How formalism shapes perception: an experiment on mathematics as a language

Jakob Kapeller*

Department of Philosophy and Theory of Science,
University of Linz,
Altenbergerstraße 69, 4040 Linz, Austria
E-mail: jakob.kapeller@jku.at
*Corresponding author

Stefan Steinerberger

Mathematical Institute,
University of Bonn,
Endenicher Allee. 60, 53115, Bonn, Germany
E-mail: steinerb@math.uni-bonn.de

Abstract: This paper discusses the role of mathematics as a conversational tool in economics. Based on the observation that mathematics is understood as an alternative language to express theoretical concepts and ideas, this paper reports experimental results trying to estimate the potential effects such a use of mathematics as a conversational tool may carry. These results refine certain intuitive statements about the role of mathematics in economic discourse and expose some unexpected effects in merit of further study. In particular, the results show that on average the mere presence of mathematics makes a problem seem more difficult, that mathematical knowledge is primarily attributed to specific training, that using mathematical expressions may decrease the proportion of people able to understand a certain argument and that mathematical arguments are more likely to convince men than women.

Keywords: mathematics; language; psychological effects; experiment; economic education.

Reference to this paper should be made as follows: Kapeller, J. and Steinerberger, S. (2013) 'How formalism shapes perception: an experiment on mathematics as a language', *Int. J. Pluralism and Economics Education*, Vol. 4, No. 2, pp.138–156.

Biographical notes: Jakob Kapeller is an economist who accidentally became a philosopher. He is working as a University Assistant at the Department of Philosophy and Theory of Science at the University of Linz. His research interests include the epistemology of the social sciences, the history of political and economic thought and heterodox economics.

Stefan Steinerberger is a PhD student at the Mathematical Institute of the University of Bonn. His research interests include the interplay of analysis and geometry in a broad sense, partial differential equations, non-linear dynamics and the applicability of formal systems in other domains.

1 Introduction

The precise relation between mathematics and economics – or rather what this relation *should* look like – is an important topic in economic discourse; it becomes even more important as the use of mathematics as a conversational tool increases. Among all social sciences, economics exhibits the strongest tendency to employ mathematical rhetoric, to a large extent because of its conception of mathematics as an alternative ‘language’. In this context we study the effects of using mathematics as a conversational tool experimentally. What kind of psychological phenomena can be observed in a mathematical setting? Will it, for instance, make an argument more convincing?

Our set-up was the following: we exposed post-graduate students from the social sciences to an unexpected problem (the Monty-Hall dilemma) and an accompanying solution. One group was given a simple ‘text-based’ solution while the other group got a ‘math-based’ version of the very same solution. Both versions of the solution were correct and complete; neither was missing any essential information. We explicitly chose this non-economic set-up to avoid a confusion of economic issues with the aspect of mathematization (as in Rubinstein, 2006a). In order to focus on the latter aspect we used a non-trivial, non-economic problem to represent the formal complexity of many theoretical arguments in economics. While bearing some affinities to Rubinstein (2006a) as well as to literature on mathematics education, the experiment underlying this paper is to the best of our knowledge the first of its kind.

The experiment is based on the famous ‘Monty-Hall problem’ (Rosenhouse, 2009), which is basically a problem in probability theory that has been used as a conceptual element in a popular game show (for a detailed description of the experimental setting see the Appendix). It is a game where the player is given three options to choose from (i.e., picking one of three boxes): two of them are blanks (the boxes are empty), while the third choice carries a prize. In the first round the player has to choose one of the three options. Then one of the two remaining options is removed from the game. All participants know that the removed option is always a blank. Now, the player has to make her final choice: she may either stick with her prior choice or change her decision and choose the remaining option. While many people believe there is a 50:50 chance to win the game, changing one’s selection leads to the prize in two out of three cases.

This paper is structured as follows: In Section 2, we briefly consider the role of mathematics in economics from a historical perspective and the general debates between more formally oriented economists and their critics. In Sections 3 and 4, we provide a detailed analysis of our experiment, its design and outcome and discuss possible implications of the results. Finally, we offer some concluding thoughts in Section 5.

2 Mathematics and economics

The use of mathematics in economics has significantly increased over the past 100 years (Debreu, 1991; Mirowski, 1991) in a way that understanding contemporary economic theory is conditional to acquiring specific mathematical skills. However, when it comes to identifying the appropriate role of mathematics in economics, a wide range of opinions can be found. While some believe that a rigorous use of mathematics could serve as a way to axiomatize the entire field of economics and thus make it precise in a manner previously believed to be reserved for natural sciences only (e.g., Lazear, 2000), others argue that the truth lies somewhere else and warn that an extensive use of mathematical trickery might be misleading and harmful. This discussion is often encountered in informal conference talks (Do economists take their own models seriously or is it just a mathematical mind-game?) or in matters of structuring curricula (How much mathematics should an economist know? Should we, polemically speaking, really trade Adam Smith for Isaac Newton?).[1] In any case it seems obvious that some mathematical abilities are a necessary prerequisite for participating in debates on economic theory. Paul Samuelson expressed this idea as early as 1952,

> "without mathematics you run grave psychological risks. As you grow older, you are sure to resent the method increasingly. Either you will get an inferiority complex and retire from the field of theory or you will get an inferiority complex and become aggressive about your dislike of it." [Samuelson, (1952), p.65]

Obviously, any advocate of rigorous axiomatization and 'ruthless' application of mathematics must reject the notion of hidden structure and truth unattainable by mathematical exactitude. From this viewpoint every relevant object or relation can be described by suitable corresponding mathematical expressions. Truly rigorous in his appraisal of mathematics in economics, Samuelson clearly embraces this position,

> "In principle, mathematics cannot be worse than prose in economic theory; in principle, it certainly cannot be better than prose. For in deepest logic – and leaving out all tactical and pedagogical questions – the two media are strictly identical. […] As slightly improved by my late teacher, Joseph Schumpeter, Fisher's statement was: 'There is no place you can go by railroad that you cannot go afoot.' And I might add, 'Vice versa!'" [Samuelson, (1952), p.56]

Samuelson adds that mathematics starts from very basic assumptions and then builds each new definition upon already existing definitions in a manner that allows reduction of any statement to more basic previously defined statements. While sounding self-evident, Samuelson ignores in silence that for all practical purposes the tower of definitions has become so large that a true reduction of a new theorem in mathematics to everyday language is, though possible in principle, practically often not feasible.

