13th International Congress on Mathematical Education
Hamburg, 24-31 July 2016

COMPARING THE RELATIVE PROBABILITIES OF EVENTS

Egan J Chernoff Ilona Vashchyshyn Heidi Neufeld

University of Saskatchewan

The purpose of this article is to contribute to the research utilizing logical fallacies, in particular the fallacy of composition, to account for normatively incorrect responses given by prospective teachers to relative probability comparisons. In breaking with traditional research, participants were presented with events related to coin flips, rather than sequences, and asked to determine which is least likely to occur as well as to justify their response. Abductive reasoning is employed to substantiate our claim that the fallacy of composition is the most probable explanation of respondents' reasoning and justifications. Results question well established assumptions regarding research on relative probability comparisons, which have suggested that participants were actually comparing events rather than sequences; this research demonstrates that even when presented with events, the majority of respondents still give normatively incorrect responses.

PRELUDE

Since 2007, there has been a (rather) systematic investigation into prospective elementary, middle, and high school mathematics teachers' relative probability comparisons of coin flip sequences (e.g., given a sequence of coin flips, participants are asked which sequence of heads and tails is more or less likely). The varied results stemming from this investigation have been presented at numerous mathematics education conferences. Without fail, the question and answer period for each and every presentation includes one particular exchange.

We now present a composite version of what is inevitably overhead during the question and answer period following said conference presentations: "Yes, your results (and analysis) demonstrate that the participants deem [using H for head and T for tail] HHHHT as less likely than HHTHT, but they're probably comparing the event [i.e., a set of outcomes] four-heads-and-one-tail to the event three-heads-and-two-tails, in which case, then, 'HHHHT' *is* less likely than 'HHTHT', which means that although their answer is incorrect for the question you gave them, they're not necessarily incorrect if they're comparing events instead of sequences." More often than not, the presenter is *not* asked to respond or comment on this line of reasoning, that is, the question is more of a statement. Responses that do occur typically make reference to research literature that has: (1) controlled the ratio of heads to tails in the sequences presented (e.g., Chernoff, 2009), (2) investigated what happens should sequences be interpreted as events (e.g., Abrahamson, 2009) and (3) resulted from investigating the idea that students are answering questions that may not have been asked (e.g., Chernoff & Mamolo, 2015). Attendees leave the presentation with air of security that participants were, in fact, comparing events instead of sequences, and doing so correctly. The courage of their conviction is strong enough for them to leave certain that not only are participants comparing events and not sequences, but also that any research conducted would simply confirm their convictions. Research that has recently been conducted (e.g., Chernoff, 2012a) suggests that the above presented conference convictions are unfounded.

INTRODUCTION

Investigations into relative probability comparisons, especially as related to coin flip sequences, have become an established thread in the last 50 years of probability education research (e.g., Abrahamson, 2009; Batanero & Serrano, 1999; Chernoff, 2009, 2012ab; Chernoff & Russell, 2012; Falk, 1981; Konold, Pollatsek, Well, Hendrickson & Lipson, 1993; Shaughnessy, 1977). In general, the aforementioned research is dedicated to accounting for normatively incorrect relative probability comparisons. Should a research participant deem, incorrectly, that one sequence of coin flips is more or less likely than another, there are now a plethora of established theories and models to account for the "error," which include the *representativeness heuristic* (Tversky & Kahneman, 1974; Shaughnessy, 1977), the *outcome approach* (Konold et al., 1993), *sample space partitions* (Chernoff, 2009), *attribute substitution* (Kahneman & Frederick, 2002; Chernoff, 2012b), *inadvertent metonymy* (Abrahamson, 2009), and others. Despite the differences in the theories and models accounting for normatively incorrect relative probability comparisons, they all have one thing in common.

A BRIEF APPRAISAL OF THE RESEARCH LITERATURE

Given the participants of this conference (i.e., #ICME13), this Topic Study Group (i.e., #TSG14: Teaching and Learning of Probability), and the category of this particular article (i.e., an invited paper for #TSG14 at #ICME13), we feel justified in taking certain liberties with this appraisal of the research literature. Instead of, once again, detailing the particulars of the Kahneman and Tversky's representativeness heuristic, followed by the particulars of Konold's outcome approach, followed by the particulars of all the other theories associated with normatively incorrect relative probability comparisons, the reader of this article is directed to eloquent summaries of the theories detailed elsewhere (see, for example, Jones, Langrall & Mooney, 2007; Kahneman, Slovic & Tversky, 1982; Shaughnessy, 1992). Rather, this appraisal of the literature will focus on the nature of the reasoning behind the aforementioned theories.

