3F	$X \times Y = Z$	$Z \div X = Y$
3R	$Z \div X = Y$	$X \times Y = Z$
	Multiplication/Exponent Conversion	
4F	$X \times X = Z$	$X^2 = Z$
4R	$X^2 = Z$	$X \times X = Z$
	Exponent/Root Conversion	
5F	$X^2 = Z$	$\sqrt{Z} = X$
5R	$\sqrt{Z} = X$	$X^2 = Z$

Table 1: Hypothesized rules for mental proofs.

Second, people may prefer explanations that involve fewer steps because such proofs more readily confer understanding (Descartes, 1954/1684; Hardy, 2004/1940; Kitcher, 1983) and are less prone to error (Hume, 1978/1738). For example, "9÷3=3 because 3+3+3=9" might be seen as a worse explanation than "9÷3=3 because 3×3=9," since the proof for the former explanation requires two steps (addition to multiplication, multiplication to division) whereas the latter requires only one step (multiplication to division), even though both proofs proceed in the same conceptual order. We test whether people scale their explanatory judgments to *proof complexity*. If so, this would be evidence not only that people use complexity as a criterion to judge explanatory quality, but also that people spontaneously construct proofs over the conceptual structure depicted in Figure 1.

Our model assumes that people evaluate these explanations by constructing and evaluating a proof of the explanatory target from the base, using the transformation rules given in Table 1. These correspond to the forward (F) and reverse (R) version of each arrow in Figure 1 (see Rips, 1983 for a related idea in propositional reasoning). Proofs are evaluated by assuming a *rule cost* is incurred for applying each rule, and that the total *proof cost* is the sum of the costs of the individual rules invoked in the proof. If people are sensitive to proof complexity, then they should prefer proofs with smaller costs. To capture the idea that people prefer explanations consonant with the conceptual order, our model allows forward and reverse rules to have different costs: We predict that reverse rules carry a higher cost than forward rules. That is, a proof has a higher cost to the extent that it uses *more* rules in general, and more *reverse* rules in particular. Equivalently, short proofs flowing with Figure 1's arrows would correspond to better explanations than long proofs flowing against the arrows (see examples below).

Method

We recruited 97 participants from Amazon Mechanical Turk in exchange for a small payment (50.5% female, M_{age} = 34.0). Participants were excluded from data analysis if they gave inappropriate answers to the check questions (N = 6; see below for details).

Participants rated a series of 30 mathematical explanations. For each explanation, participants were asked "How satisfying do you find this explanation?" on a scale from 0 ("not at all satisfying") to 10 ("very satisfying"). These explanations consisted of all possible pairings of addition, subtraction, multiplication, division, exponent, and root operations, where the constituents were 3s; examples are given in Table 2 in the Appendix. For example, across different blocks, participants completed a pair of multiplication/exponent items:

$3^2 = 9$ because $3 \times 3 = 9$ [*forward*]

$3 \times 3 = 9$ because $3^2 = 9$ [*reverse*]

Because there are 15 ways of pairing these 6 operations with each other, and two orders (forward and reverse), participants completed a total of 30 items. The forward and reverse items were presented in separate blocks, with the order of the items randomized within each block. The order of the blocks was also randomized.

Check questions were included after the test questions to detect participants who were responding randomly. These always included two items for which one of the equations was false (e.g., "4+3=7 because 4+3=2" or "743+259=1,002 because 743+259=713") and two items for which the numbers differed between the two equations (e.g., "26×47=1222 because 678−234=444"). Participants with average answers to these questions that were above the scale midpoint were excluded from data analysis.

Results

Participants were sensitive to both criteria of conceptual order and proof complexity. We first describe the results relative to the qualitative predictions of the model in order to explain how the model works, and then assess the quantitative fit at both the group and individual levels.

Qualitative model predictions. We anticipated that people would penalize explanations to the extent that the most direct proof requires applying a large number of rules (see Tables 1 and 2), and that application of 'reverse' rules would correspond to a greater penalty. These two principles are captured by (1) computing the shortest distance between the two operations in Figure 1, and (2) penalizing the explanation for each arrow along that shortest path, with arrows in the 'reverse' direction receiving a larger penalty (we call this penalty *R*) than arrows in the 'forward' direction (a smaller penalty of *F*).

