MIND, BRAIN, AND EDUCATION

From Ancient Greece to Modern Education: Universality and Lack of Generalization of the Socratic Dialogue

Andrea P. Goldin[1], Laura Pezzatti[1], Antonio M. Battro[2], and Mariano Sigman[1]

ABSTRACT—Two thousand four hundred years ago Socrates gave a remarkable lesson of geometry, perhaps the first detailed record of a pedagogical method in vivo in history [Plato. (2008). *Apología de Sócrates. Menón. Crátilo*. Madrid: Alianza Editorial]. Socrates asked Meno's slave 50 questions requiring simple additions or multiplications. At the end of the lesson the student discovered by himself how to duplicate a square using the diagonal of the given one as the side of the new square. We studied empirically the reproducibility of this dialogue in educated adults and adolescents of the 21st century. Our results show a remarkable agreement between Socratic and empiric dialogues. Even in questions in which Meno's slave made a mistake, within an unbounded number of possible erred responses, the vast majority of participants produced the same error as Meno's slave. Our results show that the Socratic dialogue is built on a strong intuition of human knowledge and reasoning which persists more than 24 centuries after its conception, providing one of the most striking demonstrations of universality across time and cultures. At the same time, they also emphasize its educational failure. After following every single question including Socrates' "diagonal argument," almost 50% of the participants failed to learn the simplest generalization when asked to double the area of a square of different size.

The dialogue between Socrates and Meno's slave (Plato, 2008) has been considered—without ever being submitted to explicit proof—one of the greatest landmarks in the history of education. The teacher is Socrates, the greatest pedagogue of antiquity; the pupil a young slave. Paradoxically, Socrates maintains that he is not teaching at all when he is asking his young pupil about the duplication of the area of a given square. When the dialogue ended, and the slave had *discovered* the solution to the geometrical question, Socrates asked Meno: "What do you think, Meno? Was there any opinion that he did not give as an answer of his own thought?" No—answered Meno—they were all his own" (Meno, 85c).

The Socratic dialogue also constitutes a magnificent experimental resource to investigate cognition in a controlled educational setup. It is structured according to precise script in which the teacher (Socrates) asks only elementary concepts in mathematics and the pupil (the slave) answers to most questions with a "yes" or "no." In the beginning of the dialogue the teacher shows the pupil a square and the objective is to make the pupil discover how to generate a new square with twice the area. During the dialogue the pupil makes errors, for instance thinking that the length has to be doubled to double the square. Subsequent questions are intended to make the student discover the error. A crucial moment of the dialogue is referred as the "diagonal argument" when the teacher points to the student the diagonal to make him reflect how this may be useful in the construction of the new square. After 50 questions, the pupil has discovered, mainly by answering yes or no to facts he already knew, how to double the area of the original square.

Here we investigate the universality of the Socratic dialogue, presenting an almost literal adaptation (with minor grammatic modifications to make it more fluid, see supporting information) to 58 participants, all native Argentinians. The dialogue was carried out verbally. For analyses, participants were grouped as adolescents (age $\leq$ 18, currently in high school) and adults (age $>$ 18, high school education completed, see "Materials and Methods" section).

[1]Laboratorio de Neurociencia Integrativa, Departamento de Fisica, FCEyN UBA and IFIBA, Conicet
[2]Academia Nacional de Educación, Buenos Aires, Argentina

Address correspondence to Mariano Sigman, Laboratorio de Neurociencia Integrativa, Departamento de Fisica, FCEyN UBA and IFIBA, Conicet; Pabellon 1, Ciudad Universitaria, C1428EGA Buenos Aires, Argentina; e-mail: sigman@df.uba.ar.

