Watson, J., Callingham, R., & Kelly, B. (2007). Students' appreciation of expectation and variation as a foundation for statistical understanding. *Mathematical thinking and learning, 9* (2), 83-130.
Watson, J., Collis, K., & Moritz, J. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9* (1), 60-82.
Watson, J., & Kelly, B. (2007). Assessment of students' understanding of variation. *Teaching Statistics, 29* (3), 80-88.

AN INFORMAL FALLACY IN TEACHERS' REASONING ABOUT PROBABILITY

Egan J. Chernoff
University of Saskatchewan
egan.chernoff@usask.ca

Gale L. Russell
University of Saskatchewan
gale.russell@.usask.ca

The main objective of this article is to contribute to the limited research on teachers' knowledge of probability. In order to meet this objective, we presented prospective mathematics teachers with a variation of a well known task and asked them to determine which of five possible coin flip sequences was least likely to occur. To analyze particular normatively incorrect responses we utilized a brand new lens – the composition fallacy – instead of the traditional lenses and models associated with heuristic and informal reasoning about probability. In our application of the new lens we were able to determine that fallacious reasoning, not just heuristic reasoning, can account for normatively incorrect responses to the task. Given the success of the new lens, we contend that logical fallacies are a potential avenue for future investigations in comparisons of relative likelihood and research in probability in general.

The general purpose of this article is to contribute to the paucity of research on (prospective) teachers' knowledge of probability (Jones, Langrall & Mooney, 2007; Stohl, 2005). More specifically, the purpose of this article is to merge the established thread of investigations into comparisons of relative likelihood (e.g., Borovcnik & Bentz, 1991; Cox & Mouw, 1992; Hirsch & O'Donnell, 2001; Kahneman & Tversky, 1972; Konold, Pollatsek, Well, & Lohmeier, & Lipson, 1993; Rubel, 2006; Shaughnessy, 1977; Tversky & Kahneman, 1974; Watson, Collis, & Moritz, 1997) with a developing thread of investigations into prospective teachers' comparisons of relative likelihood (e.g., Chernoff, 2009, 2009a, 2009b).

In order to achieve the general and specific goals detailed above, prospective teachers, as has been the case in past research, were presented with five different sequence of heads and tails – derived from flipping a fair coin five times – and were asked to declare which sequence was least likely to occur. However, unlike previous research, we utilize a brand new lens to account for certain responses; we demonstrate that certain responses fall prey to the fallacy of composition (i.e., because parts of a whole have a certain property, it is argued that the whole has that property). Further, we contend that informal fallacies, in general, create a new research opportunity for those investigating comparisons of relative likelihood.

A Review of the Literature

In mathematics education (and psychology) research, comparative likelihood responses are categorized, in a broad sense, into two particular categories: correct responses and incorrect responses. While correct responses are, for the most part, associated with normative reasoning, different individuals account for incorrect responses differently. In particular, two types of reasoning – heuristic (Tversky & Kahneman, 1974, LeCoutre, 1992) and informal (Konold, 1989) – have dominated the research literature.

Heuristic reasoning

Psychologists Daniel Kahneman and Amos Tversky (1972) asked a group of individuals whether there would be more families with a birth order sequence (using B for boys and G for girls) of BGBBBB or GBGBBG. Further, the same individuals were asked whether there would be more families with a birth order sequence of BBBGGG or GBGBBG. Kahneman and Tversky argued that individuals who declared one sequence as less likely were reasoning according to the *representativeness heuristic*, where one "evaluates the probability of an uncertain event, or a sample, by the degree to which it is: (i) similar in essential properties to its parent population; and (ii) reflects the salient features of the process by which it is generated" (p. 431). In other words, the BGBBBB sequence of births was seen as less likely than the sequence GBGBBG because the ratio of boys to girls (in the parent population) is one to one and the BBBGGG sequence of births was seen as less likely than the sequence GBGBBG because BBBGGG did not appear random. The representativeness heuristic, in essence, provided a new interpretation of normatively incorrect responses to comparisons of relative likelihood.

