Eurasia Journal of Mathematics, Science & Technology Education, 2016, **12**(4), 1009-1031

An Investigation of the Representativeness Heuristic: The Case of a Multiple Choice Exam

Egan J. Chernoff
University of Saskatchewan, CANADA
Ami Mamolo
University of Ontario Institute of Technology, CANADA
Rina Zazkis
Simon Fraser University, CANADA

•*Received 16 September 2015*•*Revised 16 November 2015* •*Accepted 13 December 2015*

By focusing on a particular alteration of the comparative likelihood task, this study contributes to research on teachers' understanding of probability. Our novel task presented prospective teachers with multinomial, contextualized sequences and asked them to identify which was least likely. Results demonstrate that determinants of representativeness (featured in research on binomial, platonic sequences) are present in the current situation as well. In identifying a variety of context-related features influencing teachers' choices, we suggest the context in which tasks are presented significantly influences probabilistic judgments; however, contextual consideration also provides researchers with potential difficulties for analyzing results. In addition, we identify strands for further research of contextual influence.

Keywords: answer key; context; multiple choice; probability; representativeness; sequences

INTRODUCTION

This research explores terrain at the intersection of two widely researched domains: the influence of contextualizing mathematics on individuals' understanding, and the mathematical knowledge and understanding of prospective teachers. Specifically, we focus our attention on a niche in the terrain – what probabilistic understandings of prospective teachers are elicited by a contextualized problem scenario? We formulate more specific research questions below, after we introduce the particular context scenario and task as the two (question and task) cannot be separated in our view. Our interest in this niche of contextual considerations for probability education is multi-fold: (i) probability is a concept which weaves its way through much of the curricula in many countries (e.g., Jones, Langrall and Mooney, 2007); (ii) probabilistic ways of reasoning tend to conflict

Correspondence: Egan J. Chernoff,
College of Education, University of Saskatchewan, 28 Campus Drive, Saskatoon SK S7N0X1, Canada.
E-mail: egan.chernoff@usask.ca
doi: 10.12973/eurasia.2016.1252a

ISSN: 1305-8223

with intuitive approaches and be influenced by context and age (e.g., Fischbein & Schnarch, 1997; Mamolo & Zazkis, 2014); (iii) the impact of contextualizing mathematics on individuals' interpretations of a problem is unpredictable (e.g., Beswick, 2011); and (iv) how teachers interpret a mathematical concept influences how they engage with it in their classrooms.

In this paper, we consider the responses of 59 prospective elementary and middle school teachers as they responded to a task which compared two multinomial sequences in the context of answer keys to a multiple choice quiz, as depicted in Figure 1. We present this task as a novel lens through which to analyse how specific, context-implied arguments may influence participants' justifications of relative likelihood. The design and rationale for the task stem from 40 years of historical development of comparative likelihood tasks which predominantly focus on familiar binomial sequences (e.g., coin flips) found in mathematics education and psychology literature. Our extension to multinomial sequences provides a situation which typically lies outside the familiar repertoire of prospective elementary and middle school teachers, and as such, plays in the somewhat treacherous terrain of muddled intuitive, contextual, and formal considerations in mathematics.

State of the literature

- There has been a limited response to the repeated calls (found in major research syntheses) for research on teachers' knowledge and beliefs about probability.
- There is an established thread of investigations, which has individuals compare the relative likelihood of binomial sequences of outcomes.
- The task found in research (in the fields of mathematics education and psychology) that has individuals compare the relative likelihood of binomial sequences of outcomes has – for over 40 years – focused predominantly on familiar binomial sequences (e.g., coin flips).

Contribution of this paper to the literature

- A novel task contributes a new perspective to the existing research by utilizing multinomial, contextualized sequences that relate explicitly to familiar ground in teaching.
- Determinants of representativeness extend from binomial to (contextual) multinomial sequences of outcomes.
- The avoidance of theoretical probabilistic considerations was, in this instance, influenced by the two characteristics of the novel task: multinomial choices and contextual setting.

Fischbein (1999) notes that in general, "for the teaching of mathematics, it is very important that the teacher understands the interactions between the intuitive, the formal and the procedural aspects in the processes of understanding" (p.28). As teacher educators, we might argue analogously that it is important for the teaching of mathematics teachers that we understand these same interactions, and further include the influence of contextualized interpretations and knowledge. As such, our research takes a step in this direction and highlights the context-related ideas that most strongly influenced participants' probabilistic judgments. Further, it draws attention to the challenges of researching in the intersection of these disparate forces, and their influence on, and manifestation in, mathematical knowledge. In what follows, we review research concerning probabilistic understanding of pupils and teachers, and situate our current study in the theoretical framing of the representativeness heuristic (Tversky & Kahneman, 1974). Participant responses are thus analysed, and we conclude with suggestions for avenues of future research.