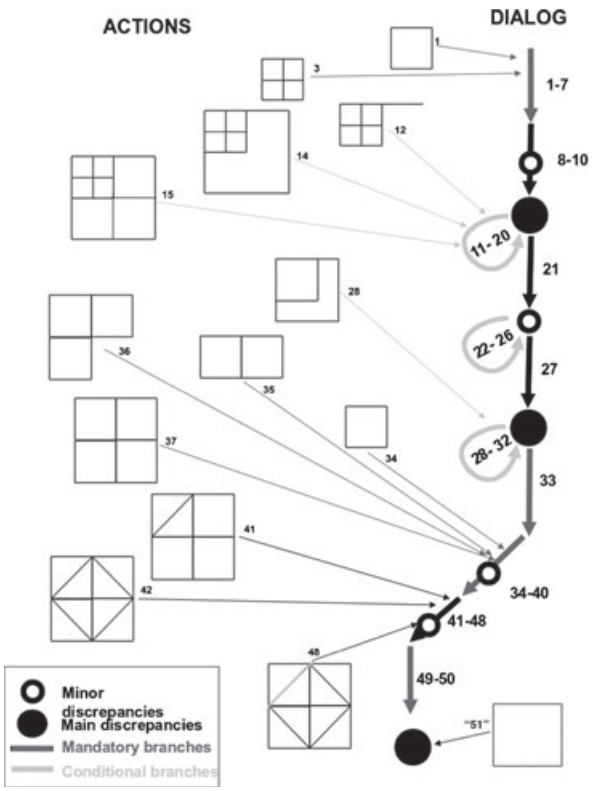

Fig. 1. Experimental structure of the Socratic dialogue. Conditional branches (light gray) diverge from questions in which the slave makes an error and are only transited if the participant makes exactly the same error. Linear branches (black) were followed sequentially unless participants made a discovery which made subsequent questions illogical. Only a few questions (dark gray) were strictly mandatory and were asked regardless of the participant answers. All sketches shown during the dialogue are pointed to the moment in the dialogue in which they were shown. Questions showing major (black, less than 60%) and minor (white, less than 33%) discrepancies between Socratic and experimental dialogue are labeled. During the course of the dialogue, the experimenter conducted a series of actions (indicated in the original dialogue) involving drawing and pointing. The drawing sketched during the dialogue is pointed to the corresponding moments.

RESULTS AND DISCUSSION

We parsed the dialogue in linear and conditional branches (Figure 1). Conditional branches (light gray) follow questions in which the slave makes an error, as in question 10:

> This (side) is two feet long: what will be the side of the other (square) which is double in size?
>
> Clearly, Socrates, it will be double (Meno, 82e).

Subsequent questions (from 11 to 20) elaborate on this mistake and are only traversed in the experiment if the participant makes exactly the same error. Linear branches were followed sequentially unless participants made a discovery which made subsequent questions illogical. For instance, in questions 41–48 Socrates elaborates the diagonal argument. If at any moment during this segment participants understood the diagonal argument and verbalized the solution, they skipped directly to question 49. Only a few questions (1–7, 33–40, and 49–50) were strictly mandatory and were asked regardless of the participant answers.

The average number of responded questions was significantly larger in the adolescent than in the adult group (adolescents, $76.6 \pm 4.1\%$ and adults, $63.9 \pm 3.1\%$; t-test: $t = 19$, $p < .0001$, $\mathrm{df} = 36$). We then measured the percent of agreement between the empiric and the Socratic dialogues. $p(\mathrm{Agreement}) = P(\mathrm{Identical}_{\mathrm{response}} | \mathrm{responded})$. Agreement between the Socratic dialogue and the experiment was remarkable (Figure 2, top panel and main image). In 28 out of 50 questions all participants, without exception, responded exactly as Meno's slave. On average, the degree of agreement was 93.7% and only in two out of 50 questions (questions 10 and 27) less than 60% of participants' responses differed from those made by Meno's slave.