Informal reasoning

Building upon certain task developments introduced by Shaughnessy (1977) (i.e., including the equally likely option and providing a response justification), Konold et al. (1993) introduced a different version of the relative likelihood task than had been seen in the past. For example, Konold et al. asked individuals "which of the following is the most likely result of five flips of a fair coin?" and provided them with the following options, "a) HHHTT b) THHTH c) THTTT d) HTHTH e) all four sequences are equally likely" (p. 395). Further, the researchers gave students a most likely version of the task followed by a least likely version. They found, for the most likely version, certain participants answered using *the outcome approach* – "a model of informal reasoning under conditions of uncertainty" (Konold, 1989, p. 59) – and for the least likely version subjects answered using the representativeness heuristic. The outcome approach represented, as had the representativeness heuristic earlier, a new interpretation of normatively incorrect responses to comparisons of relative likelihood.

Theoretical Framework

With a few exceptions (e.g., Abrahamson, 2009), there has been a lull in the past number of years in the creation and development of fresh perspectives and interpretations of normatively incorrect responses to comparisons of relative likelihood. Inspired by the notion that novel perspectives and interpretations to normatively incorrect responses to comparisons of relative likelihood have, in the past, established new domains of research (e.g., Konold's most likely version leading to the outcome approach), we introduce the fallacy of composition to account for certain responses to comparisons of relative likelihood. In doing so, contend that informal fallacies, in general, may provide a original research domain; however, given pagination limitations associated with the present article, we decided to limit our scope and, further, our theoretical framework. As such, our impending analysis of results will consist of one particular fallacy: the fallacy of composition.

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.,

Wiest, L. R., & Lamberg, T. (Eds.). (2011). Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Reno, NV: University of Nevada, Reno.

the parts) are sturdy; buildings (i.e., the whole) are made of bricks; therefore, buildings are sturdy (which is not necessarily true). Applying the fallacy of composition framework to existing research (for example, and more specifically, Kahneman and Tversky's (1972) research involving the birth of boys and girls where participants deemed the sequence BGBBBB less likely than the sequence GBGBBG): the birth of a boy or a girl (i.e., the parts) occurs at a ratio of 1 to 1; birth order sequences (i.e., the whole) are made of births of boys and girls; therefore, birth order sequences should have a 1 to 1 ratio of boys to girls, which is not necessarily true. Beyond the application of the fallacy of composition framework to existing research, as presented, we will also demonstrate, in the analysis of results, certain participants in our research associated certain properties of individual coin flips (e.g., the ratio of heads to tails) to be true for sequences derived from individual coin flips – demonstrating our new perspective for normatively incorrect responses to relative likelihood comparisons.

The Task

As presented in Figure 1 below, the task given to participants was both similar and different to the task utilized by Konold et al. (1993), which we will comment on, in turn.

Which of the following sequences is the *least likely* to occur from flipping a fair coin five times:

a) HHTTH
b) HHHHT
c) THHHT
d) HTHTH
e) THHTH
f) all five sequences are equally likely to occur
Justify your response...

Figure 1. The relative likelihood task

Fundamentally, and is the case with all subsequent research since Konold et al.'s version of the task (e.g., Cox & Mouw, 1992; Chernoff, 2009; Hirsch & O'Donnell, 2001; Rubel, 2006), the structure of our task is similar Konold et al.'s. For example, participants are asked to pick out a particular sequence as more or less likely or to declare that all sequences are equally likely to occur and, subsequently, are asked to justify their response. The minor differences, however, between our current version of the task and Konold et al.'s version lie in the number of sequences available to choose from and certain sequences have been replaced with others.

Participants

Participants in our research were (n =) 147 prospective mathematics teachers enrolled in a methods course designed for teaching elementary school mathematics. More specifically, the 147 participants were comprised of five classes, each containing approximately 30 students, which were all taught by the same instructor. Of note, the topic of probability had not been covered in the course they were taking part in at the time of the research and, further, it was determined, afterwards, that none of the individuals involved in the research had ever answered an alternative version of the relative likelihood task. Individuals were given as much time as necessary to complete the task.