The real issue is, hence, whether theoretical reducibility alone is already sufficient to guarantee the existence of a meaning-preserving translation from a mathematical to a verbal expression. The answer given by Prof. Willard Gibbs at a Yale faculty meeting to the question as to whether students should be allowed to skip languages in favour of mathematics was, as famously quoted by Samuelson in the title page of his *Foundations of Economic Analysis* (Samuelson, 1948), that 'Mathematics is a language'. This assertion implies that mathematics can be seen as an equivalent to verbal language from a translational perspective. Similarly, accounts of economics models as a means of telling 'stories' by using mathematical expressions as conversational tools also invoke the idea that mathematics is a device to express our thoughts in – a device which is as good and unbiased as any other means of communication (Gibbard and Varian, 1978; McCloskey, 1998; Krugman, 1998; Sugden, 2000; Morgan, 2001). Eventually, we are left with the question of whether a true translation between two languages, or specifically mathematical and ordinary language, is really possible.

A pragmatic approach towards translation theory is that a perfectly equivalent translation is generally not feasible but that one can get pretty close by finding related concepts, ideas or notions (Quine, 1960). In this context one should not ignore the possibility of more subtle mechanisms being at work: the Sapir-Whorf hypothesis[2] famously states that an exact translation of languages is impossible because we do not use language to merely express our thoughts but conversely our entire mode of thinking is influenced by the language we use. As stated by Sapir (1949, p.162),

> "Human beings do not live in the objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language which has become the medium of expression for their society. [...] No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds."

Of course, this is taking the analogy too far. Additionally, nobody learns 'math' as a natural language but only much later and under circumstances, which do not resemble the learning of a language at all.[3] Nonetheless, it can hardly be denied that the conceptual frameworks we employ have a tremendous influence on our work. In the case of mathematics and economics, a simple example is the use of purely mathematical tools to work on an economic equation without reflecting its real-life meaning (i.e., a merely 'syntactic' approach as opposed to a comprehensive 'semantic' reading). Backhouse (1998) addresses this issue by illustrating how the meaning of central theoretical concepts in economics, like the invisible hand or the freedom to invest one's own capital, have changed in the process of their formalisation. In a similar vein, Rubinstein (2006a) conducted an experiment based on a layoff-decision, where the underlying optimisation problem was presented in a tabular form (for the control group) or in functional form (for the treatment group). His results document a shift in favour of the profit-maximising solution if the problem is represented in a functional form. From this he concludes that 'presenting the problem formally, as we do in economics, seems to obscure the real-life complexity of the situation for most students (including math students)' [Rubinstein, (2006b), p.879].

In sum, we agree that mathematics could be perceived as 'some kind' of language, although the comparison is rather lopsided. While both are used for communication in some form or another, a series of idiosyncrasies appear when looking at the details, which make it very difficult to take the analogy any further.

Criticism of an overly formalistic approach in economics is usually centred around the following arguments:[4] one, choosing a system of axioms is often inappropriate (partly because this ignores the possibility of paradigm-shifts, partly because already simple axioms entail a hidden structure that cannot be easily checked). Two, mathematical precision is not suited to treat an imprecise system, especially if chaotic elements such as humans are involved. Three, an all too mathematical approach will lead to a subtle form of bias (one will prefer to study models where the mathematics is simple). Finally, not unrelated to the first point, mathematics mostly delivers so-called 'closed systems' while the economy is sometimes seen as an ever-evolving open system. Additionally one could add sociological effects, i.e., if your peers expect your theoretical argument to be formulated mathematically, failing to fulfil this disciplinary convention could lead to adverse effects. King (2002, p.193) reports a case, where a submitted manuscript was initially rejected (by the referees of *Econometrica*), whereas a slightly modified version presenting the same argument in mathematical form was immediately accepted by the same journal.

We suggest broadening this debate by adding a rather pragmatic perspective and studying the *effects of using a 'mathematized' language on individuals in concrete examples*. Such an analysis could advance our understanding of economic discourse in general and economic education in particular. Thus, in what follows we investigate the psychological effects of mathematical formulations in complex arguments via a web-based experiment.

3 The experimental design

We are interested in how educated lay people like economics students, journalists and officials, or researchers from other fields perceive communication that involves mathematics. We chose to arrange an experimental survey among post-graduate students in the field of social and economic sciences to test the effects of mathematical arguments on the readers' perception.

3.1 The experimental setting

The experiment is based on a simple procedure: We present a basic version of the Monty-Hall problem to the participants and offer them a (correct) solution to the problem. Afterwards respondents answer a series of questions related to the problem and the solution and provide some personal data. While we present exactly the *same problem* to all respondents, the solutions differ: the control group is confronted with a verbally formulated solution, which relies on simple logical arguments and is easy to grasp (Krauss and Wang, 2003), while the treatment group is provided with a 'mathematical' translation of the above solution, where the basic argument is still clearly identifiable. Both solutions are correct and contain the same arguments: they only differ in 'language'. Solutions have been assigned randomly to the participants.