BACKGROUND

Shaughnessy's seminal (1992) review of research in the teaching and learning of stochastics concluded with "a wish list" (p. 488) for future research. Included in the list, was a call for investigation into teachers' conceptions of probability. Fifteen years later, Jones, Langrall, and Mooney (2007), in a more recent review of research in the teaching and learning of probability, included "Stohl's (2005) review[, which]

Which of the following, Answer key 1 or Answer key 2, is least likely to be the answer key for a 10 question multiple choice math quiz? Explain your answer.

Answer key 1: A C C B D C A A D B

Answer key 2: C C C B B B B B B B

Figure 1. Answer key iteration of a comparative likelihood task

concluded that there had been limited response to Shaughnessy's call for research on teachers' knowledge and beliefs about probability" (p. 945).

Despite this dearth of research, there has been an emerging thread of investigations into prospective teachers' knowledge of probability (e.g., Chernoff, 2009, 2011; Chernoff & Zazkis, 2011) and, also, continues an established thread of investigations, which has individuals compare the relative likelihood of binomial sequences of outcomes (Borovcnik & Bentz, 1991; Cox & Mouw, 1992; Hirsch & O'Donnell, 2001; Kahneman & Tversky, 1972; Konold, 1995; Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993; Rubel, 2006, 2007; Shaughnessy, 1977, 1981; Tversky & Kahneman, 1974; Watson, Collis, & Moritz, 1997).

These findings notwithstanding, our interest is on the "determinants of representativeness" (Kahneman & Tversky, 1972, p.431) as they relate to multinomial, contextual sequences of outcomes. Alternatively stated, we are interested in the features of multinomial, contextualized sequences to which individuals attend as they compare the relative likelihood of two events. These "determinants" shed light on individuals' interpretations of probabilistic situations based on their perceptions of what "should be" more, less, or equally likely, and why. Kahneman and Tversky note that the "notion of representativeness is best explicated by specific examples" (p.432), and we discuss this in greater detail below. To the best of our knowledge, research that has looked at multinomial sequences has identified that a "representativeness bias" (Fischbein, 1999, p.26; see also Chernoff & Mamolo, 2015) exists, though it has not delved into the specific characteristics of samples or events that make them appear as representative to the observer (e.g., what makes something "seem more random" in a particular context). For instance, Fischbein and Schnarch (1997) examined the responses of students in grades 5, 7, 9, and 11, as well as prospective mathematics teachers, to a question on the relative likelihood of winning the lottery when one player has chosen 6 consecutive numbers (from a total of 40) and the other has chosen numbers at random. Participants were given the option of answering that one or the other had a greater chance of winning, or that their chances were the same. This question was deliberately chosen to correspond to misconceptions regarding representativeness, and they observed that correct responses improved "as the student ages and finally overcomes the primitive, global, intuitive heuristic of representativeness" (p.103).

In his work on intuition, Fischbein (1999) noted that they "are very sensitive to the influences of the context, especially because no logical, formal support intervenes" (p.36). As such, we wondered if, and how, a change in context might influence prospective teachers' reliance on a heuristic of representativeness. In general, the influence of contextualising mathematics has been difficult to pin down. Beswick (2011) notes that contextualising mathematics problems is "underpinned

by a particular view of the nature of mathematics… [and] it is a philosophical rather than empirical stance" (p.380). She reviews several studies which examine student responses to "context problems" (though none of them address issues in probability) and refers to the work of Boaler (1993) who described "contexts as potentially a bridge or a barrier in relation to transferring mathematical knowledge" (Beswick, 2011, p.373). Beswick calls for further research into the use of contextualised mathematics problems.

Acknowledging Fischbein and Schnarch's observation that, "Probability does not consist of mere technical information and procedures leading to solutions. Rather, it requires a way of thinking that is genuinely different from that required by most school mathematics" (1997, p.104), we see a need for further research into the influence of context on probabilistic judgements – particularly in those tasked with teaching these concepts.

THEORETICAL CONSIDERATIONS

According to Watson and Kelly (2009), "As the volume of research into probabilistic understanding has increased over the years, various types of reasoning have been identified in explaining inappropriate decision-making when choices are made among events in terms of relative likelihood" (p. 27). Also in this particular volume of research, various types of tasks, which have been denoted both collectively and independently as "the comparative likelihood task" (Chernoff, 2009, p. 19), have been utilized. Stemming from investigations involving the comparative likelihood task, hereafter referred to as the CLT, two particular types of reasoning dominate the literature. The representativeness heuristic (Tversky & Kahneman, 1974), which accounts for responses to least likely versions of the CLT (e.g., which of the following sequences is least likely to occur); and the outcome approach (Konold, 1989), which accounts for most likely versions of the CLT (e.g., which of the following sequences is most likely to occur). Given that our present research is based upon a least likely version of the CLT, we take some time to develop in detail the foundations of the representativeness heuristic.