For example, consider explaining a root formula in terms of division (e.g., "√9=3 because 9÷3=3"). According to our model, people would rate this explanation by producing a mental proof of '√9=3' from '9÷3=3'. As noted in Table 2, this requires the application of three rules: 3R (to derive '3×3=9' from '9÷3=3'), 4F (to derive '3^2=9'), and finally 5F (to derive '√9=3'). Two of these rules are forward and one is backward, so the total penalty is 2*F* + 1*R*—since this is a relatively high penalty, we would expect this explanation to be rated poorly. In contrast, addition would be seen as an excellent explanation of subtraction, because a subtraction formula (e.g., '9–3–3=3') can be derived from addition ('3+3+3=9') using only one forward rule (1F), leading to a penalty of only 1*F*. The penalty scores for several of the explanations are given in Table 2 in the Appendix, along with the rules required to perform these proofs.

This model captures several patterns in the means (Table 3 in the Appendix). First, for each operation, we can consider which explanation was rated highest (i.e., the highest mean in each row of Table 3). For the addition operation, which is not conceptually dependent on any of the other operations, its highest rated explanations were subtraction and multiplication—the closest downstream operations. For both subtraction and multiplication, addition is the highest rated explanation, consistent with the topology of Figure 1, wherein both operations depend directly on addition. Similarly, for explaining division and exponentiation, multiplication is highest rated, consistent with Figure 1, in that both operations depend directly on multiplication. Finally, for roots, exponentiation was seen as the best explanation, again consistent with the direct dependence of roots on exponents.

More generally, our model predicts a central role of multiplication and a peripheral role of subtraction. As Figure 1 shows, multiplication is a central node in the conceptual structure of arithmetic—most roads lead to (or from) multiplication—but subtraction is on the periphery. This prediction is borne out by the data. Multiplication is both the most easily explained operation (i.e., the highest mean in the rightmost column of Table 3) and the operation that explains the most (i.e., the highest mean in the bottom row of Table 3). In contrast, subtraction is least easily explained and explains the least.

Group-level model fitting. We model the results in terms of the sum of the rule costs, shown in Table 2. This analysis assumes that the cost of each rule is determined only by whether it is a forward or reverse rule. Thus, one free parameter *R*/*F* is used, reflecting the extent to which *R* rules were penalized more heavily than *F* rules.

We modeled the explanation ratings in terms of the summed rule costs, where only the *R*/*F* parameter was free to vary. These scores were good predictors of the explanation ratings, $r(28) = -.86$, $p < .001$. The best fitting value for the *R*/*F* parameter was 1.18, indicating that the explanatory cost of applying reverse rules that go against the conceptual grain of mathematics is 18% higher than the explanatory cost of applying forward rules. This supports our conjecture that forward explanations (e.g., explaining subtraction in terms of addition) are preferred to their logically equivalent reverse explanations (e.g., explaining addition in terms of subtraction).

This asymmetry between forward and reverse rules is also evident from looking at the means in Table 3. For example, explanations of subtraction in terms of addition were rated more satisfying than explanations of addition in subtraction, since the former grounds an operation in a more psychologically basic operation whereas the latter does the opposite. Since there are five rules in Table 1, there are five directly reversible pairs, as well as four pairs of operations (addition/division, addition/exponentiation, addition/root, and multiplication/root) that are connected by applying two or more rules in the same direction (see Proof column in Table 2). Averaging across these pairs, the forward explanations were seen as more satisfying than the reverse, $t(90) = 3.90$, $p < .001$.

Individual-level model fitting. Our model also captures individual participants' explanatory judgments. To test the proof complexity factor, we calculated, for each participant, the correlation between the explanatory judgment for each of the 30 items and the number of rules required for that item's proof (i.e., the sum of the *F* and *R* columns in Table 2). This parameter-free model captured a substantial amount of the variance within each participant's response pattern, with a mean correlation of –.46 between number of rules and explanatory judgment (Fisher-transformed to a *z*-score before averaging, and inverse-transformed back to a correlation). Furthermore, almost all participants (95.6%) had a negative correlation, demonstrating that the excellent model fit at the group level is not due to a small subset of participants, but instead generalizes across almost all participants.

Although this parameter-free model is useful in showing that considerable within-subject variability can be explained via proof complexity, it is less useful for testing asymmetries between the forward and backward rules, since this requires estimating the relative penalties associated with each rule. To do so, we conducted a linear regression for each participant, using ten dummy-coded variables to represent whether each of the ten rules figures in each item's proof. For example, for the item explaining division in terms of subtraction, the dummy variables for rules 1R, 2F, and 3F were set to 1, and all others set to 0. For each participant, we calculated the regression weights for each rule, reflecting the relative penalty associated with each rule (thus, all regression weights would be expected to be negative), and these weights (averaged across participants) are depicted in Figure 2.