The problem is fairly well-suited for a general approach since it is highly unlikely that our respondents are completely familiar with the used mathematical expressions, while the basic message of our mathematical solution can be deciphered without fully understanding these symbols. However, some basic intuition in combination with careful reading alone allows the respondents to follow the basic steps of the argument rather independently of their specific educational background. We explicitly choose a setup where the presented mathematical expressions are complicated to decipher *literally* for reasons of internal and external validity; with respect to the former, we wanted to make our stimulus as unambiguous as possible. With respect to the latter, we argue that such a setup resembles a broad class of events relevant for the economic community. Among these are teaching, interdisciplinary discourse, discourses with lay-people and the application of scientific results to practical problems, possibly in conjunction with practitioners from other fields, like politicians or managers. To some, our mathematical solution will seem to be 'overly formalistic'. However, in practical terms the same characterisation holds for many economic models (Rubinstein, 2006b), which again points to our concerns for external validity. Since we are primarily interested in the effect conveyed by mathematical expression we opted for a rather neutral choice problem not associated with any specific discipline in the social sciences. This avoids unwelcome framing effects stemming from the possibility that the participants might recognise that they have solved similar problems in their fields of specialisation in the past. We also refrained from using an economic problem for much the same reason: an economic background would have posed a much more complicated and contextually embedded stimulus by itself and, therefore, possibly invalidated the conclusions on a more general level.

The questionnaire consisted of 11 questions (ten closed, one open). With the exception of the first question, which asked whether the participants were already familiar with the problem, the respondents could choose between four (in one case five) incremental and symmetrical options. The problem and both solutions are located in the appendix; the specific questions can be found in Tables 1 and 3-11.

3.2 The hypotheses

We created a series of hypotheses and formulated appropriate questions. These hypotheses can be structured along four dimensions: the *perceived difficulty* of the problem, the *attribution of* abilities, *the comprehensibility* of the solution as well as its *credibility.*

With respect to the first dimension, we postulate a framing effect (Goffman, 1974) created by the use of a mathematical idiom. We propose that if a problem is framed mathematically, e.g., by offering a mathematical solution, the problem will appear more difficult to the respondents. Mathematical framing, thus, should increase the perceived difficulty.

With respect to the attribution of abilities, we refer to the concept of *signalling* (Spence, 1974). Actors may signal their internal, and thus unobservable, abilities to other people by creating or acquiring representative artefacts (like university degrees or test scores). In this context we propose that the use of mathematics signals a specific competence to a given audience. In order to test for these effects we included three separate questions, which allow searching for signalling-effects on different levels (general 'respect,' past education, internal ability; see also Table 1). Additionally the open question, at the end of the questionnaire, asked for a written evaluation of the abilities of the solution's author.

Two different aspects motivate our hypotheses concerning comprehensibility: first, we propose that the mathematical solution is harder to grasp for the audience as compared to the verbal solution. This hypothesis is a rather trivial, but nonetheless carries direct implications for economic conversation. We expect this effect to be stronger for female participants because women tend to underestimate their mathematical abilities relative to men which would lead to a lower perceived understanding per se (i.e., without 'real' differences in understanding, see Hyde et al., 1990, or Eccles et al., 1993). Second, we assume that a mathematical solution is more surprising to our respondents.

In dimension 4, we explore the possibility of an increased trustworthiness instilled by the utilisation of mathematics as a language. It is generally accepted that expert knowledge is superior to lay knowledge. In this context we hypothesise that the use of mathematics is perceived as indicating expertise: respondents should judge the credibility of the mathematical solution more favourably.

These dimensions and hypotheses (and the associated questions) are summarised in Table 1.

Table 1 Hypotheses and associated questions

Dimensions and hypotheses		*Associated questions*	
Dimension 1:	*Difficulty of the problem*		
H1.1	If the solution is presented mathematically, the problem is perceived as more difficult.	Q1.1:	How do you perceive the difficulty of the problem?
		Q1.2:	How many people, do you think, can find correct answer without any help?
Dimension 2:	*Attribution of abilities*		
H2.1	If the solution is presented mathematically, the solution's author is perceived to be more intelligent/better educated.	Q2.1:	Do you have any respect for the intellectual achievement embodied in the solution?
		Q2.2:	Would one need a specific education in order to solve this problem?
		Q2.3:	How do you judge the intellectual abilities of the author in general?
Dimension 3:	*Comprehensibility and perception of the solution*		
H3.1:	If the solution is presented mathematically, the solution is less comprehensible.	Q3.1:	Did you understand the solution?
H3.2:	If the solution is presented mathematically, the solution is (even) less comprehensible for women as compared to men.	Q3.1:	Did you understand the solution?
H3.3:	If the solution is presented mathematically, the solution is more surprising for the respondents.	Q3.2:	Did you expect such a solution?
Dimension 4:	*Credibility of the solution*		
H3.4:	If the solution is presented mathematically, the solution is more credible and thus rather perceived as correct.	Q4.1:	Do you think the solution is correct?
H3.5:	If the solution is presented mathematically, the solution is perceived as more 'scientific'.	Q4.2:	Do you think the solution is of a scientific character?

3.3 A note on two pretests

We pretested the setting two times with small groups of students ($n_1 = 20$; $n_2 = 26$). The results from both pretests are comparable and showed, with one significant exception, no remarkable differences. The only striking difference was that whereas in the first pretest a huge majority of students accepted the suggested solution, in the second pretest a small majority objected strongly to it. Besides discrepancies in age and the fact that the second group consisted of fresh post-graduate students, while members of the first were experienced graduate students, there was only one major difference between the two pretests: the first was conducted by a 25-year-old research assistant, while the second by a 60-year-old university professor. The answers to the open question at the end of the questionnaire indicated that the students distrusted the professor and suspected him of delusion (which was not the case with the research assistant). Based on this experience we concluded that the experimenter might influence the answering behaviour of the participants in unexpected ways, so we decided to conduct the survey via a web-application in order to eliminate any potentially biasing influence generated by 'suspect' experimenters.