The "heuristics and biases approach" began with a survey of 84 participants at the 1969 meetings of the Mathematical Psychology Society and the American Psychological Association" (Kahneman & Frederick, 2002, p.49) and the initial body of research pertaining to the heuristics and biases program quickly grew. Central to Tversky and Kahneman's initial research was the notion of judgment under uncertainty. Tversky and Kahneman (1974) raised the question: "How do people assess the probability of an uncertain event or the value of an uncertain quality" (p. 1124)? Whether referred to as judgment under uncertainty or intuitive judgments of probability (Tversky & Kahneman, 1974), subjective probabilities (Kahneman & Tversky, 1972) or probability estimates (Jones & Thornton, 2005; Kahneman & Tversky, 1972), "perhaps the most general conclusion, obtained from numerous investigations, is that people do not follow the principles of probability theory in judging the likelihood of uncertain events" (Kahneman & Tversky, 1972, p. 430). More specifically, Tversky and Kahneman (1974) found that "people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations" (p. 1124).

Tversky and Kahneman (1974) described three heuristics – representativeness, availability, and adjustment and anchoring – that are used in the subjective assessment of probabilities. Furthermore, the authors described particular biases associated with employment of each of the aforementioned heuristics. While a number of diverse tasks were used in the research, it became apparent that particular types of tasks were associated with particular heuristics: "One oft-used task that evokes representativeness asks subjects to compare likelihoods of

sequences of outcomes that have been generated by a binomial process" (Shaughnessy, 2003, p. 219). Alternatively stated, the representativeness heuristic is frequently observed in research employing the CLT.

Kahneman and Tversky, in examining how "people replace the laws of chance by heuristics" (1972, p. 430), produced an initial investigation into what they called the representativeness heuristic. According to their findings, an individual who follows the representativeness heuristic "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). Kahneman and Tversky denote the aforementioned features of the representativeness heuristic collectively as the determinants of representativeness, that is, "characteristics of samples, or events, that make them representative, and demonstrate their effects on subjective probability" (p. 431). Individually, there are two determinants of representativeness called the "similarity of sample to population" (p. 433) and the "reflection of randomness" (p. 433). Creating a nested notion of determinants, the authors further detailed determinants specific to each of the determinants of representativeness. The determinants of representativeness (i.e., similarity of sample to population and reflection of randomness) and their respective determinants (i.e., the determinants of similarity of sample to population and the determinants of reflection of randomness) are now detailed.

Kahneman and Tversky (1972) theorized that events are considered more probable when appearing more representative and, similarly, events are considered less probable when appearing less representative. In order to test their theory, the authors focused their attention on different variations of the CLT. Kahneman and Tversky developed and presented individuals with birth sequences that, while equally likely, might not be interpreted by the participants as "equally representative" (ibid.) Of the three sequences presented – GBGBBG, BGBBBB and BBBGGG – the sequence BGBBBB was considered less likely than GBGBBG because BGBBBB does not reflect the ratio of boys to girls found in the parent population. In other words, since there are two outcomes, in this instance, the determinant of the (determinant of the) similarity of sample to population is the expectation of having the same number of boys and girls in the sequence. Of note, this determinant of the similarity of sample to population determinant was not given a specific name, but is, instead, referred to by its determinants of representativeness name: the similarity of sample to population. Further, BBBGGG was deemed less likely than GBGBBG because BBBGGG did not reflect the random nature associated with the birthing of boys and girls. While the findings supported the authors' initial hypotheses – more representative sequences would be judged more likely – they also declared similarity of sample to population as a necessary but not sufficient determinant of representativeness. As a result, they further investigated (determinants of) the other determinant of representativeness: the reflection of randomness.

In an investigation designed to reveal determinants of the reflection of randomness (which is one of the two determinants of representativeness) Kahneman and Tversky (1972) declared that "two general properties, irregularity and local representativeness, seem to capture the intuitive notion of randomness" (p. 433). Alternatively stated, the determinants of the reflection of randomness are irregularity and local representativeness. To explicate this point, the authors demonstrated that alternating sequences, such as a perfect alteration of heads and tails, are too regular and, thus, would not correspond to the results of a random process. That is, a sequence such as HTHTH would, in essence, trigger the application of the determinant of irregularity because it does not appear irregular and, as such, must not be random. Local representativeness, on the other hand, is the

belief that "the essential characteristics of the parent population are represented not only globally in the entire sample, but also locally in each of its parts" (p. 434). That is, individuals, when examining a short sequence of coin tosses, would expect not only the ratio of heads to tails to be close to one, but also, short runs (e.g., having the same side of a coin continuously appear), which would directly correspond to frequent switches or alterations (e.g., having alternating sides interspersed with runs when flipping a coin). "People view chance as unpredictable but essentially fair" (p. 434) and are likely to, for example, expect the coin flip sequence THHTHT to be more likely to occur than HTHTHT.