Results and Analysis

As presented in Figure 2 below, the majority of participants' responses fell into three different categories.

Students	***HHTTH***	***HHHHT***	***THHHT***	***HTHTH***	***THHTH***	***equally likely***
Total (147)	***1***	***38 (26%)***	***1***	***15 (10%)***	***1***	***91 (62%)***

Figure 2. Breakdown of responses from the 147 participants

Most participants, that is, 91 of the 147 (or 62%), responded correctly, that each of the five sequences were equally likely to occur. (However, not all 91 participants who answered correctly provided proper or, for that matter, normatively correct response justifications.) Fifteen (or 10% of the) participants responded incorrectly that the sequence HTHTH was least likely to occur and 38 (or 26% of the) participants declared that HHHHT was least likely to occur. For the analysis of results and in order to bolster the claim that the fallacy of composition can be used to account for particular responses to comparisons of relative likelihood, certain response justifications from the 38 individuals that responded HHHHT as least likely to occur are analyzed in detail.

The Ratio of Heads to Tails

As seen in the responses of Dave and Jerry, presented below, they both reference the ratio of heads to tails in the sequences they are presented and, further, declare that the sequence B (i.e., HHHHT) is least likely to occur because it has a heads to tails ratio of 4 to 1, whereas all other sequences have a ratio of 3 to 2.

Dave: I think B is least likely to occur because the others are all 2 only B is 1.

Jerry: B, it's the only one with one T and four Hs, the rest have 3Hs and 2Ts.

However, neither Dave nor Jerry elaborate beyond "that" one sequence has a different ratio of heads to tails, which leaves one to infer as to "why" they would think a 4 to 1 ratio of heads to tails is less likely. Based upon previous research (e.g., Tversky & Kahneman, 1974), one may argue that Dave and Jerry find the equally likely sequences to not be equally representative and, further, that the sequence with a heads to tails ratio of 4 to 1 is less representative, because the parent population of coin flips would have a ratio of 1 to 1, and, thus, less likely. However, the eight responses presented below tell a slightly different story.

Emboldened in the response justifications below, three individuals mention, in one form or another, that the 4 to 1 ratio is less likely because of the notion of equiprobability.

Gustav: It's more unlikely to have 4 heads and one tail **because there is a 50% chance**.

Igor: **Because the likeliness of both is the same**. It's unlikely that out of 5 the ratio would be 4:1 instead of something like 2:3.

Farhan: **because it's 50/50** so it will probably be 3H2T or 3T2H so B has the least likelihood of happening (4H1T).

However, none of the three individuals explicitly reference where the equiprobability they are using is coming from, which is not the case with the justifications of Hermine, Keanu, and Ferdinand.

Hermine: **Because the chance for H and T are both ½**. So it will be least likely to have H, H, H, H, T.

Keanu: **Both heads and tails have an equal chance for getting flipped**. I think B as in B H had more chance.

Ferdinand: I think that HHHHT is the least sequence that is more likely to occur, **because the chances are when you flip a coin once, the possibility of getting a tail is ½ because there are 2 sides on a coin.**

From the responses above, the notion of 50-50, 1/2, 1 to 1 ratio or equiprobability used in their justifications is derived from an individual coin, which, as Keanu declares, "ha[s] an equal chance of getting flipped."

Taking their justifications one step further, Duncan and Andrew make reference to individual coin flips, which each have a 50-50 chance of landing heads and tails.

> *Duncan:* I chose B because there are more H's than T's and they are not spread out. **A likely guess would be a 50/50 chance for each, eg: H, T, H, T,...**
>
> *Andrew:* b is least likely because **if each flip gives a 50/50 chance of resulting in heads/tails**, it is more likely that in five flips more tails would result than one. I mean it is not as likely that one result would dominate over the other (but possible).