3.4 Participation

The survey was conducted via a web-application. An e-mail-request asking to participate in a survey on an 'interesting puzzle' was sent to about 650 post-graduate students in the field of social and economic sciences at the University of Linz. We did not disclose our specific motivation within the invitation to participate to avoid biased responses. The survey was open from the 25th of November 2009 to the 31st of January 2010 (however 90% of our respondents answered within a time frame of two weeks after we sent our request, while nobody responded in January). Three hundred seven PhD students looked at the survey and 158 completed it. After excluding 12 incomplete questionnaires, 146 responses remained, which were used as the basis for our statistical analysis. While the 'termination rate' seems rather high, the disconnections are equally distributed over both groups and do not bias our results (there are 76 respondents in the control group and 70 in the treatment group). Additionally, we compared the personal data (age or field of study) from our questionnaires to compare our sample's characteristics with the characteristics of the basic population we requested to participate. In this context no striking deviations appeared as is indicated by Table 2 (numbers in brackets show the distribution of participants between control group and treatment group).

Since we reframed the standard wording of the problem in order to impede cheating via internet search engines, a simple online query for key terms would not deliver any plausible result. However, the average time respondents needed to complete the survey was quite low (median: 6 minutes 45 seconds) – it seems therefore improbable that many participants even tried to search the web for possible hints on the correctness of the alleged solution. While 40% of the participants declared they already knew the problem, we could, with one exception, observe no decisive differences between those respondents who already knew it and those who did not. This exception relates to the fact that respondents already familiar with the problem indicated a better understanding of the proposed mathematical solution, which is the natural result for counterintuitive problems. Since those respondents, who already knew the problem, are fairly equally distributed over both groups (30/26) no biasing effect should occur. Additionally, we explored whether the respective field of study biases our estimations with regard to gender-specific effects, but found no evidence for such a potentially confounding effect.

Table 2 Distribution of respondents with respect to sex and field of study compared to the characteristics of the full population

Field of study	*Sample*				*Population*			
	Male		*Female*		*Male*		*female*	
	Number	*% of all participants*	*Number*	*% of all participants*	*Number*	*% of all participants*	*Number*	*% of all participants*
Economic sciences	40 (19/21)	28.17%	37 (19/18)	26.06%	301	36.75%	209	25.52%
Business informatics	12 (6/6)	8.45%	3 (2/1)	2.11%	81	9.89%	13	1.59%
Socio-economics	11 (7/4)	7.75%	11 (6/5)	7.75%	40	4.88%	50	6.11%
Sociology	3 (1/2)	2.11%	12 (5/7)	8.45%	29	3.54%	57	6.96%
Business education	3 (1/2)	2.11%	4 (4/0)	2.82%	23	2.81%	13	1.59%
Statistics	4 (3/1)	2.82%	2 (1/1)	1.41%	3	0.37%	0	0.00%
Total	73 (37/36)	51.41%	69 (37/32)	48.59%	477	58.24%	342	41.76%

Notes: In order to determine their field of study respondents were asked what kind of graduate study they completed (multiple answers were possible; in such cases we always counted the more specialised subject). These numbers are in turn compared to university data on the specific dissertation field of the whole population, which means that the data is only roughly comparable. This difference explains some deviations: first, university statistics list all PhD students (819) while we only sent our questionnaire to those students with a valid e-mail address (650). Second, some people have studied statistics (and are, thus, recorded as statisticians by our questionnaire), but achieve their PhD in other fields.

4 Experimental results and discussion

The statistical estimations that follow are all based on a comparison of mean values between the treatment group and the control group. Mann-Whitney U-tests were used to investigate whether the differences in mean values are statistically significant. Thereby, the Mann-Whitney U-test represents the equivalent to a standard t-test, but is specifically tailored to handle ordinal data. The resulting p-values are given in parentheses and interpreted as usual, i.e., *, ** and *** represent the 10%, 5% and 1% significance level respectively. Additionally, all tables show mean values for the subsets of women and men (again U-tests have been used to compare the differences in terms of statistical significance). The unreflected application of p-values in hypothesis testing is known to be problematic (see Gelman and Stern, 2006; Ziliak and McCloskey, 2008) – this is why we put an emphasis on presenting mean values as well as their differences. Since we always supply the scale of possible responses it is possible to compare the differences in mean values with the respective scale allowing for a basic inspection of the effect of size. For instance, if the scale of possible responses is 3, a difference of 0.3 already implies a jump of 10% relative to the total length scale. All tests based on concrete hypotheses are one-tailed and represented in *italics*, other tests are two-tailed and have a regular format. The number of respondents belonging to each group (n) is noted in parentheses. The results will be compared with the hypotheses developed in the preceding section.

4.1 Dimension 1: Perceived difficulty of the problem

The experimental data unambiguously confirms the hypothesis that the presentation of a mathematical instead of a verbal solution makes the associated problem appear more difficult – comparing mean values indicates an effect size of almost 10% of the global scale. This is in spite of the underlying problem being *exactly* the same for both groups. This effect is even stronger (an effect size of nearly 20% of the global scale) if the question generally asks whether others could solve the problem without help (Table 4) instead of focusing on individually perceived difficulty (Table 3). The obvious conclusion from this result is that the use of mathematics in economic discourse in general and economic education in particular may render concrete problems or questions *more complicated than they actually are*.

While this result may seem trivial, it is of special relevance when the reader is not familiar with the specific mathematical concepts embedded in a certain (sub)discipline, which is the case for most students in economics as well as for professionals from other fields and 'lay intellectuals'. Furthermore, while simple economic models might appear difficult to students, complex models of a special kind might seem equally difficult to an economics professional from an alternative subfield. This comparison indicates that perceived difficulty strongly depends on the reader's context. Nonetheless, most professional economists know that many models look more complex at first sight than when studied with moderate intensity. They have, thus, a psychological advantage when compared to people with less professional experience, especially their students.