The representativeness heuristic is "a key psychological contribution" (Jones & Thornton, 2005, p. 73) to probability education and, according to Shaughnessy (1992), probability research (past and present) in mathematics education has been heavily influenced by Tversky and Kahneman. For example, both the continuing research on the relative likelihood of sequences of outcomes (e.g., Chernoff, 2009, 2011; Borovcnik & Bentz, 1991; Cox & Mouw, 1992; Hirsch & O'Donnell, 2001; Rubel, 2006, 2007; Shaughnessy, 1977, 1981; Watson, Collis, & Moritz, 1997) and the research on the perceived randomness of sequences of outcomes (e.g., Abrahamson, 2009, Batanero, Green & Serrano, 1998, Batanero & Serrano, 1999; Falk, 1981, Falk & Konold, 1997; Green, 1983, 1988; Schilling, 1990; Toohey, 1995) consistently reference and confirm the importance of (1) the determinants of similarity of sample to population, which is a determinant of representativeness, and (2) local representativeness and irregularity, which are determinants of the reflection of randomness, which is a determinant of representativeness.

TASK DESIGN AND RATIONALE

As mentioned, the rationale behind the design of our version of the CLT is based upon the historical progression of the CLT found in mathematics education and psychology literature. Although the CLT has undergone many developments, many of the CLT's features have remained constant. As such, the developments and consistencies of the CLT are next commented on in turn.

Task developments

Although not the first individuals to conduct an experiment comparing the relative likelihood of sequences of outcomes (e.g., Cohen & Hansel, 1958), researchers of, and literature in, probability education consider – by acclamation – the early seventies research of Tversky and Kahneman (e.g., Kahneman & Tversky, 1972; Tversky & Kahneman, 1974) as the canonical research into the relative likelihood of sequences of outcomes. Kahneman and Tversky (1972) asked individuals whether (from a group of 72 families with 6 children) there would be more families with the birth order sequence of BGBBBB or GBGBBG. Further, in determining that "order information…is not simply ignored" (p. 34), they conducted a second, related version of the CLT, asking whether more families would have a birth order sequence of GBGBBG or BBBGGG. Alternatively stated, the initial investigation into the relative likelihood of sequences of outcomes was, actually, two versions of the CLT: A task was first presented that would address the similarity of sample to population determinant for representativeness and, second, a related CLT was presented to address the reflection of randomness determinant. These two versions of the CLT, as presented, would undergo a number of changes as they became a fixture in the field of mathematics education research.

Shaughnessy's (1977) research introduced two important developments to the CLT. First, in comparing the chances of occurrence of different sequences, Shaughnessy's version of the CLT gave students the option of choosing "(c) about the same chance" (p. 309) as one of the response items. Second, Shaughnessy asked

participants to "give a reason for your answer" (p. 309). Despite these two new developments, one thing remained the same: the CLT remained as two separate versions (in which the first version of the CLT addressed the similarity of sample to population determinant of representativeness and the second version addressed the reflection of randomness determinant). However, a single task that would address both determinants of representativeness was, soon, forthcoming.

Konold et al.'s (1993) iteration(s) of the CLT unified, for the first time, in one task, both the similarity of sample to population and the reflection of randomness determinants of the representativeness heuristic.

As seen in Figure 2, three of the options presented to participants have a ratio of 3 heads to 2 tails, while option c) has a ratio of 4 heads to 1 tail, which is consistent with earlier tasks investigating the similarity of sample to population determinant (of representativeness). Moreover, the options containing 3 heads and 2 tails, present a variety of switch and run considerations, which is also consistent with earlier versions of the CLT examining the reflection of randomness determinant. Further, Konold et al.'s (1993) versions of the CLT, that is, the task, adopts Shaughnessy's (1977) developments to Kahneman and Tversky's (1972) task, which not only provides the equally likely option, but "also asked subjects to provide a written justification for their answer" (p. 396).

Which of the following is the most likely result of five flips of a fair coin?

a) HHHTT

b) THHTH

c) THTTT

d) HTHTH

e) All four sequences are equally likely

Figure 2. Konold et al.'s (1993, p. 395) iteration of the CLT

Other than Chernoff's (2009) alteration, which maintains the same ratio of heads to tails in all four sequences presented (i.e., all sequences, for five flips of a fair coin, presented have 4 heads and 1 tail), all other iterations of the CLT found in mathematics education (e.g., Hirsch & O'Donnell, 2001; Rubel 2006, 2007) adopt the same framework as Konold et al.'s (1993) iteration of the task, which possess the following elements: (1) the most or least likely is to be identified; (2) addressing both the similarity of sample to population and reflection of randomness in one task; (3) presenting an equally likely option; and (4) having individuals justify their responses. Although the CLT has undergone many developments, as shown, many of its salient features have remained the same.

Historically, the response justification has garnered much of the attention from researchers. In fact, a number of researchers have critiqued the inferential nature of Kahneman and Tversky's (1972) initial research, which was due to the absence of a response justification opportunity in their versions of the CLT. Nevertheless, when Shaughnessy (1977) asked participants to supply a reason or justify their responses, not only was it "…possible to gain [deeper] insight into the thinking process of the subjects as they answered the questions" (p. 308), but, also, it reinforced the research results that were once inferred from Kahneman and Tversky's forced response items. Picking up on the power of the response justifications, other

research began to investigate the connection between sequence selection and response justifications.