The reasoning associated with relative probability comparison theories (e.g., the representativeness heuristic, the outcome approach, sample space partitions, and others) is based on Peirce's (1931) notion of abductive reasoning. Upon hearing or reading a response Q given by a participant, a researcher infers a rule P which implies Q (e.g., if the participant is using the representativeness heuristic, then she will answer Q) and a consequence P (e.g., that the participant is using the representativeness heuristic). Peirce describes abductive inferences as inferences to the *best* explanation. In a more general sense, facts are used as a starting point, a particular hypothesis (derived from inferences and used to best explain the facts observed) is then presented, and if it is the case that if the hypothesis were true it would best or most likely explain the observed facts, there exists reason to suspect that the theory hypothesized is true (Lipton, 1991).

We contend that relative probability comparison theories were created in a similar fashion. Case in point, Chernoff (2009) observed through participants' (unconventional) written descriptions of the sample space that certain individuals' relative probability comparison responses were in accord with (what Chernoff called) a *subjective sample space* – a sample space partitioned according to the participant's interpretation of the sequence of coin flips presented (e.g., switches, runs, longest run, switches and longest run). The observed facts, the written responses, acted as the starting point. It

was then hypothesized that certain individuals answering the question were employing a subjective sample space in answering the task. Then, it was demonstrated that if the newly hypothesized model were true, the results observed would follow accordingly (which occurred through the use of Chernoff's [2009] notions of a *meta-sample space* and *event description alignment*). As such, there existed reason to suspect that the subjective sample space hypothesis was true, that is, to accept the claims of the newly hypothesized model. Subsequently, the notion of subjective sample spaces (as was the case with the representativeness heuristic, the outcome approach, inadvertent metonymy, and other theories) exists in the research literature as a theory or model (specific terminology has not yet been adopted) to explain how certain individuals respond to particular probability tasks.

As evidence mounts for a theory or model there is, on behalf of respective research communities, more and more reason to suspect that the theory is true (or, at least, the best possible explanation for an observed phenomenon); that is, that the participants are employing a particular theory in their answering of a question (see, for example, Chernoff [2012b] for a discussion of the reification of the representativeness heuristic in the field of mathematics education). However, it should be noted that one cannot claim with *certainty* that if an individual declares, in this instance, that one sequence is more or less likely to occur that they have indeed done so according to a particular theory or model. The introduction of new theories and models questions the abductive notion of the 'best' possible explanation(s) associated with previous research.

RESEARCH DESIGN AND RATIONALE

The standard models for the relative probability comparisons in the field of mathematics education (e.g., the representativeness heuristic, and the outcome approach) have had a long run as the best (abductively speaking) models accounting for normatively incorrect, inconsistent and sometimes inexplicable answers to certain probability questions. Within this long and storied run, new research continues to investigate normatively incorrect responses to relative probability comparisons (e.g., inadvertent metonymy, sample space partitions and others).

An emerging thread of research associated with the teaching and learning of probabilistic thinking suggests that logically fallacious reasoning, specifically the use logical fallacies, best (again and hereafter, abductively speaking) accounts for certain normatively incorrect responses to probabilistic tasks. Put simply, *the fallacy of composition* occurs when an individual infers something to be true about the *whole* based upon truths associated with *parts* of the *whole*. For example: Bricks (i.e., the parts) are sturdy; buildings (i.e., the whole) are made of bricks; therefore, buildings are sturdy (which is not necessarily true). Chernoff and Russell (2012) utilized the fallacy of composition to analyze responses when participants (prospective mathematics teachers) were asked to determine which of five possible coin flip sequences were least likely to occur. The fallacy of composition accounted for normatively incorrect responses to the task. More specifically, the researchers demonstrated that participants tended to reference the equiprobability of the coin, note that the sequence is comprised of flips of a fair coin and, as such, fallaciously determined that the sequence of coin flips should also have a heads to tails ratio of one to one. In other words, the properties associated with the fair coin (the parts), which make up the sequence (the whole), are also expected in the sequence. As such, the fallacy of composition, in addition to the traditional theories, models and frameworks associated with heuristic and informal reasoning, accounted for certain incorrect responses.