Questions in the Socratic dialogue belong to two classes: (1) *Meno-Correct*, in which the slave response is correct and (2) *Meno-Incorrect*, in which the slave makes an error (see question 10 referred above as an example). The close agreement between the Socratic dialogue and the experiment is particularly remarkable in *Meno-Incorrect* questions for which the possible space of responses is virtually infinite. In fact, the vast majority of discrepancies result from mathematically correct responses.

It is notable that the errors committed by Meno's slave in the dialogue were largely ubiquitous in young thinkers of the 21st century. This is impressive even when analyzing question 10 referred above which elicits the greater disagreement. When asked this question, 55.2% of our participants responded, as Meno's slave, that the side had to be doubled to double the area; 41.4% responded correctly and only 3.4% responded a distinct but still erred response.

Agreement with the dialogue in question 10 was significantly greater for adolescents, of whom 61.9% responded that the line had to be doubled. Remarkably, 51.4% of adults, all of whom had high school education and most of whom had university education, responded that the side had to be doubled to double the area. The difference in agreement was significant between adolescents and adults ($p < .05$).

The second greatest discrepancy between empiric and Socratic dialogues occurred in the 27th question of the dialogue. Previous questions had lead Meno's slave (and our participants) to agree that the size of the length of the line which doubled the area of the square had to be somewhere between 2 and 4 ft (the original side had length of 2 ft). When asked about which of all values should be the length doubling the area, the slave response is: *Three*. Thirteen out of