Task consistencies

In addition to CLT features identified above, two other consistencies have remained despite CLT iterations and developments: (1) the use of binomial sequences and (2) the use of (what we will define below as) platonic sequences.

The majority of research on the relative likelihood of sequences of outcomes, which utilize the CLT (Chernoff, 2009, 2011; Cox & Mouw, 1992; Hirsch & O'Donnell, 2001; Kahneman & Tverksy, 1972; Konold et al., 1993; Rubel, 2006, 2007; Shaughnessy, 1977, 1981), thus far, utilizes, exclusively, binomial sequences. Although certain research does investigate sequences where the probability of success does not equal failure (e.g., Konold et al., 1993; Watson & Kelly, 2009), these, potentially, multinomial experiments, such as the rolling a six-sided die, are reduced to binomial sequences by, for example, painting the sides of the die one of two colors. Historically, research comparing the relative likelihood of sequences of outcomes has utilized but one of the following two types of CLTs: flips of a coin (e.g., HTHHTH) or the birth of males and females (e.g., BGBBGB).

To help frame our discussion, we now present two types of sequences, which have been denoted as platonic and contextualized (Chernoff, 2011). Platonic sequences, which currently dominate research literature, are characterized by their idealism. For example, a sequence of coin flips derived from an ideal experiment – where an infinitely thin coin, which has the same probability of success as failure, is tossed repeatedly in perfect, independent, identical trials – would represent a platonic sequence. On the other hand, contextualized sequences, which are less represented in the current research literature, are characterized by their pragmatism. For example: the sequence derived from the severed left and right feet (i.e., LLLLLR), which were washing up on the shores of British Columbia, Canada in late 2007 and early 2008 ("6th human foot found", 2008); the sequence of six numbers obtained when buying a (North American) lottery ticket (e.g., 4, 8, 15, 16, 23, 42); the answer key to a true or false mathematics quiz (e.g., TFTTF); the answer key to a multiple choice mathematics quiz (e.g., ACCBDCAADB); and others would represent contextualized sequences.

Despite researchers' current interest in platonic sequences, this platonicity, which Taleb (2007) explains can result in the "tendency to mistake the map for the territory" (p. xxv), of the sequences, which are now completely devoid of context, occurred over time. For example, in the 'beginning', Tversky and Kahneman (1972) incorporated a frequentist or experimental perspective with their sequence of six children, when declaring, in their CLT, that "72 families" (p. 34) were surveyed. As the CLT migrated from psychology to mathematics education, platonicity began to take hold. Shaughnessy's (1977) research witnesses the first explicit move towards idealism. In his version of the CLT, Shaughnessy declares, up front, that "the probability of having a baby boy is about 1/2" (p. 309) and, in the second version of his CLT states, "(same assumptions as [task 1])" (p. 309). However, the assumptions explicitly stated in Shaughnessy's version of the CLT become implicit in subsequent versions. By the beginning of the 1990s, platonification of the experiment and sequences, in essence the entire CLT, was complete. For example, the research of Chernoff (2009, 2011), Cox & Mouw (1992), Hirsch & O'Donnell (2001), Konold et al. (1993), and Rubel (2006, 2007) all utilize the following phrase: A fair coin is tossed x times; which, in these cases, is intended to mean (from a rationalist perspective) that an infinitely thin coin, which has the same probability of success as failure, is tossed repeatedly in perfect, independent, identical trials to produce the sequences of coin flips (or births of males and females). In other words, the map is mistaken for

the territory. However, despite the platonification of the CLT, pragmatic responses were a main-stay for individuals responding to the CLT and are part of the motivation behind our new iteration of the CLT.

Task and research questions

The domination of binomial and platonic sequences found in the literature may be limiting the investigation into the nature of participants' probabilistic considerations. As such, our novel CLT contributes a new perspective to the existing research by utilizing multinomial, contextualized sequences that relate explicitly to familiar ground in teaching. Our iteration of the relative likelihood task, that is, CLT, presented below, asked individuals to compare the relative likelihood of two different answer keys to a 10 question multiple choice quiz and to explain their answer. Further, in an attempt to allow for a flow of future investigations that parallel the historical developments of the binomial, platonic CLT and related research, we, with our task, start at "the beginning" by embracing certain features (e.g., forced response between two options, no equally likely option) found in early research involving the CLT.

Given our interest in teacher's responses and justifications to this task, we pose the following research questions to guide our investigation:

1. What determinants of representativeness are present in teachers' considerations of a contextualized comparative likelihood task? Do (and, if so, how are) prior findings regarding determinants of representativeness extend to multinomial sequences? Specifically, we focus on whether the similarities of sample to population, and the reflection of randomness in the sample, also influence judgments on the relative likelihood of two multinomial sequences. For example, we are interested in whether or not individuals expect each of the possible responses to appear approximately one quarter of the time, if four choices available.
2. What impact does the contextualizing of sequences have on participants' responses? In particular, what specific context-implied arguments are used in the justifications of responses? For example, do individuals think differently about the notion of independence when examining an answer key to a multiple-choice test as opposed to a sequence of coin flips?