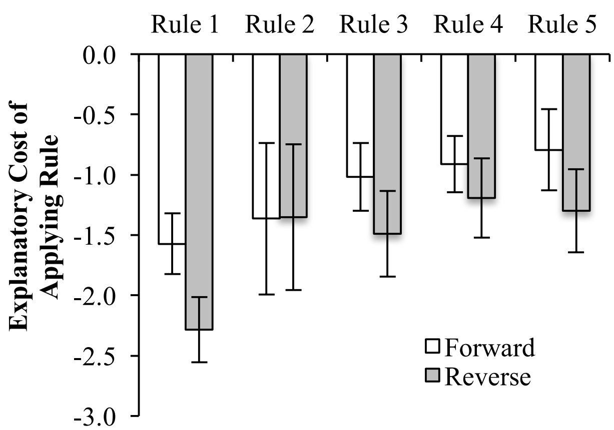

Figure 2: Regression coefficients on each rule.

Note. These coefficients represent the explanatory 'cost' of a given rule appearing in the proof of the explanation. Error bars represent 95% confidence intervals, calculated over participants.

Mirroring the group-level findings, Figure 2 reveals higher costs for reverse rules than for forward rules, leading to more negative regression coefficients for the reverse rules. This was true for rules 1F and 1R (95% CI [0.28, 1.15] for the difference in regression coefficients), rules 3F and 3R (95% CI [0.13, 0.81]), rules 4F and 4R (95% CI [-0.02, 0.58]; marginally significant), and rules 5F and 5R (95% CI [0.17, 0.85]). This difference was not significant for the addition/multiplication rules 2F and 2R (95% CI [-0.41, 0.39]), perhaps because repeated addition of the same addends is uncommon except in the context of multiplication. Overall, these findings are consistent with the best-fitting value of the R/F parameter of 1.18 in the group-level analysis, indicating a higher explanatory cost for reverse rules than for forward rules.

Discussion

Mathematical knowledge is critical for explaining patterns in both the physical and symbolic worlds, and for building an understanding conceptually dependent mathematical facts. Here, we proposed that people evaluate mathematical explanations (e.g., "9–3–3=3 because 3+3+3=9") by building a proof from the explanatory base ("3+3+3=9") to the explanatory target ("9–3–3=3") using a set of transformation rules (e.g., deriving subtraction from addition). Supporting this idea, participants preferred explanations that obeyed the conceptual order of mathematics and which required fewer derivational steps.

Where might these intuitions come from? One possibility is that they are rooted in a more basic understanding of the natural numbers (e.g., Carey, 2009; Dehaene, 1997; Rips, Bloomfield, & Asmuth, 2008) that begins to emerge early in development. For example, addition and subtraction are intimately related to counting, both in development (Rips et al., 2008) and in mathematics (Tao, 2016). This is because the natural numbers are constructed by using the *successor* function (e.g., 9 is the successor to 8). Such psychologically and mathematically primitive mechanisms may underlie later-emerging explanatory intuitions.

Alternatively, could it be possible that people simply parroted explanations as introduced in school? This possibility is unlikely for two reasons. First, multiplication was strongly preferred over subtraction as an explanation. This pattern is consistent with our claims about conceptual structure but conflicts with this alternative account, since subtraction is typically learned before multiplication. Second, we doubt most people have *ever* heard (for example) division explained in terms of addition, exponential, roots, etc., so differences across these explanations must be due to a chaining mechanism of the type we proposed.

Might analogous results hold beyond arithmetic explanations? Indeed, people have a rich intuitive understanding of other mathematical domains such as geometry (Dillon, Huang, & Spelke, 2013), suggesting that people have intuitive theories of Platonic regularities across a variety of domains. Moreover, the proof construction and evaluation principles may be the same used in more general deductive reasoning processes (Rips, 1994; but see Johnson-Laird & Byrne, 1991), in which case our method may generalize. Our studies focused on simple arithmetic operations (e.g., Ashcraft, 1992), but future work could extend this inquiry to other areas of mathematics (such as geometry), populations (such as children or expert mathematicians), or domains (such as dependencies among physics concepts or among mental states) to further map our intuitive theories.