For both individuals, since each individual coin flip has a 50-50 chance of heads or tails, the sequence with a 4 to 1 ratio of heads to tails is least likely to occur because the ratio is furthest away from 50-50 or 1 to 1.

As found in past research, responses reference the ratio of heads to tails as the reason why the coin flip sequence with a 4 to 1 ratio (i.e., HHHHT) is least likely to occur; however, we break tradition from previous research and demonstrate that the fallacy of composition – not just Tversky and Kahneman's (1974) sample to parent population determinant of representativeness – can account for the above ten responses. For example, all responses declare that the ratio of heads to tails for flips of a fair coin is 1:1 (i.e., the bricks); further, they note that the sequence (i.e., the building) is comprised of five flips of a fair coin; therefore, the sequence (i.e., the building) should also have a heads to tails ratio of 1:1 (i.e., the bricks), which, simply, is not true. As such, the expectation of a 1:1 ratio of heads to tails for individual flips of a fair coin (not the sample to parent population) leads the individuals to declare that the sequence with a head to tails ratio of 4:1 (the furthest "away" from 1 to 1) is least likely. A similar situation is revealed when looking at those responses, which traditionally would have been analyzed according to Tversky and Kahneman's reflection of randomness determinant.

The Appearance of Randomness

Aaron, Candy, Zoni, and Geena, declare that the sequence HHHHT is least likely to occur because of the "long" run of heads. However, the responses do not get into detail as to why the run of 4 heads in a row make the sequence less likely.

> *Aaron:* The chances of getting the same one four times is least.
>
> *Candy:* because the chances of getting 4 same sides are small.
>
> *Zoni:* B. It's more unlikely to flip four of the same sides 4 times in a row.
>
> *Geena:* B. Because that has the most of one side of the coin, which makes it less likely.

Peter, on the other hand, provides some insight, in his response below, as to why the sequence with a run of 4 heads is less likely.

> *Peter:* There is less of a chance of getting the same answer four times in a row. It's more likely to get a variety of answers.

Peter, and perhaps Aaron, Candy, Zoni, and Geena are expecting some mixture of heads and tails and not a run of 4 heads in a row. Presented in terms of previous research, Peter and the other individuals are focusing on the sequences appearance of randomness (Tverksy & Kahneman, 1974). More specifically, the run of 4 heads in a row is not locally representative, which demotes the sequences appearance of randomness and, further, demotes its likelihood.

A large number of responses, however, provide a very particular reason for the why the run of 4 heads in a row in less likely to occur than the other sequences. Chiefly, the sequence HHHHT is not likely to occur because the fair coin, which is used for individual flips that make up each of the sequences, is equally likely to result in heads or tails. References to the equiprobability (i.e., the 1 to 1 ratio of heads to tails) associated with the coin (and the flip of the coin) are emboldened in the response justifications below. Further, the response justifications are categorized in two different (sub) groups.

The first group, made up of the responses from Velma, Wally, Nina, and Carol, all reference that the four heads appear in the sequence. Further, they argue that it would be less likely to have four heads in the sequence because, as Wally mentions, “each side has an equal chance.”

> *Velma:* **Because there are 2 sides the percentage is 50-50** so that means it is more likely to have 3 tails or 3 heads. 4 heads is possible, but if you flip a coin, I think 4 heads isn’t going to happen a lot.
>
> *Wally:* B **because each side has an equal chance**. Therefore, it’s very hard to flip a coin and get the same side 4 times.
>
> *Nina:* I first thought the answer was F, but my gut feeling was telling me the correct answer was B. **To consistently flip heads when you have equal opportunity to flip tails** has made me chose B.
>
> *Carol:* B is least likely because it would be hard to flip a coin 5 times in a row and have it land on heads **because coins are pretty much equally weighted throughout and there is just as much chance of it landing on tails as heads** and heads is probably heavier since the design has got more metal on it. 50/50 chance.