Participants and data collection

Data for this study was gathered by asking participants, 59 prospective elementary and middle school teachers, that is, teachers of students (aged approximately 4 to 13 years old), to respond, in writing (with no time limitations) to the task in Figure 1 presented earlier. Consequently, the data is comprised of two components: choice of sequence and response justification.

While the participants' choice of sequence provides hard data, the response justifications will be the main focus of the impending analysis of results, which allows us "to account for….inconsistencies, [because] it is critical to understand the beliefs and reasoning processes that underlie the various answers that subjects give" (Konold et al., 1993, p. 393). We now turn to the results and analysis.

RESULTS

The results show that the majority (81%) of the 59 participants in our study chose answer key 2 (hereafter referred to as AK2 and, similarly, answer key 1 as AK1) as least likely to be the answer key for a 10 question multiple-choice quiz. More specifically, as seen in Table 1, 23 participants (74%) in Class A and 25 participants (89%) in Class B chose AK2 to be least likely.

Table 1. Numerical breakdown of responses

Class	AK1 least likely	AK2 least likely	AKs equally likely
A (31 students)	6*	23	2
B (28 students)	1	25	2
Total (59 students)	7 (12%)	48 (81%)	4 (7%)

The seven responses, which indicated AK1 was least likely, were comprised of 6 individuals from Class A and one individual in Class B. However (note the asterisk in Table 1), 3 of the 6 six AK1 least likely responses from Class A were, based upon their response justifications, deemed invalid. As such, 4 (not 7) of the 59 participants (7%) responded AK1 as least likely to occur. Furthermore, despite not being presented as an option in our iteration of the task, 4 of the 59 participants (7%) responded that AK1 and AK2 were equally likely to occur.

While a numerical breakdown of the responses of each class is presented in Table 1, given the similar themes in the identified responses, we have chosen not to distinguish between the two classes in our analysis of said results.

ANALYSIS OF RESULTS

In what follows we organize the data around our two research questions and analyze: (i) the influence of multinomial sequences on determinants of representativeness, and (ii) the influence of context on determinants of representativeness. As noted earlier, Kahneman and Tversky (1972) declared two determinants of representativeness: similarity of sample to population and reflection of randomness. As such, our analysis of both trends (i) and (ii) similarly parses responses based on these two determining factors. Given the distribution of data presented in Table 1, we focus primarily on response justifications which identified AK2 as least likely, and identify instances where the same considerations led to a different conclusion. We note that only one of the 59 participants gave a response which hinted at an argument based on theoretical considerations of probability, identifying the two events as equiprobable. Zonie wrote that "They are both [AK1 and AK2] equally likely because there is the same amount of options (A, B, C, D) for both answer key A and B." While the data is insufficient to conclude with certainty that Zonie took a theoretical approach, we do note that he was the only participant who seemed to distinguish between the options available to choose (A, B, C, D) and the options chosen (e.g., C, B, for AK2) when determining the relative likelihood of the two events. We further note that of the considerations that focused on expectations for multinomial sequences or for context, only the latter yielded responses of "equally likely" or "AK1 least likely".

Determinants of representativeness and multinomial sequences

As evidenced below in the response justifications for ten particular individuals who deemed AK2 as least likely, "the features that determine the similarity of a sample to its parent population" (Kahneman & Tversky, 1972, p. 33) extend, we contend, from binomial to multinomial sequences of outcomes. While all nine individuals declared the proportion of multiple choice answers presented in (the sample) AK2 (i.e., 3 Cs and 7 Bs) does not reflect the appropriate proportion of multiple choice answers for the population, different individuals presented the notion in different ways, as seen in the responses from Adam and Ben.

Adam: AK2 because there is too little variety of answers.

Ben: AK2 is least likely because it has only C's and B's!

Other responses, such as those of Donna and Eva, presented below, declare, probabilistically, that the sample of multiple choice answers witnessed in the

answer key does not reflect the proportion of multiple choice answers one would find in the population.

> *Donna*: AK2 does not seem as a fair representation of probability
> *Eva*: because the probability of having no A's or D's is slim.

We infer the use of the term "probability" found in the responses above is used to declare, implicitly, that in the population a "fair representation of probability," would have an equal or nearly equal number of each of the multiple choice answers; because each answer has a 25% chance of occurring and, as such, each answer should be found approximately 25% of the time.

While implicitly stated in certain responses, Frank, Gary, Harvey, and Ike are explicit in their use of percentages and expected frequencies when declaring that an answer key would have an equal distribution of available answers.