Table 3 Perceived difficulty of the problem, question 1.1

Q1.1:	*How do you perceive the difficulty of the problem? (1 – very high, 4 – very low)*			
	Total	*Women*	*Men*	*Difference*
Verbal	2.50 (76)	2.51 (37)	2.49 (37)	0.02 (0.662)
Mathematical	2.24 (70)	2.26 (32)	2.22 (36)	0.04 (0.983)
Difference	0.26 *(0.009***)*	0.25 *(0.013**)*	0.27 *(0.064*)*	-

Table 4 Perceived difficulty of the problem, question 1.2

Q1.2:	*How many people, do you think, can find correct answer without any help? (1 – more than 75%, 4 – less than 25%)*			
	Total	*Women*	*Men*	*Difference*
Verbal	2.84 (76)	2.77 (37)	2.92 (37)	–0.15 (0.344)
Mathematical	3.41 (70)	3.35 (32)	3.47 (36)	–0.12 (0.322)
Difference	–0.57 *(0.000***)*	–0.58 *(0.003***)*	–0.55 *(0.006***)*	-

A pedagogical implication arising from this result is that one should present key economic problems as stand-alone issues first and introduce a verbal solution in the second step. If a mathematical solution is presented after the two former aspects are understood, the obscuring effect of mathematical expressions on the initial problem should be minimal.

4.2 Dimension 2: Attribution of abilities

The questions related to our second hypothesis, namely, that the author of the mathematical solution is perceived to be more intelligent or knowledgeable, delivered mixed results: While the respondents do not show any additional respect for the author of the mathematical solution or perceive her to be more intelligent, there is a substantial effect showing that the participants acknowledge that a certain specific skill or educational background is required to solve the problem (see Table 6). Consequently, students from the treatment group disproportionally often mention 'mathematical knowledge'[5] in the open question as a necessary condition for solving the problem (28 do so in the treatment group; 14 in the control group). Thus, the results imply that *the mere usage of mathematics suggests that mathematical tools (and skills) are necessary to solve the given problem*, even if this is actually not the case. While some might view this as an appropriate incentive structure, it is – for many central concepts in economics – essentially false, since mathematical representations of economic concepts typically represent only a reduced form of the latter.

Putting these results in a broader context, one could argue that mathematical symbols carry expert power: if a certain problem (or solution) is formulated mathematically, people perceive the problem to be a matter of (technical) experts – there is no bad conscience when leaving these problems to the designated and specifically educated people. This aspect might partially explain why economics exhibits a non-negligible amount of institutional power in the political sphere since mathematics may be established as a demarcation criterion between serious consultancy and partisan advice.

Table 5 Respect for the intellectual achievement embodied in the solution, question 2.1

Q2.1:	*Do you have any respect for the intellectual achievement embodied in the solution? (1 – yes, very much, 4 – no, not at all)*			
	Total	*Women*	*Men*	*Difference*
Verbal	2.18 (76)	2.08 (37)	2.30 (37)	–0.22 (0.155)
Mathematical	2.14 (70)	2.24 (32)	2.06 (36)	0.18 (0.623)
Difference	0.04 *(0.255)*	–0.16 *(0.390)*	0.24 *(0.073*)*	-

Table 6 Perceived educational preparation necessary to solve the problem, question 2.2

Q2.2:	*Would one need a specific education in order to solve this problem? (1 – yes in any case, 4 – definitely not)*			
	Total	*Women*	*Men*	*Difference*
Verbal	2.76 (76)	2.69 (37)	2.84 (37)	–0.15 (0.251)
Mathematical	2.07 (70)	2.12 (32)	2.03 (36)	0.09 (0.495)
Difference	0.69 *(0.000***)*	0.57 *(0.005***)*	0.81 *(0.000***)*	-

Table 7 Subjects' estimation of the author's intellectual ability, question 2.3

Q2.3:	*How do you judge the intellectual abilities of the author in general? (1 – very high, 5 – very low)*			
	Total	*Women*	*Men*	*Difference*
Verbal	2.54 (76)	2.51 (37)	2.57 (37)	–0.06 (0.700)
Mathematical	2.54 (70)	2.65 (32)	2.44 (36)	0.21 (0.173)
Difference	0.00 *(0.420)*	–0.14 *(0.331)*	0.13 *(0.155)*	-

There is also a gender-specific pattern incorporated in this issue: While female respondents tend to attribute a higher degree of respect and intellectual ability to the author of the verbal solution, male respondents show the opposite pattern. Although women as well as men give similar answers when it comes to the role of educational requirement, the differences in the former two aspects (personal respect and intellectual ability) might partially explain why women tend to evade economics-curricula: if they interpret mathematics, in contrast to their male counterparts, as intellectually inferior as compared to verbal explanations, they might see mathematical literacy as an 'empty skill', making them think twice before engaging in economics. This last conjecture, however, is in no way *directly* supported by the data.

4.3 Dimension 3: Comprehensibility and perception of the solution

Overall, respondents found the mathematical solution less comprehensible than the verbal solution. This result is mainly driven by the answers of those respondents who were not familiar with the problem and its solution in advance: as one would expect the difference between the treatment group and control group is smaller among those participants, who were already familiar with the problem (the difference in means was equal to 0.09 for those participants familiar with the solution and 0.63 within the rest of the sample). Table 8 additionally shows that this effect is slightly stronger for female participants.