The above four responses, we contend, fall prey to the fallacy composition. For example, all four responses declare that the coin, which has two equally likely sides, has a 50-50 chance of landing on heads or tails (i.e., the bricks); further, all participants note that the sequence (i.e., the building) is made of flips of that fair coin; therefore, the sequence (i.e., the building) should also have a heads to tails ratio of 1 to 1. As was also seen in the ratio of heads to tails responses earlier, the equiprobability of the coin and coin flips (i.e., the bricks) correlates with an expectation of “heads to tails equiprobability” for the sequence, which is not found in the sequence with four heads and, as such, is deemed less likely.

The second group, made up of the response justifications from Quinn, Reno, Susan, Terry, and Uma, also reference the number of heads, that is, four, found in the sequence. They do, however, also reference that the four heads occur in a row. Further, all five individuals declare that the sequence with four heads in a row is less likely to occur because, as Quinn declares, “there is a fifty-fifty chance it will land on heads or tails.”

> *Quinn:* **Because there is a fifty-fifty chance it will land on heads or tails**. So for it to land on head four times straight is possible, but the least likely to happen.
>
> *Reno:* I think HHHHT is least likely to occur **because there is a 50% chance to land on H or T** so you probably wouldn’t get H 4 times in a row.
>
> *Susan:* B is least likely to occur because there is a very low chance to get 4 head in a row. **It is a 50% chance to get H or T** so it is the most unlikely sequence.
>
> *Terry:* The reason why is **because a coin has 2 sides, so there is a 50/50 chance of one or the other**, so it is unlikely for so many heads to appear in a row. It would be more likely to have 3 or 2 heads (or tails).
>
> *Uma:* I believe this is **because there is a fifty-fifty percent chance that it will land heads or tails** and 4 head in a row is very unlikely to happen.

All five responses above – which, in the past, would have been accounted for with the reflection of randomness determinant of the representativeness heuristic – we contend, are also falling prey to the fallacy composition. All responses declare that the coin, which has two equally likely sides, has a 50-50 chance of landing on heads or tails or has a heads to tails ratio of 1 to 1 (i.e., the bricks); further, the sequence (i.e., the building) is derived from flips of a fair coin (i.e., with a heads to tails ratio of 1 to 1); therefore, the sequence (i.e., the building) should also have a heads to tails ratio of 1 to 1 and, as such, the sequence with too many heads is deemed least likely to occur. As was seen in the ratio of heads to tails and in the "four of one side responses," the heads to tails ratio of the coin (i.e., the bricks) correlates with an expectation of the same heads to tails ratio for the sequence, which is not found in the sequence with four heads (in a row).

Concluding Remarks

Demonstrated in the analysis of results, the fallacy of composition accounts for certain responses (i.e., certain normatively incorrect responses) to the relative likelihood task (which we have introduced in this article). Also as demonstrated in the analysis of results above, all responses presented can be framed within the fallacy of composition. In particular, participants: make note of the 1 to 1 ratio of heads to tails of the coin (alternatively expressed throughout as: "fifty-fifty," "equiprobable," "the possibility of getting a tail is 1/2"); note that sequence is made up of flips of said fair coin; and, as such, determine (fallaciously) that the sequence of coin flips should also have a heads to tails ratio of 1 to 1. In other words, the properties associated with the fair coin (i.e., the brick), which make up the sequence (i.e., the building), are expected in the sequence. Subsequently, when looking at potential responses, the one sequence with a heads to tails ratio furthest away from 1 to 1 (i.e., the only sequence with a heads to tails ratio of 4 to 1 when compared to all others sequences with a heads to tails ratio of 3 to 2) is deemed the least likely to occur. The respondents are taking their knowledge of equiprobability in one context and transferring it to another without validation of equiprobability in the new and expanded context. It is in this reasoning that the fallacy of composition lays.