> *Frank*: There are 4 possible letters, so each should show up around 25% of the time. This is true of AK1 (A:30% B:20% C:30% D:20%). So AK2 (C:30% B:70%) is least likely.
> *Gary*: The probability of AK2 is unlikely because there is only a 25% chance of the same letter to come up again. Since there is a 75% chance of different letters coming up, it is more likely that a different letter will come up, not the same letter repeated.
> Harvey: assuming answer key is selected randomly, high frequency of same answers unlikely.
> *Ike*: AK2 upon first impression because it doesn't appear to be random, since each question has 1 in 4 chances of either being A, B, C, or D. The likelihood/probability of that occurring is low.

For us, Ike's response epitomizes the similarity of sample to population determinant of representativeness when extended to non-binomial sequences. For Ike, AK2 is less likely than AK1 because AK2, by not having an even distribution of answers A, B, C and D, "doesn't appear random". We contend that AK2, with only Bs and Cs, is not representative of the similarity of sample to population and is influencing Ike's perception of randomness, which, in turn, is influencing his probabilistic judgment because, according to Kahneman and Tversky's (1972) confirmed hypothesis, less representative sequences are deemed less likely.

Of note, "To be representative, it is not sufficient that an uncertain event be similar to its parent population. The event should also reflect the properties of the uncertain process by which it was generated, that is, it should appear random" (p. 35). As such, we now shift our attention to the reflection of randomness determinant of representativeness and multinomial sequences.

As was the case with the similarity of sample to population determinant of representativeness, the reflection of randomness determinant of representativeness extends from binomial to multinomial sequences of outcomes. Kahneman and Tversky's (1972) notion of local representativeness, which states "a representative sample is one in which the essential characteristics of the parent population are represented globally in the entire sample, but also locally in each of its parts" (p. 36), influences the relative likelihood of the answer keys for two participants, whose answers are featured in the responses below.

> *Jen*: AK2 [...] because the probability of sequential answers being identical is low.
> Kate: AK2 is the least likely because there would not be so many "B" answers in a row.

For Jen, the chances of "sequential answers being identical," i.e., 3 Cs and then 7 Bs, is not representative for part of the answer key and, thus, not likely to be the answer key. Similarly for Kate, "the so many 'B' answers in a row" is not locally representative and, thus, she concludes AK2 is least likely to be the answer key.

Although stated differently by Jen and Kate, for both, the long run of 7 Bs found in the latter section of AK2 is not locally representative. That is, the essential characteristics of the parent population are not found in the BBBBBBB section or part of the answer key. As such, the entire sequence, which, for them, does not appear random, is not representative of a 10 question multiple-choice answer key. Given the answer key is not seen to be locally representative, AK2 is considered less likely to be the answer key.

The reflection of randomness determinant of representativeness, however, is one of two general properties that "seem to capture the intuitive notion of randomness" (Kahneman & Tversky, 1972, p. 35). To determine whether or not the reflection of randomness determinant of representativeness, i.e., irregularity based on the absence or presence of a pattern, extends from binomial, platonic sequences of outcomes, we now examine responses under contextual considerations.

Determinants of representativeness and contextual sequences

According to Kahneman and Tversky (1972), "As is true of the similarity of sample to population, the specific features that determine apparent randomness differ depending on context" (p. 35). As such, we now examine the determinants of representativeness, both similarity of sample to population and reflection of randomness, by taking into account certain contextual considerations associated with our answer key task from a variety of perspectives, including: answer key, personal experience, teacher, student, and combined (or multiple) perspectives. The main goal in demonstrating the diverse perspectives is presenting the subject of our investigation (i.e., determinants of representativeness and contextual sequences) in a greater context.

The responses from six individuals that deemed AK2 as least likely took the answer key into consideration and, in doing so, provided us with insights into personal perceptions associated with a perceived innate structure of answer keys. For example Mary and Oliver (like Jen and Kate above) reference the length of runs for answers C and B, which are found in AK2. However (unlike Jen and Kate), both respondents qualify that long runs are not an innate feature to answer keys.

> *Mary*: answer keys usually do not have a constant answer for consecutive questions in a row. They are usually mixed up upon each other and will occasionally be in a row.
>
> *Oliver*: AK2 is least likely because there is far more repetition (3 C's / 7B's) than one would normally find on a typical multiple choice quiz.

Similarly, the justification given by Quinn, like that of Adam and Ben above (and others), makes reference to the lack of variety of answers found in AK2. However, unlike Adam and Ben (and others), Quinn also qualifies that the lack of variety is not an innate feature of answer keys.

> *Quinn*: #2 because it seems there's only two possible answers for this quiz – not multiple choice.

For all six of the above respondents, the specific features that demote the likelihood of AK2 are the long run of Bs found in the latter half of the answer key and the lack of variety amongst the available answers. According to prior research, the long run of Bs and lack of variety would be attributed, respectively, to the local representativeness and similarity of sample to population determinants of representativeness. However, as evidenced in all responses justifications above, the specific features demoting the likelihood of AK2 are based on personal perceptions of features innate to answer key structure. As a result, it becomes difficult to discern, in this instance, whether the determinants of representativeness, perceived answer key structure, or both, are the root cause for these individuals choosing AK2 least likely. However, what can be discerned for these individuals is that the perceived

structure of multiple choice answer keys epitomizes the similarity of sample to population and local representativeness.