Table 8 Comprehensibility of the solution, question 3.1

Q3.1:	*Did you understand the solution? (1 – yes, definitely, 4 – no, definitely not)*			
	Total	*Women*	*Men*	*Difference*
Verbal	1.51 (76)	1.62 (37)	1.41 (37)	0.21 (0.375)
Mathematical	1.93 (70)	2.09 (32)	1.78 (36)	0.31 (0.083*)
Difference	–0.42 *(0.003***)*	–0.47 *(0.017**)*	–0.37 *(0.032**)*	-

The result here is fairly trivial; however, since the effect is quite strong for women one could once again draw conclusions related to gender-specific aspects in choosing one's (primary) field of study (see also Chipman et al., 1992). Another interpretation is, of course, that men are more ashamed of their lack of logical and mathematical skills leading to an incentive to misrepresent their grasp of the solution. The implications for economic education are rather similar to those already advanced when discussing the results in dimension 1: one should try to establish a firm understanding of key economic problems and convey the main solution in a verbal form. If these aspects are understood, a more mathematical approach can be introduced more easily and effectively.

Additionally, the respondents were asked whether they had expected such a solution. The answers show that the participating PhD students are more surprised by the mathematical formulation of the solution, that is, they did not expect such an answer (with an effect size of 14% compared to the global scale). This effect stems to a large extent from the female respondents in our sample for whom the effect is much stronger.

Table 9 Surprise effect of the solution, question 3.2

Q3.2:	*Did you expect such a solution? (1 – yes, definitely, 4 – no, definitely not)*			
	Total	*Women*	*Men*	*Difference*
Verbal	2.26 (76)	2.26 (37)	2.27 (37)	–0.01 (0.758)
Mathematical	2.70 (70)	2.88 (32)	2.53 (36)	0.35 (0.116)
Difference	–0.44 *(0.002***)*	–0.62 *(0.001***)*	–0.26 *(0.125)*	-

4.4 Dimension 4: Credibility of the solution

Our last set of hypotheses is related to the credibility of the mathematical solution. We asserted that a technical argument bears some kind of 'mythical' authority to an audience not used to communication via mathematical expressions. Most interestingly these hypotheses generally failed. This result was driven by a substantial mistrust of our female respondents, while the male participants accepted the mathematical rather than the verbal solution (although there is only a small and statistically insignificant 'pro-maths'-effect). At the same time there was no notable difference between respondents who already knew the problem, and those who did not, on both questions. While there is no evidence on the causes for this gender-specific effect in our study, one is inclined to speculate that women are maybe more sensitive for a possible delusive usage of mathematics in arguments, or that a worse understanding of mathematics is furthering distrust in mathematical operations. In any case, the results imply that using mathematical concepts does not necessarily render a specific content more credible or convincing for an educated lay-audience.

Table 10 Perceived correctness of the solution, question 4.1 (six respondents choosing the option 'do not know' have been excluded from the sample)

Q4.1:	*Do you think the solution is correct? (1 – yes, definitely, 4 – no, definitely not, 5 – do not know)*			
	Total	*Women*	*Men*	*Difference*
Verbal	2.24 (74)	2.11 (36)	2.39 (36)	–0.28 (0.332)
Mathematical	2.38 (66)	2.58 (29)	2.20 (35)	0.38 (0.108)
Difference	–0.14 *(0.200)*	–0.47 *(0.023**)*	0.19 *(0.290)*	-

Table 11 Scientific character of the solution, question 4.2

Q4.2:	*Do you think the solution is of a scientific character? (1 – yes, definitely, 4 – no, definitely not)*			
	Total	*Women*	*Men*	*Difference*
Verbal	2.32 (76)	2.28 (37)	2.35 (37)	–0.07 (0.530)
Mathematical	2.31 (70)	2.50 (32)	2.14 (36)	0.36 (0.183)
Difference	0.01 *(0.448)*	–0.22 *(0.229)*	0.21 *(0.127)*	-

However, a relatively large number of respondents used the open question in our questionnaire to (explicitly or implicitly) comment on the question related to the correctness of the solution. Looking through the answers of respondents who strongly objected to the suggested solution we got the impression that the use of mathematical expressions created a significant 'fear of being deluded': a quarter of those respondents, who strongly objected to the solution, suspected a 'delusion through math' (3 out of 12; some who only objected weakly also did), while no one among those who strongly objected to the verbal solution (14) made a similar reference to delusion or trickery. Taking into account that many people in general deem the solution to be incorrect or at least very counterintuitive, we may have, up to a point, provoked such a skeptical reaction within our sample by combining this specific problem with a mathematical 'code'. This impression is strengthened by the observation that the answers on questions 3.1 and 4.1 are moderately correlated (Spearman's Rho of 0.394): if the respondents understand the solution (which is harder for the mathematical translation), they also tend to judge it to be correct. Thus, in summary, the results presented in this section seem to be the least reliable and imply the need for further experimental testing based on a different scenario.

Our colleagues thought these surprising results could be due to idiosyncrasies determined by the choice of our subsample.[6] Since the use of mathematics is often invoked as a demarcation criterion between science and non-science and many social sciences experienced a historical struggle for acceptance to characterise their work as 'scientific,' post-graduate students in the social sciences might be more reluctant and skeptic to grant authority to mathematical formulations when compared to the everyman on the street. While it is possible that such a 'general sample bias' has occurred, this is, of course, solely speculation. A replication of this study with a different population could shed more light on this intuitive claim.

5 Concluding thoughts

Overall, we believe to have shown that it is worthwhile to focus on the practical implications of the use of mathematics in economics. We are optimistic that our investigation and perhaps further related work are indeed capable of shedding some light on the psychological effects of mathematics in scientific debates in general and economic discourse in particular. The following table gives an overview of our experimental results.