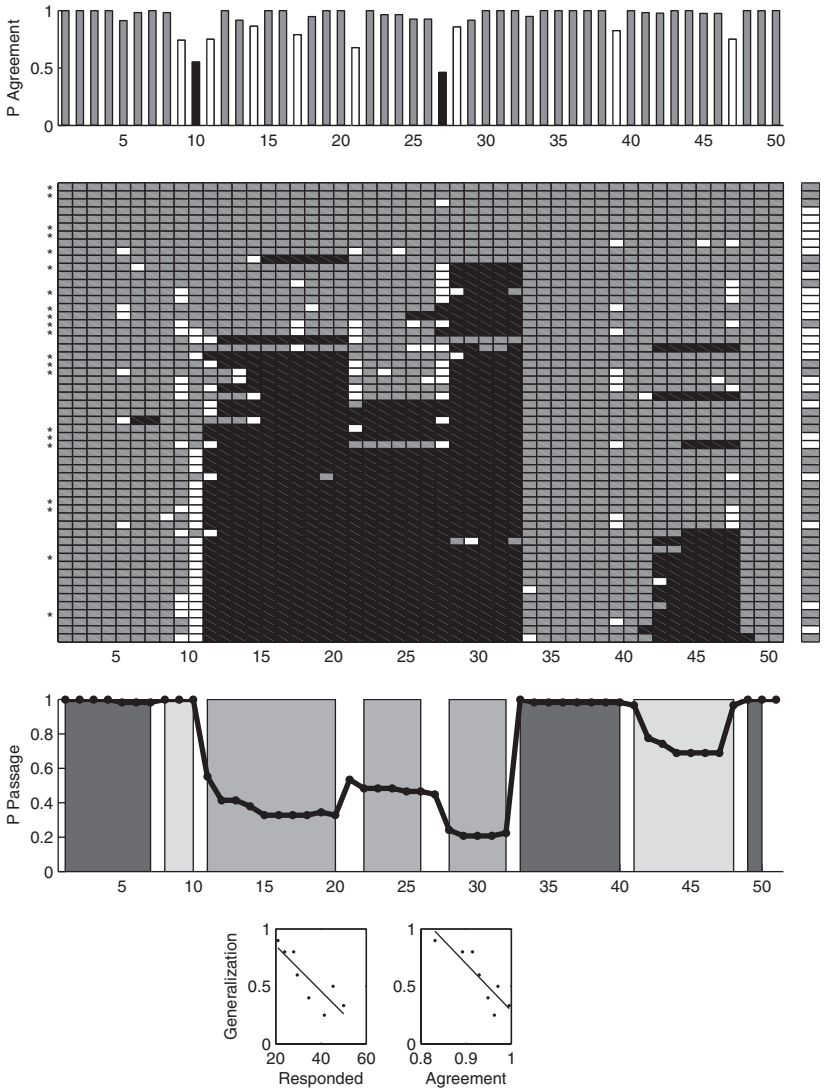

Fig. 2. Agreement between experimental and Socratic dialogue. Top panel: p(agreement) for all 50 questions of the dialogue. Main image: Responses of all participants in the study. Each column corresponds to a question, each line indicates a participant. White squares indicate discrepancies, black squares skips, and gray squares agreement. Participants were sorted according to the degree of average agreement. The rightmost column indicates whether participants could generalize knowledge to a new square a few minutes after completion of the dialogue. Stars at the left of each line indicate that the participant was adolescent; unmarked lines correspond to adult participants. Third row: Probability of responding (regardless of agreement) for each question. Colored shadows indicate distinct branches of the dialogue, as in Figure 1. Bottom panels: Probability of generalization as a function of the total number (left) and agreement (right) of responded questions.

21 teenagers and 13 out of 37 adults responded to this question. Of these, 53.9% of adults and 38.5% of adolescents made the same error as the slave, assuming that an answer which ranged between 2 and 4 ought to be 3. This result—while only marginally significant due to the small sample of responding participants—is in line with developmental and ethnographical studies which have demonstrated a cultural influence in a linear and symbolic bisection of the number line[1] (Dehaene, Izard, Spelke, & Pica, 2008; Feigenson, Dehaene, & Spelke, 2004; Lemer, Dehaene, Spelke, & Cohen, 2003).

After completing the 50 questions participants were shown a new square of a different size. They were asked the same original question: to double the area of the square. To our great surprise, more than half of the adolescents (57.1%) and almost one third of the adults (32.4%), after having followed the dialogue for about 15 min, failed to respond correctly. The probability of generalizing knowledge decreased when participants followed more closely the dialogue, as testified by the clear and significant negative correlation of the total number of responded questions and the probability of generalizing the learned knowledge to a new square ($R^2 = .74$; $F = 17.06$; $p < .01$; Figure 2, lower panel). Note that the participants who follow the dialogue closer are those that make the same errors as the slave and hence go through conditional branches. Hence,

this result also implies (as can be seen from the figure) that participants who responded to the critical branching questions as Meno's slave show less generalization to a new problem of the same kind. The degree of agreement also correlated negatively with generalization probability ($R^2 = .77$; $F = 20.32$; $p < .01$).

We have shown that a contemporary replication of the Socratic dialogue on geometry with an illiterate young Greek slave of the time of Plato shows remarkably similar results in a group of high school and college students, reflecting human cognitive universals traversing time and cultures. The Socratic dialogue is probably one of the most emblematic examples of education in a minimally guided environment, in which learners must discover or construct essential information for themselves. Socrates acts as a midwife, breaking the problem into components and prompting answers to each part separately (maieutics), a teaching strategy contemporarily referred to as *piloting* (Marton, 1997). Our observation of a lack of generalization in a strict Socratic dialogue extends a broad literature which has questioned the efficacy of unguided education, despite their broad popularity and intuitive appeal (Chen & Klahr, 1999; Kirschner, 2006; Klahr, 2009; Kuhn, 2005; Tikva, 2010).

Our results also question the efficacy of the modern educational system. What is it about modern education that still leads to many of the same shortcomings as in Socrates' time?

A first answer to this question may come from large-scale projects which might provide a finer grained understanding of the effect of distinct educational factors. Our current data was insufficient to perform a regression of all possible educational (degree of education, type of education, educational scores, etc.) and more general (age, gender, IQ, etc.) factors which may predict performance. This is why we broadly grouped the data in two categories: adults who had completed high school and adolescents. Our results show that education has a modest but significant impact on reasoning, reflected by the dialogue: adolescents had more agreement with the dialogue, especially in question 10 (62% in adolescents vs. 50% in adults) in which they erroneously responded that the side had to be doubled to double the area.

