האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

A COMMENTARY ON MEL RUTHERFORD'S 'ON THE USE AND MISUSE OF THE "TWO CHILDREN" BRAINTEASER'

By

MAYA BAR-HILLEL

Discussion Paper # 551 May 2010

מרכז לחקר הרציונליות

CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Givat-Ram, 91904 Jerusalem, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Abstract

Rutherford (2010) criticizes the way some people have analyzed the 2-children problem, claiming (correctly) that slight nuances in the problem's formulation can change the correct answer. However, his own data demonstrate that even when there is a unique correct answer, participants give intuitive answers that differ from it systematically -- replicating the data reported by those he criticizes. Thus, his critique reduces to an admonition to use care in formulating and analyzing this brainteaser -- which is always a good idea -- but contributes little what is known, analytically or empirically, about the 2-children problem.

A Commentary on Mel Rutherford's

'On the Use and Misuse of the "Two Children Brainteaser"'

Maya Bar-Hillel

Reading Rutherford's (2010) critique of the 2-children problem, I had a déjà-vu experience. In the 1980's, I had had a similar exchange with Nathan (1986), who wrote a similar critique to which I responded (Bar-Hillel, 1989). My response here will bear some similarities to that earlier one, though the former is more comprehensive and contains data as well as argument.

Rutherford's complaint is two-fold. On the one hand, he critiques the analytic discussions of the 2-children problem as a probability riddle, and on the other hand he critiques the empirical research that used the 2-children problem to study certain aspects of people's intuitive probability reasoning. Let us consider both critiques.

Regarding the first critique, Rutherford is generally right (though I don't agree with all his examples). Some conditional probability problems, and the 2-children riddle among them, have been occasionally presented in somewhat ambiguous terms, allowing for more than one answer to be legitimately defended, depending on how one interprets and resolves the ambiguity. Granted[1]. Rutherford lists some examples, offering his opinion of whether they do or do not compel a unique correct answer. Whether or not one accepts Rutherford's specific classification of all his examples (I personally do not), Rutherford himself regards some formulations (e.g., Feller's, see Rutherford's Version 1 and 2; and Ghahramani's, see Versions 11 and 12) as beyond reproach. They have, he

[1] This was already granted in Bar-Hillel and Falk (1982) and in Bar-Hillel (1989).

asserts, a unique correct answer. So be it. He suggests that a problem be called a brainteaser if "the intuitive answer differs from the formal correct answer". I do not attach the same importance to the question whether a problem is truly deserving of this label, but neither do I have any objection to Rutherford's definition. So far -- so good.

Things are more complicated regarding the second critique. In order to classify a problem as a brainteaser according to Rutherford's definition, we must ascertain what "the intuitive answer" is, to check whether it "differs" from the "formal correct answer" (all quotes taken from Rutherford's footnote 1, defining a brainteaser). This, presumably, is where empirical data is needed (even just a thought experiment, or the intuition of some writer, are forms of empirical data). The researchers who have over the years collected intuitive answers to the 2-children problem (and other related problems) were not testing whether it is worthy of being called a brainteaser -- that has usually just been taken for granted. Nonetheless, their studies provide the required data. Oddly, Rutherford chooses not to report any of these data, apparently because he disapproves of the particular wordings of the problem (calling them "unfortunate", "ambiguous", "indeterminate", and even "improper") that were used to solicit people's intuitive answers, and so considers them dismissible. Instead, he collects his own data, using formulations he does approve. And what does he find?

In Version 11, to which the correct answer is 1/2, all the respondents gave the correct answer. But in Version 12, to which the correct answer is 1/3, only 2 respondents gave the correct answer (and 23 still gave the incorrect 1/2). Rutherford concludes (grudgingly, it seems) that Version 12 "could be used as a brainteaser", while continuing to complain that it is not the version cognitive psychologists have used. This objection notwithstanding, Rutherford actually replicates the findings he chose to ignore.

It seems to me that Rutherford's little study is just what is required to show that the alleged misdeeds of those cognitive psychologists who stand accused by him are of little consequence. It is clear from his own results that it is not the researchers' sloppiness, nor the fact that their questions do not deserve to be called brainteasers, that accounts for their findings. So beyond some finger wagging, what have we learned? Here is what everyone can agree upon (by "everyone" I mean Rutherford, on the one hand, and myself -- representing those whose work he criticized -- on the other).

On the analytic side: There are versions of the 2-children problem to which the correct answer is 1/3, there are versions to which the correct answer is 1/2, and there are versions to which one can plausibly defend either the answer 1/3 or the answer 1/2, because their wording is not "determinate" enough. These versions may only differ by slight, yet critical, nuances.

On the empirical side: People are generally insensitive to these answer-altering nuances. How do we know? Because the common and intuitive answer to *all* 2-children versions is 1/2 -- whether it is the right answer (Version 11), the wrong answer (Version 12), or neither (all those improper versions other researchers have been using for decades, which Rutherford asserts to have no unique answer). Moreover, as Rutherford has noted himself, included among the people who have been insensitive to these critical nuances are people who should know better, namely some of those who have written, academically or in the popular press, about these riddles (myself included).