Table 12 Experimental results compared to the original hypotheses

Dimensions and hypotheses		*Experimental result*
Dimension 1:	*Difficulty of the problem*	
H1.1	If the problem is formulated mathematically, the problem is perceived as more difficult.	Confirmed
Dimension 2:	*Attribution of abilities*	
H2.1	If the problem is formulated mathematically, the solution's author is perceived to be more intelligent/better educated.	Partially confirmed: respondents acknowledge primarily the necessity of a specific education/special training.
Dimension 3:	*Comprehensibility and perception of the solution*	
H3.1:	If the problem is formulated mathematically, the solution is less comprehensible.	Confirmed
H3.2:	If the problem is formulated mathematically, the solution is (even) less comprehensible for women as compared to men.	Confirmed
H3.3:	If the problem is formulated mathematically, the solution is (even) less comprehensible for women as compared to men.	Confirmed
Dimension 4:	*Credibility of the solution*	
H4.1:	If the problem is formulated mathematically, the solution is more credible and thus rather perceived as correct.	Not confirmed
H4.2:	If the problem is formulated mathematically, the solution is perceived as more 'scientific'.	Not confirmed

These results, assuming additional tests would further strengthen their credibility, could have a wide range of applications. First, it would fundamentally affect the notion of mathematics as a neutral language, which is accessible by anyone on the same grounds.

Mathematical language in economics, as we have seen, is not necessarily neutral or objective, but instead carries a wide range of connotations and additionally conveys a certain hidden message about the author's purported abilities (see H2.1). It might furthermore unnecessarily obscure certain arguments (see H1.1), which confirms some worries brought forth by those, who are critical of formalism in economics.

The implications for economic education are diverse: We might interpret the results as a cautionary tale and perhaps as a warning not to start introductory or intermediate courses in economic theory with a large list of 'basic formulas' (as some textbooks tend to do). Instead, it could be argued, let the student see what economics is really about, namely economic problems (like unemployment, scarcity, efficiency or poverty) before going into mathematical details (Hey, 2005). On the other hand, an increase in mathematical literacy in general might well lead to an increased skepticism with regards to the traditional mathematical apparatus used in neoclassical economics potentially raising awareness for alternative mathematical techniques and thereby fostering a pluralist approach to economics (as argued by Keen, 2009).

Another important and related point is raised by the apparent gender-differences in our results. The women in our sample tend to be more critical when it comes to using equations as a 'rhetorical crutch' and such a critical quality is certainly of worth for the future development of economics. On the other hand, this predisposition might lead to some kind of self-selection bias, which induces women to 'opt out of economics'. However, it also implies that the increase in female faculty we are currently witnessing in economics might have an impact on academic economic conversation in general.

Acknowledgements

The authors would like to thank Johann Burgstaller, Leonhard Dobusch, Volker Gadenne, Patrick Hanslmaier and Doris Weichselbaumer for providing valuable comments on earlier drafts of this paper and two anonymous referees for a series of helpful suggestions.

References

Backhouse, R.E. (1998) 'If mathematics is informal, then perhaps we should accept that economics must be informal too', *Economic Journal*, Vol. 108, No. 451, pp.1848–1858.

Chick, V. (1998) 'On knowing one's place: the role of formalism in economics', *Economic Journal*, Vol. 108, No. 451, pp.1859–1869.

Chick, V. and Dow, S.C. (2001) 'Formalism, logic and reality: a Keynesian analysis', *Cambridge Journal of Economics*, Vol. 25, No. 6, pp.701–721.

Chipman, S.F., Krantz, D.H. and Silver, R. (1992) 'Mathematics anxiety and science careers among able college women', *Psychological Science*, Vol. 3, No. 5, pp.292–295.

Colander, D. and McGoldrick, K. (Eds.) (2009) *Educating Economists the Teagle Discussion on Re-evaluating the Undergraduate Economics Major*, Edward Elgar, Cheltenham, UK.

Debreu, G. (1991) 'The mathematization of economic theory', *American Economic Review*, Vol. 81, No. 1, pp.1–7.

Eccles, J., Wigfield, A., Harold, R.D. and Blumenfeld, P. (1993) 'Age and gender differences in children's self- and task perceptions during elementary school', *Child Development*, Vol. 64, No. 3, pp.830–847.

Enyedy, N., Rubel, L., Castellón, V., Mukhopadhyay, S., Esmonde, I. and Secada, W. (2008) ‘Revoicing in a multilingual classroom’, *Mathematical Thinking and Learning*, Vol. 10, No. 2, pp.134–162.

Gelman, A. and Stern, H. (2006) ‘The difference between ‘significant’ and ‘not significant’ is not itself statistically significant’, *American Statistician*, Vol. 60, No. 4, pp.328–331.

Gibbard, A. and Varian, H.R. (1978) ‘Economic models’, *Journal of Philosophy*, Vol. 75, No. 11, pp.664–677.

Goffman, E. (1974) *Frame Analysis: An Essay on the Organization of Experience*, Harvard University Press, Cambridge (MA).

Hey, J.D. (2005) ‘I teach Economics, not Algebra and Calculus’, *Journal of Economic Education*, Vol. 36, No. 3, pp.292–304.

Hyde, J.S., Fennema, E., Ryan, M., Frost, L.A. and Hopp, C. (1990) ‘Gender comparison of mathematics attitudes and affect’, *Psychology of Women Quarterly*, Vol. 14, No. 3, pp.299–324.

Keen, S. (2009) ‘Mathematics for pluralist economics’, in Reardon, J. (Ed): *The Handbook of Pluralist Economics Education*, pp.150–167, Routledge, London.

King, J.E. (2002) *A History of Post-Keynesian Economics Since 1936*, Edward Elgar, Cheltenham, UK.

Krauss, S. and Wang, X.T. (2003) ‘The psychology of the Monty Hall problem: discovering psychological mechanisms in solving a tenacious brain teaser’, *Journal of Experimental Psychology: General*, Vol. 132, No. 1, pp.3–22.

Krugman, P. (1998) ‘Two cheers for formalism’, *Economic Journal*, Vol. 108, No. 451, pp.1829–1836.