Despite this modest improvement it must be emphasized that knowing that doubling the length actually quadruples the area is a basic aspect of mathematics that all our participants have gone through (see Table 1). So the question still remains. Why a significant fraction still fails? Is there an educational system which would lead to less failures or do they reflect basic aspects of human cognition?

A second strategy to respond to these questions is by cross-cultural studies. At this stage our study is only examining the Argentinian educational system. The Socratic dialogue might turn out to be an interesting vehicle to probe the efficacy of certain aspects of distinct educational curriculum. Indeed, it is possible that the key training is not necessarily on mathematical aspects of the problem but in concentration during reasoning, focus of attention, and, more specifically in the presence and limits of heuristics in problem solving (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974). Heuristic methods are used to speed up the process of finding a good enough solution, where an exhaustive search is impractical, using an intuitive judgment, or common sense. While heuristic may turn out to be under many circumstances better than explicit reasoning (Dijksterhuis, Bos, Nordgren, & van Baaren, 2006) in mathematical problems, responding beyond heuristic, beliefs, and cognitive biases constitutes

Table 1
Content of Mathematical Knowledge Expected At Each Age Based On the Educational Curriculum of Argentina

	Number of participants	*Number of years spent in high school (HS)*	*Number of participants that failed Q51*	*Topics expected to be learned when finishing that year in formal education*
Adolescents	2	0 (starting HS)	2	Number lines. Recognition and use of integers and rational numbers. Proportions. Linear equations. Pythagoras. Volumes of bodies, properties of geometric figures
	1	1	0	Equations. Deepening in handling numbers. Lines, angles, Thales theorem. Relations between variables, graphics
	4	2	2	Real and complex numbers. Equations and inequalities. Algebraic expressions with integers. Polynomials factorization. Systems of linear equations. Probability and statistics
	1	3	1	Construction and interpretation of graphs. Linear, quadratic, rational, exponential and logarithmic functions. Probability and statistics
	13	4	7	Limit of a function, continuity of a function. Derivatives. Integrals. Probability and statistics
Adults	15	5	2	Exact or biological sciences studied after HS
	22	5	10	Nonexact nor biological sciences studied after HS

an important aspect of learned rationality (Gigerenzer & Reinhard, 2002). It is possible that most of the participants which responded incorrectly (as Meno's slave) to the 10th question did so because they relied on heuristic trying to find a fast solution, even if they were not explicitly speeded or pressed to respond in a short time. The belief which may drive this heuristic is that "everything is linear and proportional" hence to "double the area just double the length" leading to the most frequent error. While speculative at this stage, this suggests that explicitly informing students how behavior is often guided by beliefs might have a stronger impact than deeper emphasis in geometry.

This argumentation has an interesting corollary. We found that participants who responded to the 10th question as Meno's slave had significant greater difficulty to generalize to a similar problem of the same kind, even to a virtually identical problem (presenting a new square, only of different size). We can speculate that participants who are simply responding heuristically to the dialogue based on priors and beliefs are not undergoing the reasoning processes which might result in concrete learning that might resist generalization.

The specific geometric failure reflected by Meno's slave and by our participants (lack of understanding the scaling of the length and of the area) may reflect the limitations in an abstract system of encoding geometrical structures. This is inline with proposed mechanisms for matching concepts through analogies, as in the structure-mapping theory (Gentner, 1983) which posits that analogy is characterized by the mapping of relations between objects, rather than its attributes. While these rules have the desirable property that they depend only on syntactic properties of the knowledge representation, they may mislead the relevant content. In our specific example in the abstract (nongeometrical, nonconcrete) mapping of a line to a square, the only preserved relation is its "diameter" and hence the natural inference is that doubling the length of one should double any other feature of the other.