The ontological implications of this work are within the domain of philosophy. For now, we merely contrast two possible views. According to the dominant *Platonist* view (e.g., Frege, 1974/1884), mathematical truths are 'out there' in the world. On the Platonist view, our results reflect aspects of mathematical structure that have been internalized from the world. However, others with *Kantian* views argue that mathematical cognition reflects structure imposed *on* the world by our minds rather than anything intrinsic in the world (Kant, 1998/1781; Mill, 2002/1843; see Lakoff & Núñez, 2000). On the Kantian view, our results reflect the intrinsic structure of our minds themselves, which we impose on the world.

As for the instructional implications of these findings, we believe mathematics educators are best-positioned to make the assessment. However, we do make some tentative suggestions. First, mathematical proof may not be intrinsically unintuitive—it may instead be the level of abstraction of many proofs that masks intuitive understanding. If so, introducing simple deductive proofs of simple arithmetic relationships at an earlier educational stage could lay an intuitive foundation for more formal proofs later on (see Carpenter, Franke, & Levi, 2003). Second, people use the conceptual structure of mathematics to understand derivative concepts in terms of more basic ones. Educators may wish to emphasize these abstract connections, in conjunction with more concrete applications, in order to tap into this intuitive understanding; for example, explaining division both as a way to divide resources and as the inverse of multiplication. Finally, our methodology might be used to assess the explanatory trade-offs between different kinds of examples. Studying explanatory preferences in adults may provide a simple laboratory for testing out explanatory methods that might be used in educational settings, prior to undertaking expensive and risky intervention studies. This method could be used not only to illuminate mathematical understanding, but also the conceptual structure of other domains.

References

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, *44*, 75–106.

Bolzano, B. (1817). Purely analytic proof of the theorem that between any two values which give results of opposite sign there lies at last one real root of the equation. (Trans. S. Russ.)

Bonawitz, E. B., & Lombrozo, T. (2012). Occam's rattle: Children's use of simplicity and probability to constrain inference. *Developmental Psychology*, *48*, 1156–1164.

Carey, S. (2009). *The origin of concepts*. Oxford, UK: Oxford University Press.

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking mathematically: Integrating arithmetic and algebra in elementary education*. Portsmouth, NH: Heinemann.

Common Core State Standards Inititative. (2010). *Common Core State Standards for Mathematics*. Washington, DC: National Governors Association Center for Best Practices and the Council of Chief State School Officers.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. Oxford, UK: Oxford University Press.

Dedekind, R. (1995). *What are numbers and what should they be?* (H. Pogorzelski, W. Ryan, & W. Snyder, Trans. and Ed.). Orono, ME: Research Institute for Mathematics. (Original work published 1888.)

Descartes, R. (1954/1684). Rules for the direction of the mind. In E. Anscombe & P. T. Geach (Trans.), *Descartes: Philosophical Writings*. London: Pearson.

Dillon, M. R., Huang, Y., & Spelke, E. S. (2013). Core foundations of abstract geometry. *Proceedings of the National Academy of Sciences*, *110*, 14191–14195.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*, 15822–15826.

Frege, G. (1974/1884). *The foundations of arithmetic: A logico-mathematical enquiry into the concept of number*. (J. L. Austin, Trans.). Oxford, UK: Blackwell.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science*, *7*, 567–580.

Hardy, G. H. (2004/1940). *A mathematician's apology*. Cambridge, UK: Cambridge University Press.

Hume, D. (1978/1738). *A treatise on human nature*. Oxford, UK: Oxford University Press.

Johnson, S.G.B., Valenti, J.J., & Keil, F.C. (2017). Opponent uses of simplicity and complexity in causal explanation. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction: Essays in cognitive psychology*. Hillsdale, NJ: Erlbaum.

Kant, I. (1998/1781). *Critique of pure reason*. (P. Guyer, & A. W. Wood, Trans. and Ed.). Cambridge, UK: Cambridge University Press.

Kitcher, P. (1975). Bolzano's ideal of algebraic analysis. *Studies in the History and Philosophy of Science*, *6*, 229–269.

Kitcher, P. (1983). *The nature of mathematical knowledge*. Oxford, UK: Oxford University Press.

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from*. New York, NY: Basic Books.

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology*, *55*, 232–257.

Mill, J. S. (2002/1843). *A system of logic, ratiocinative and inductive*. Honolulu, HI: University Press of the Pacific.

Rips, L. J. (1983). Cognitive processes in propositional reasoning. *Psychological Review*, *90*, 38–71.

Rips, L. J. (1994). *The psychology of proof.* Cambridge, MA: MIT Press.