Lazear, E.P. (2000) ‘Economic imperialism’, *Quarterly Journal of Economics*, Vol. 115, No. 1, pp.99–146.

McCloskey, D.N. [1998(1985)] *The Rhetoric of Economics*, University of Wisconsin Press, Madison.

Mirowski, P. (1991) ‘The when, the how and the why of mathematical expression in the history of economic analysis’, *Journal of Economic Perspectives*, Vol. 5, No. 1, pp.145–157.

Mitchell, J.M. (2001) ‘Interaction between natural language and mathematical structures: the case of ‘Wordwalking’’, *Mathematical Thinking and Learning*, Vol. 3, No. 1, pp.29–52.

Morgan, M.S. (2001) ‘Models, stories and the economic world’, *Journal of Economic Methodology*, Vol. 8, No. 3, pp.361–384.

Quine, W.V.O. (1960) *Word and Object*, MIT Press, Cambridge, Mass.

Rosenhouse, J. (2009) *The Monty Hall Problem: The Remarkable Story of Math’s Most Contentious Brainteaser*, Oxford University Press, Oxford.

Rubinstein, A. (2006a) ‘Skeptic’s comment on the studies of economics’, *Economic Journal*, Vol. 116, No. 510, pp.C1–C9.

Rubinstein, A. (2006b) ‘Dilemmas of an economic theorist’, *Econometrica*, Vol. 74, No. 4, pp.865–883.

Samuelson, P.A. (1948) *Foundations of Economic Analysis*, Harvard University Press, Cambridge, Mass.

Samuelson, P.A. (1952) ‘Theory and mathematics – an appraisal’, *American Economic Review*, Vol. 42, No. 2, pp.56–66.

Sapir, E. [1949(1929)] ‘The status of linguistics as a science’, in Sapir, E. (Ed): *Culture, Language and Personality*, pp.65–77, University of California Press, Berkeley, CA.

Spence, M. (1974) *Market Signalling*, Harvard University Press, Cambridge, Mass.

Sugden, R. (2000) ‘Credible worlds: the status of theoretical models in economics’, *Journal of Economic Methodology*, Vol. 7, No. 1, pp.1–31.

Ziliak, S.T. and McCloskey, D.N. (2008) *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, University of Michigan Press, Ann Arbor.

Notes

1 See, for example, the occurrences of this debate in Colander and McGoldrick (2009) or the *Economic Journal* (1998).

2 Named after the linguists Edward Sapir and Benjamin Lee Whorf and also sometimes called the linguistic relativity principle. Its status within the linguistic community has been largely that of a discredited theory starting with the rise of Chomsky's universal grammar in the 1950s. In the late 1980s there has been renewed interest in this topic and modified and weakened versions of this hypothesis are still inspiring psychological experiments and research in the field.

3 There are, however, noteworthy interactions between the way mathematical problems are verbally communicated and framed in natural language and the mathematical understanding acquired in the course of this communication process (Mitchell, 2001; Enyedy et al., 2008).

4 See Chick (1998), and Chick and Dow (2001).

5 We counted all answers incorporating one of the following words in evaluating the author's abilities: 'mathematics', 'mathematical', 'statistics', 'statistical' and 'probability theory'.

6 We are especially thankful to Volker Gadenne and Doris Weichselbaumer for this specific hint.

Appendix

Problem and solution(s) within the experimental design

The *problem* was presented to all participants in the following form (the questions from the questionnaire can be found in Table 1 and Tables 3 to 11).

> Assume you participate in a game of chance, where you have the chance of winning 1 Million Euros. The game is played as follows: Three hats are lying on a table; beneath one of them you find the money while you won't find anything under the other two hats. In the first step of the game you have to choose one of the hats. In succession the game-master takes one of the other two hats and removes it from the game, however, he will only remove a hat containing no money. Now you are given a chance: You can either stick with the hat you have chosen previously or change to the other hat. Assuming you wish to win the money: Should or should you not switch hats or does it make no difference at all?

The striking and at first counterintuitive solution is that it is indeed better to switch hats. The verbal solution, presented to the control group, was the following.

> Let us play through the possibilities. Assume you have chosen the first hat in the beginning of the game. Such a situation can be depicted as follows:
>
> First: The money is indeed hidden under the first hat. Thus, it makes sense to stick with your original guess in this case.
>
> Second: If the money is under the second hat, then the game-master has to remove the third hat and it would be preferable to switch hats.

> Third: The same holds true for the third case, where the money is hidden under the third hat and the game-master has to remove the second hat. Thus, switching hats is again preferable in this case.
>
> This means that switching hats will make sure that you win the money in two of three cases, while sticking with your original choice leads to success in only one of three cases.

Contrary to this, the mathematical solution, which has been presented to the treatment group, has the following form.

> In order to be able to deal with the 'random choosing' of a hat, we define a probability space (Ω, A, P) with $\Omega = \{1, 2, 3\}$, A = 2^Ω and $P(\{1\}) = P(\{2\}) = P(\{3\}) = 1/3$. We now create two random instances (w_1, w_2): the hat you have chosen in the beginning of the game and the hat, which contains the money. Since $|\Omega| < \infty$, it is sufficient to play through the possibilities. Due to reasons of symmetry, we can assume that in the beginning of the game the first hat was chosen. We distinguish between three cases:
>
> I. $(w_1, w_2) = (1, 1)$. Switching hats is not preferable.
>
> II. $(w_1, w_2) = (1, 2)$. The game-master will remove hat 3 from the game. Switching hats will thus guarantee that you win the money.
>
> III. $(w_1, w_2) = (1, 3)$. The game-master will remove hat 2 from the game. Switching hats will thus guarantee that you win the money.
>
> If you change hats, you thus win with probability 2/3 whereas sticking with your initial hat will only be effective with probability 1/3. It is thus preferable to change hats.