The remarkable detail of the Socratic dialogue made it possible to follow every step of the geometrical reasoning and places the Socratic dialogue, 23 centuries after its conception, as an ideal experimental vehicle to explore the physiological correlates of teaching and untangle one of the most distinctive elements of human culture (Battro, 2007, 2010; Strauss, 2005).

MATERIALS AND METHODS

Fifty-eight subjects (mean age 26.3 ± 1.7 years old) participated in the study. Participants were grouped as adults ($N = 37$; ages from 21 to 64 years; high school completed) and adolescents ($N = 21$, ages from 12 to 18 years; currently in high school).

To avoid subtle misinterpretations, participants were informed that there were no hidden tricks in the drawings or the questions, that is, if the experimenter drew a square, even if lines were not perfectly equal and perfectly parallel then, following common sense, they would agree that this was a square. The dialogue was carried out verbally. As all participants in this study were native Argentinians, the dialogue was converted to colloquial Argentinian Spanish, using the conjugation "vos" instead of "tú".

After each question, if the participant response coincided with the Socratic dialogue the experimentalist simply advanced to the next question. If the participant responded correctly where Meno's slave made an error, the researcher skipped subsequent conditional branches.

All dialogues were tape-recorded. Analyses were done offline. Audio tapes were transcribed and each answer was coded on a binary basis, as 1 if it coincided with the Socratic dialogue and 0 if it differed.

Acknowledgments —This work was funded by the Human Frontiers Science Program. We thank Mike Posner, Stan Dehaene, and David Klahr for reading the manuscript and for useful comments. We also thank the members of Expedición Ciencia for stimulating discussions and ExpC2010 for their participation in the experiment.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Full detail of the dialogue.

Figure S1. Representative drawings made by participants to the 51st question.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

NOTE

1 Interestingly, the majority of the adults who did not respond three did not do so because they realized that $3x3 = 9$ and hence reasoned that the number had to be between two and three and thus it ought to be 2.5.

REFERENCES

Battro, A. M. (2007). Homo educabilis: A neurocognitive approach. In M. S. Sorondo (Ed.), *What is our real knowledge of the human being?* (Vol. Scripta Varia 109, pp. 46–60). The Vatican: Pontifical Academy of Sciences.

Battro, A. M. (2010). The teaching brain. *Mind, Brain, and Education*, *4*, 28–33.

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, *70*(5), 1098–1120.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science*, *320*(5880), 1217–1220.

Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science*, *311*(5763), 1005–1007.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, *7*, 155–170.

Gigerenzer, G., & Reinhard S. (Eds.). (2002). *Bounded rationality: The adaptive toolbox.* Cambridge, MA: MIT Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.

Klahr, D. (2009). To every thing there is a season, and a time to every purpose under the heavens. What about direct instruction? In T. M. D. E. Sigmund Tobias (Ed.), *Constructivist instruction: success or failure*? (pp. 291–310). New York: Taylor & Francis.

Kuhn, D. D. Jr. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, *16*(11), 866–870.

Lemer, C., Dehaene, S., Spelke, E., & Cohen, L. (2003). Approximate quantities and exact number words: dissociable systems. *Neuropsychologia*, *41*(14), 1942–1958.

Marton, F. B. S. (1997). *Learning and awareness.* Mahwah, NJ: Erlbaum.

Plato. (2008). *Apología de Sócrates. Menón. Crátilo.* Madrid: Alianza Editorial.

Strauss, S. (2005). Teaching as natural cognitive ability: Implications for classroom practice and teacher education. In B. P. David & S. H. White (Eds.), *Developmental psychology and social change: Research, history and policy* (pp. 368–388). New York: Cambridge University Press.

Tikva, J. B. (2010). Socratic teaching is not teaching, but direct transmission is: Notes from 13 to 15-year olds' conceptions of teaching. *Teaching and Teacher Education*, *26*, 656–664.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.