As I see it, the bottom line of Rutherford's paper is: You people out there have been using probability brainteasers -- specifically, the 2-children problem -- improperly (Sections 1-3), but hey! -- when I do it the right way, I find the very same results (Sections 4-5). In this light, Rutherford's point seems to reduce to a chiding that care should be exercised in the wording and analysis of probability problems. This is advice one cannot but endorse. However, I would like to take issue with some of his more specific complaints.

First, I wish to disagree with Rutherford's intuitions about the various versions he critiques. Rutherford calls Feller's 1950 formulation ("Consider families with exactly two children … Given that a family has a boy…") "concise", "unambiguous", "clear", "correct". In contrast, he calls vos Savant's formulation (Version 4: "If a woman has exactly two children, at least one of whom is a boy …") "incorrect" and a "misuse" of the problem. I am hard pressed to see the difference between the two formulations. Saying to respondents "If a family [or woman] has exactly two children" is tantamount to inviting them to "Consider women [or families] with exactly two children"[2] . "Given that a family has a boy" is equivalent to "at least one is a boy". Indeed, the latter is the best explanation in ordinary language of what the technical term "given" means.

Second, I wish to disagree with Rutherford's dismissal of informally posed questions. Rutherford condemns all versions which he thinks allow for more than one answer (although I doubt either the researchers or the participants would agree with him; each only sees a single answer). To be sure, for better or for worse, math textbooks rarely present ambiguous word problems, let alone do so deliberately. But life presents

[2] I equate "women" with "families", because the valid question of the unit sampled is whether 2 siblings are sampled together, as they are when a woman, family, parent, mother, father, etc. is sampled, or are sampled separately, as when it is the individual children who are the sampling unit.

us with problems that are not formulated like textbook problems, either when posed by circumstances or when posed by others or when we pose them ourselves. It behooves us to find a way of using the formal math we know to solve the informal, ambiguous, improper, poorly defined problems with which life presents us (see discussion in Bar-Hillel 1989). Rutherford concedes that even in Feller's Version 1, which he strongly endorses, 1/2 is the correct answer *only if we make two assumptions* -- equiprobability and independence of the sexes -- which the endorsed version does not explicitly make. These are reasonable, and natural, assumptions to make. But it is likewise reasonable, and natural, to assume in almost all 12 versions in Rutherford's paper (with the exception of Feller's Version 2 and Rutherford's Version 12) that it is the family (or parent) which is the "unit of selection", rather than the child. I bet that is what most participants assume. I also bet they make this assumption unawares. Finally, I bet that if it were argued that in some formulations the child could also conceivably be the "unit of selection", they'd wonder: "But what difference does that make??" [3]

And herein lies the "paradox" of the 2-children problem. *That* is what makes it the popular riddle it is. So many versions of the problem, all reduced by respondents to a single problem, the one that first comes to mind, appropriately or not: "If some kid is a boy, what's the probability that his sibling is also a boy?".

For several decades, the "two children" problem (and its cousins, the two aces problem, Bertrand's Box paradox, the two cards problem, the Monty Hall problem, the 3 prisoners problem, etc.) have intrigued and delighted people, among them not just riddle-loving lay people but also professional mathematicians, statisticians, psychologists,

---

[3] Alas, I don't have the data to back up this bet.

philosophers and economists. This interest has spawned a large literature (although, disappointingly, very few important insights[4]. Although some of these reputable scholars have indeed occasionally made embarrassing assertions, Rutherford's paper trivializes the real conundrum underlying this literature: Who do so few people realize the difference between various versions of this entire family of problems? Why do they readily answer questions that Rutherford rejects as improper, for having no unique answer? Why do they favor the answer of 1/2 irrespective of the exact formulation? And why is it so hard to debug the erroneous solutions? Alas, Rutherford's paper has not advanced us towards understanding this conundrum.

References

Bar-Hillel, M. 1989 "How to solve probability teasers". *Philosophy of Science,* 56: 348-358

Bar-Hillel, M. and Falk, R. 1982. "Some teasers concerning conditional probabilities". *Cognition* 11: 109-122.

Falk, R. 1992 "A closer look at the probabilities of the notorious three prisoners". *Cognition,* 43(3): 197-223

Fox, C.R. and Levav, J. 2004. "Partition-Edit-Count: Naive extensional reasoning in judgment of conditional probability". *Journal of Experimental Psychology: General* 133(4): 626-642.

Nathan, A. 1986. "How not to solve it". *Philosophy of Science*, 53: 114-119.

Rutherford, M.D. 2010 "On the use and misuse of the "Two children" brainteaser. *Pragmatics and Cognition*.

[4] However, see Falk (1992); Fox and Levav (2004).