In breaking from the tradition of presenting sequences, which is found in all previous research on relative probability comparisons, Chernoff (2012a) and Chernoff and Russell (2012) demonstrated that certain individuals (prospective mathematics teachers), when asked to identify which event (i.e., outcome or subset of the sample space) from five flips of a fair coin was least likely to occur, did not use the representativeness heuristic (proposed by Kahneman & Tversky, 1972) or the outcome approach (proposed by Konold et al., 1993). Rather, they too were observed to utilize the fallacy of composition. In this instance, the coin flips [the parts] are equiprobable, and the events [the whole] are comprised of coin flips; therefore, the events are deemed to be equiprobable (which, again, is not necessarily true). Worthy of note, perhaps surprisingly, the fallacy of composition accounted for both normatively correct and incorrect responses to the new relative likelihood comparison task.

TASK AND RESEARCH QUESTIONS

This article extends recent investigations utilizing logical fallacies, specifically the fallacy of composition, as a theoretical framework to account for certain normatively incorrect responses given by prospective mathematics teachers. In addition, the present study eschews the more traditional approach of presenting sequences and, instead, continues with recent investigations where participants are presented with events and asked to determine which is least likely to occur. As is the case with all prior research in this domain, the abductive approach was applied throughout the study.

The purpose of this study is to establish whether the fallacy of composition best accounts for certain incorrect responses given by prospective mathematics teachers when comparing the relative probabilities of given events. To achieve this goal, three different versions of a task familiar to those in the research literature were presented to 363 prospective teachers of mathematics.

RESULTS AND ANALYSIS

The current study, involving 363 prospective mathematics teachers, is comprised of three separate, yet similar studies consisting of 106, 162 and 95 participants, denoted as Study I, Study II and Study III, respectively. The task given to participants in the different studies varies only slightly in nature and parallels the developmental history of the sequence version of the relative probability comparison task found in the field of mathematics education research (see, e.g., Chernoff & Russell, 2012).

Study I

As presented in Table 1 below, of the 106 participants that took part in Study I, 73 (approximately 69% of) respondents incorrectly established that the event ‘five heads and one tail’ was equally likely as the event ‘four heads and two tails.’

Which of the following is the least likely result of six flips of a fair coin?
a) five heads and one tail **[23/106]**
b) four heads and two tails **[10/106]**
c) both results (a and b) are equally likely to occur **[73/106]**
Please justify your response…

Table 1: Task and results from Study I

Among the 73 participants that incorrectly established that the two events were equiprobable, 14 responses were flagged as evidencing use of the fallacy of composition. However, given pagination limitations, only five of the fourteen responses are now presented in detail.

Sam:	Both results a and b are equally likely to occur because the probability of being heads or tails on every flip is equal (50/50).
Diane:	Both are equal in likeliness to occur. There is a 50/50 chance that each flip will get a head or tail, so the probability is equal for every option.
Coach:	Each individual flip has the same odds/probability of being either heads or tails (50/50). Therefore all combinations of heads and tails are equally likely to occur.
Cliff:	The probability of a coin being heads or tails when flipped is always 50/50. This means that all combinations should be equally probable.
Norm:	Both results are equally likely to occur because the probability of flipping a tail or a head is equal.

Observe that Sam, Diane, Coach and Cliff all reference the 50/50 "probability" (Sam, Coach and Cliff), "odds" (Coach) and "chance" (Diane), whereas Norm references the "probability" as "equal," of the coin being heads or tails. All five respondents, however, are clear in establishing that it is "because" (Sam and Norm) (or "means" [Cliff], "Therefore," [Coach] or "so" [Diane]) a coin has an equal probability of heads or tails that the "results" (Sam and Norm), "combinations" (Cliff and Coach), "option[s]" (Diane) should be "equally probable" (Cliff) or "equally likely / equal in likeliness to occur" (Cliff, Coach, Sam and Diane). Despite these slight differences in terminology, the core of the responses are strikingly similar: the coin is equiprobable and, as a result, the events are equiprobable. Stated in fallacy of composition parlance, all five individuals infer that equiprobability is true for the events [the whole] because it is true of the coin [the part] which is part of the event [the whole]. Similar findings associated from a similar task were found in Study II, which is presented below.

Study II

As is shown in Table 2 below, of the 162 participants that took part in Study II, 114 (approximately 70% of) respondents incorrectly established that both events, 'three heads and two tails,' and 'four heads and one tail,' are equally likely to occur.

Which of the following (events) is the least likely to result from five flips of a fair coin?
a) three heads and two tails **[13/162]**
b) four heads and one tail **[35/162]**
c) both outcomes are equally likely to occur **[114/162]**
Justify your response. In other words, explain your answer…

Table 2: Task and results from Study II

Of the 114 normatively incorrect responses, 5 evidenced use of the fallacy of composition when determining the relative probability of the events in the task, which are now presented.