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences*, *31*, 623–687.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook for research on mathematics teaching and learning* (pp. 334–370). New York, NY: Macmillan.

Tao, T. (2016). *Analysis I* (3rd Ed.). Berlin: Springer.

Whitehead, A. N., & Russell, B. (1910). *Principia mathematica* (Vol. 1). Cambridge, UK: Cambridge University Press.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*, 749–750.

Appendix

Operation Explained	Operation Used to Explain	Stimuli	Proof	*F*	*R*
Addition	Subtraction	3 + 3 + 3 = 9 because 9 – 3 – 3 = 3	1R	0	1
	Multiplication	3 + 3 + 3 = 9 because 3 × 3 = 9	2R	0	1
	Division	3 + 3 + 3 = 9 because 9 ÷ 3 = 3	3R, 2R	0	2
	Exponent	3 + 3 + 3 = 9 because 3^2 = 9	4R, 2R	0	2
	Root	3 + 3 + 3 = 9 because √9 = 3	5R, 4R, 2R	0	3
Subtraction	Addition	9 – 3 – 3 = 3 because 3 + 3 + 3 = 9	1F	1	0
	Multiplication	9 – 3 – 3 = 3 because 3 × 3 = 9	2R, 1F	1	1
	Division	9 – 3 – 3 = 3 because 9 ÷ 3 = 3	3R, 2R, 1F	1	2
	Exponent	9 – 3 – 3 = 3 because 3^2 = 9	4R, 2R, 1F	1	2
	Root	9 – 3 – 3 = 3 because √9 = 3	5R, 4R, 2R, 1F	1	3
Multiplication	Addition	3 × 3 = 9 because 3 + 3 + 3 = 9	2F	1	0
	Subtraction	3 × 3 = 9 because 9 – 3 – 3 = 3	1R, 2F	1	1
	Division	3 × 3 = 9 because 9 ÷ 3 = 3	3R	0	1
	Exponent	3 × 3 = 9 because 3^2 = 9	4R	0	1
	Root	3 × 3 = 9 because √9 = 3	5R, 4R	0	2
Division	Addition	9 ÷ 3 = 3 because 3 + 3 + 3 = 9	2F, 3F	2	0
	Subtraction	9 ÷ 3 = 3 because 9 – 3 – 3 = 3	1R, 2F, 3F	2	1
	Multiplication	9 ÷ 3 = 3 because 3 × 3 = 9	3F	1	0
	Exponent	9 ÷ 3 = 3 because 3^2 = 9	4R, 3F	1	1
	Root	9 ÷ 3 = 3 because √9 = 3	5R, 4R, 3F	1	2
Exponent	Addition	3^2 = 9 because 3 + 3 + 3 = 9	4F, 2F	2	0
	Subtraction	3^2 = 9 because 9 – 3 – 3 = 3	4F, 2F, 1R	2	1
	Multiplication	3^2 = 9 because 3 × 3 = 9	4F	1	0
	Division	3^2 = 9 because 9 ÷ 3 = 3	4F, 3R	1	1
	Root	3^2 = 9 because √9 = 3	5R	0	1
Root	Addition	√9 = 3 because 3 + 3 + 3 = 9	5F, 4F, 2F	3	0
	Subtraction	√9 = 3 because 9 – 3 – 3 = 3	5F, 4F, 2F, 1R	3	1
	Multiplication	√9 = 3 because 3 × 3 = 9	5F, 4F	2	0
	Division	√9 = 3 because 9 ÷ 3 = 3	5F, 4F, 3R	2	1
	Exponent	√9 = 3 because 3^2 = 9	5F	1	0

Table 2: Mental proofs and penalty scores for all explanations.

		Operation Used to Explain						
		Addition	Subtraction	Multiplication	Division	Exponent	Root	**Average**
Operation Explained	Addition	—	6.37	7.96	5.70	5.69	4.47	**6.04**
	Subtraction	7.14	—	4.45	4.93	4.01	3.82	**4.87**
	Multiplication	8.11	4.20	—	7.12	8.02	6.01	**6.69**
	Division	5.76	4.65	7.46	—	5.70	5.45	**5.80**
	Exponent	6.12	3.65	8.75	5.26	—	6.49	**6.05**
	Root	5.27	3.43	7.11	5.45	7.44	—	**5.74**
	Average	**6.48**	**4.46**	**7.15**	**5.69**	**6.17**	**5.25**	

Table 3: Explanatory ratings for each pair of operations.