Discussion

Research involving comparisons of relative likelihood has, historically, been focused on accounting for individuals' responses – both correct and incorrect. Through investigating normatively incorrect responses, research has developed a variety of theoretical models (e.g., the representativeness heuristic, the outcome approach, and the equiprobability bias) to account for incorrect, sometimes incomprehensible, responses. In a more general sense, the incorrect responses have, in the past, been accounted for by contending that individuals were employing heuristic or informal reasoning in their justifications. In more recent years, there has been a lack of developments and new perspectives to response justifications associated with comparisons of relative likelihood. We have, however, in this article, presented a fresh perspective, the composition fallacy, to account for certain response justifications. This novel perspective, we contend, opens a new area of investigation for future research on comparisons of relative likelihood: the use of logical fallacies. While, in this article, we have demonstrated the descriptive power of the fallacy of composition, more research in the new domain will lead researchers to determine to what extent informal logical fallacies can describe response justifications to comparisons of relative likelihood and, in doing so, determine to what extent logical fallacies are a part of teachers' knowledge of probability. Through the identification of these logical fallacies within teachers' probabilistic knowledge, teacher educators can assess the origins of teachers' non-normative (e.g., heuristic, informal, fallacious) probabilistic reasoning. Once assessed, educators can then address the teachers' knowledge and model effective strategies for assessing and responding to their future students' non-normative probabilistic reasoning.

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.

Borovcnik, M., & Bentz, H. (1991). Empirical research in understanding probability. In R. Kapadia & M. Borovcnik (Eds.), *Chance encounters: Probability in education* (pp. 73-106). Dorecht, The Netherlands: Kluwer.

Chernoff, E. J. (2009). Explicating the multivalence of a probability task. In S. L. Swars, D. W. Stinson & S. Lemons-Smith (Eds.), *Proceedings of the Thirty-First Annual Meeting of the North-American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 653-661). Atlanta, GA: Georgia State University.

Chernoff, E. J. (2009a). Sample space partitions: An investigative lens. *Journal of Mathematical Behavior, 28*(1), 19-29.

Chernoff, E. J. (2009). The subjective-sample-space. In S. L. Swars, D. W. Stinson & S. Lemons-Smith (Eds.), *Proceedings of the Thirty-First Annual Meeting of the North-American Chapter of the International Group for the Psychology of Mathematics Education* (Vol. 5, pp. 628-635). Atlanta, GA: Georgia State University.

Cox, C., & Mouw, J. T. (1992). Disruption of the representativeness heuristic: Can we be perturbed into using correct probabilistic reasoning? *Educational Studies in Mathematics, 23*(2), 163-178.

Hirsch, L. S., & O'Donnell, A. M. (2001). Representativeness in statistical reasoning: Identifying and assessing misconceptions. *Journal of Statistics Education, 9*(2). [Online: *http://www.amstat.org/publications/jse/v9n2/hirsch.html*]

Jones, G. A., Langrall, C. W., & Mooney, E. S. (2007). Research in probability: Responding to classroom realties. In F. K. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning*, (pp. 909-955). New York: Macmillan.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Konold, C. (1989). Informal conceptions of probability. *Cognition and Instruction, 6*(1), 59-98.

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.

Lecoutre, M-P. (1992). Cognitive models and problem spaces in "purely random" situations. *Educational Studies in Mathematics, 23*(6), 557-569.

Rubel, L. H. (2006). Students' probabilistic thinking revealed: The case of coin tosses. In G. F. Burrill & P. C. Elliott (Eds.), *Thinking and Reasoning with Data and Chance: Sixty-eighth yearbook* (pp. 49-60). Reston, VA: National Council of Teachers of Mathematics.

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*, 285-316.

Stohl, H. (2005). Probability in teacher education and development. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 345-366). New York: Springer.

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

Wiest, L. R., & Lamberg, T. (Eds.). (2011). Proceedings of the 33rd Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Reno, NV: University of Nevada, Reno.

Watson, J. M., Collis, K. F., & Moritz, J. B. (1997). The development of chance measurement. *Mathematics Education Research Journal, 9*, 60-82.