Carla:	There is a 50/50 chance of flipping either Heads or Tails. Therefore there is a chance that it could easily be either a or b.
Woody:	[…] Flipping 3 heads and two tails is just as likely as flipping 4 heads and one tail because theres a 50/50 chance of flipping heads or tails […]

Rebecca: You have an equal chance (50%) of getting heads as you do tails; therefore, you have an equal likelihood of getting any combination of heads and tails.

Frasier: When flipping a coin there are only 2 outcomes: heads or tails. Therefore, there is a 50% chance of getting heads for one flip and 50% chance of getting tails for the same toss. It is just as likely to get 4 heads and one tail as it is to get 3 heads and 2 tails because looking at each indidivual toss the coin has equal outcomes of going heads or tails.

Lilith: Both results are equally likely to occur because you have flipped the coin five times, there is a chance each time that you can get either heads or tails, so there is an equal chance of coming out with the outcome of a) or b).

Observe that Carla and Woody referenced the "50/50 chance" of getting heads or tails, while Rebecca and Frasier stated the probability as a percentage. Lilith, however, does not mention the exact probability of a single coin flip, but simply references the "chance each time that you can get either heads or tails." Each of the five respondents stated the equiprobability of the coin, and concluded ("therefore" [Carla, Rebecca, Frasier], "because" [Woody] or "so there" [Lilith]) that "a or b" (Carla and Lilith), or "4 heads and one tail [and] 3 heads and 2 tails" (Frasier and Woody) or even "any combination of heads or tails," (Rebecca) could "easily be," (Carla), are just as/equally "likely" (Woody, Frasier and Lilith), or have "equal likelihood" (Rebecca). As in Study I, although the respondents showed slight variations in their language, infer the equiprobability of the events from the equiprobability of the coin.

Study III

As Table 3 below shows, of the 95 participants that took part in Study III, 70 (approximately 74% of) respondents incorrectly established that all of the given events ('four heads and two tails,' 'five heads and one tail,' 'three heads and three tails,' and 'two heads and four tails') were equally likely.

Which of the following events is least likely to result from flipping a fair coin six times?
a) four heads and two tails **[1/95]**
b) five heads and one tail **[18/95]**
c) three heads and three tails **[6/95]**
d) two heads and four tails **[0/95]**
e) all four events are equally likely **[70/95]**
Give a reason for (i.e., justify) your answer…

Table 3: Task and results from Study III

Among the 70 participants that incorrectly established that the two events were equiprobable, 7 responses were flagged as evidencing use of the fallacy of composition. Six of these responses are presented in detail below.

Paul: All four events are equally likely to occur because it is a fair coin meaning 50% chance it will be heads and 50% chance it will be tails leading to any of the outcomes.

Robin: Each time you flip the coin there is a fifty/fifty chance so any of those events are equally likely.

Evan: All equal because there will always be a 50/50 chance of flipping H or T on each flip.

Kelly: Because when you flip a coin you have a 50% chance of it being a head or a tail so each combination could be equally likely to occur? I'm not sure…

Eddie: All four events are equally like because flipping a head or a tail is ½ or 50%.

Gary: All four events are equally likely because each time you flip a coin there is a 50% chance of a heads and 50% of getting a tails.

As in the previous two studies, the above responses explicitly evidence utilization of the fallacy of composition when determining the relative probability of the events presented. Five responses refer to the part ("a fair coin" for Paul, "each time you flip" for Robin and Eddie, "each flip" for Evan and "flip a coin" for Kelly). All six responses include the characteristic of the part, the equiprobability of a single coin flip, stated as a "50% chance" of heads/tails (by Paul, Kelly and Gary), a "50/50 chance" (by Robin and Evan), or the somewhat more ambiguous "a head or a tail is 1/2 or 50%" (by Eddie). Again, the inference about the wholes, here referred to as "all four events" (by Paul, Eddie and Gary), "any of those events" (by Robin), "all" (by Evan), or "each combination" (by Kelly), are "equally likely" (according to Paul, Robin, Kelly, Eddie and Gary) or "equal" (for Evan) is reached "because" (as per Paul, Evan, Kelly, Eddie and Gary) the part (a single flip) is equiprobable.

DISCUSSION AND CONCLUDING REMARKS

Rather than being asked to compare sequences of coin flips, prospective mathematics teachers in three separate studies (363 participants in total) were asked to compare events related to coin flips. In these studies, 69%, 70%, and 73% of participants, respectively, incorrectly responded that all events were equally likely to occur. Moreover, 26 responses in total (of the 363 participants and the 257 participants that deemed events equally likely) were identified as evidencing use of the fallacy of composition. As such, the fallacy of composition, in addition to the traditional theories, models and frameworks associated with heuristic and informal reasoning, has been shown to account for certain incorrect responses to questions involving not only coin sequences, but also events related to coin flips (Chernoff, 2012a; Chernoff & Russell, 2012).

We have also argued that, despite the differences in the various theories and models accounting for normatively incorrect relative probability comparisons, they all have one thing in common: namely, the reasoning associated with these theories is based on Peirce's (1931) notion of abductive reasoning, which involves inferences to the best explanation of observations. We have described how probability comparison theories may have been developed according to this reasoning, and have utilized a similar approach in reaching our own conclusions about responses given by prospective mathematics teachers when tasked with comparing the relative probabilities of given events. However, although there is growing evidence to suggest that the use of the fallacy of composition may best explain certain responses to relative probability comparisons, we can only use the term "best" in an abductive sense. In other words, we cannot claim with *certainty* that if an individual declares that, in this instance, given events are equally likely to occur that they have indeed done so by employing the fallacy of composition. The same applies, of course, for other theories and models related to probability comparisons, such as the representativeness heuristic, the outcome approach, and the equiprobability bias.

REFERENCES

Abrahamson, D. (2009). Orchestrating semiotic leaps from tacit to cultural quantitative reasoning: The case of anticipating experimental outcomes of a quasi-binomial random generator. *Cognition and Instruction*, *27*(3), 175-224. doi:10.1080/07370000903014261

Batanero, C., & Serrano, L. (1999). The meaning of randomness for secondary school students. *Journal for Research in Mathematics Education, 30*(5), 558-567, doi:10.2307/749774

Chernoff, E. J. (2012a). Logically fallacious relative likelihood comparisons: the fallacy of composition [Special issue: National Year of Mathematics]. Experiments in Education, 40(4), 77-84.

Chernoff, E. J. (2012b). Recognizing revisitation of the representativeness heuristic: an analysis of answer key attributes [Themed issue: Probability in Reasoning About Data and Risk]. ZDM - The International Journal on Mathematics Education, 44(7), 941-952.

Chernoff, E. J. (2009). Sample space partitions: An investigative lens. *Journal of Mathematical Behavior, 28*(1), 19–29. doi:10.1016/j.jmathb.2009.03.002

Chernoff, E. J. & Mamolo, A. (2015) Unasked but answered: Comparing the relative probabilities of coin flip sequence attributes. *Canadian Journal of Science, Mathematics and Technology Education*, *15*(2), 186-202. doi:10.1080/14926156.2015.1031410

Chernoff, E. J., & Russell, G. L. (2012). The fallacy of composition: Prospective mathematics teachers' use of logical fallacies. *Canadian Journal of Science, Mathematics and Technology Education*, *12*(3), 259-271. doi:10.1080/14926156.2012.704128

Falk, R. (1981). The perception of randomness. In C. Laborde (Ed.), *Proceedings of the Fifth International Conference for the Psychology of Mathematics Education* (pp. 222–229). University of Grenoble.

Jones, G. A., Langrall, C. W., & Mooney E. S. (2007). Research in probability: Responding to classroom realities. In F. Lester (Ed.), *Handbook of research on mathematics teaching and learning* (2nd ed., Vol. 2, pp. 909–955). Greenwich, CT: Information Age Publishing and National Council of Teachers of Mathematics.

Kahnemann, D., & Frederick, S. (2002). Representativeness revisited: attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, and D. Kahnemann (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, England: Cambridge University Press.

Kahneman, D., Tversky, A., (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology, 3*(3), 430-454. doi:10.1016/0010-0285(72)90016-3

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*(5), 392–414.

Lipton, P. (1991). *Inference to best explanation.* New York, NY: Routledge.

Peirce, C. S. (1931). Principles of philosophy. In C. Hartshorne & P. Weiss (Eds.), *Collected papersof Charles Sanders Peirce* (Vol. 1). Cambridge, MA: Harvard University Press. Retrieved from http://www.hup.harvard.edu/catalog.php?isbn=9780674138001&content=toc

Shaughnessy, J. M. (1977). Misconceptions of probability: An experiment with a small-group, activity-based, model building approach to introductory probability at the college level. *Educational Studies in Mathematics*, *8*(3), 295-316.

Shaughnessy, J. M. (1992). Research in probability and statistics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465–494). New York: Macmillan.

Tversky, A., Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131, http://dx.doi.org/10.1126/science.185.4157.1124