Instead of qualifying responses with references to the structure of answer keys, six individuals who deemed AK2 as least likely made reference to their personal experiences with answer keys, and, further, provided insight into their experiences. As eloquently stated by Raymond:

> *Raymond*: Every multiple choice answer key over a twenty-year career in academics has looked more like AK1. I am using probability to make an educated guess that AK2 is less likely.

The reference to a lack of variety, seen below in the responses of Sue and Tara, and the reference to the long run of the one answer, seen below in the response of Uma, exemplify responses from other individuals who referred to personal experience with answer keys.

> *Sue*: I say that AK2 is least likely mostly for the fact that I have personally never had an exam (like this) where only 2 answers are correct.
>
> *Tara*: From my experience with multiple choice exams, the answers never line up one after the other, like in AK2. The multiple choice exams I studied for, such as math, have always looked more like AK1, where there is a variety of answers such as ACCBD instead of CCCBBB.
>
> *Uma*: there are too many answers that are the same ex. cccbbb. This (as a student) always made me confused. If the answers are all in a line like that, it makes the student feel like they did something wrong.

Sue, Tara, and Uma describe specific features that demote the likelihood of AK2 relative to AK1, which, once again, are "too many answers that are the same" and the lack of variety in the answers.

According to established research, the local representativeness and similarity of sample to population determinants of representativeness would attribute for the response of AK2 being deemed less likely in this instance. However, their response justifications indicate that their reasons for demoting the likelihood of AK2 are based on a perceived innate structure of answer keys, which is based upon personal experience or involvement, in some instances for more than 20 years, with answer keys. As a result, it becomes complicated (especially given the isomorphic similarities between (1) the determinants of representativeness applied to answer keys and (2) the perceived innate structure of answer keys) to discern whether or not the determinants of representativeness account for the responses. Perhaps determinants of representativeness may account for the responses, but, perhaps, the perceived innate structure of answer keys accounts for the responses – or both. Nevertheless, it is possible for us to assert that, for these individuals, the innate structure of answer keys, based on personal experience, epitomizes the similarity of sample to population and local representativeness (which is a determinant of the reflection of randomness) determinants of representativeness.

Six individuals (three who deemed AK2, one who deemed AK1 as least likely, and two who deemed AK1 and AK2 were equally likely) projected themselves into a teacher's perspective with their response justifications. Their responses provided unique insight into how they perceive teachers' use of and experience with answer keys.

> *Val*: AK2 because there is a more likely variation in answers provided by the teacher.
>
> *Wendy*: Normally teachers or instructors who set up answer keys tend to highlight a number of letters and use variation. A teacher would very rarely choose B to represent an answer 7 times in a row, as they usually seem to make random answer keys according to correct letters.

> *Xavier*: AK2 as test makers are (possibly) less likely to put strings of answers with the same letter.

On the one hand, Val, like those individuals whose responses are detailed and analyzed above, makes reference to the lack of variety found in the AK2. Further, she declares that the lack of variety would not exist because of the teacher, who would take the variety of multiple-choice answers into account during her making of the quiz. Wendy and Xavier, on the other hand, note that there would not be a string or run of identical consecutive answers because this is not something that "instructors who set up answer keys" or "test makers" would do. In contrast, Anwar also interpreted the task in terms of what a teacher would do, but his pragmatic considerations had a different flavour:

> *Anwar*: I think AK1 would be the least likely. Why? Because it is easier to correct, the questions can be altered to fit the numbers, so it is easier to correct.

The lack of variety and long run of one answer, seem to be interpreted as a positive feature on one hand, and a negative one on the other, depending on whether the teacher's perspective is one of the test maker or the test marker. While Anwar justifies his response by referring to what would be convenient when using an answer key, Val, Wendy, and Xavier justify their responses by referring to how they believe teachers construct answer keys. Similarly, the respondents who answered "equally likely" also focused on the creation of answer keys.

> *Lydia*: I think both answer keys are likely. I want to say AK2 is least likely because I think the person designing the test would feel inclined to mix up the answers. BUT either answer key is acceptable because if the students know their test material they will feel confident answering 7 B's in a row.
>
> *Marnie*: My answer is that they are both equally probable, because a multiple choice quiz is created by a teacher who can create any pattern of letters he or she wishes. The letter/answers are not random.

Both Marnie and Lydia bring in a perspective of a teacher who can design the test as "he or she wishes". In addition, Lydia also acknowledges a perspective of a student, who should not get confused by "7 B's in a row". That is, while Lydia's decision is that both answer keys are equally likely, when referring to a student perspective she acknowledges the unconventional distribution of answers in AK2. We discuss such an approach further below.

For all three options – AK1 least likely, AK2 least likely, or equally likely – determining whether lack of variety and long runs, for these individuals, are a result of the determinants of representativeness or the perceived innate structure of answer keys based upon how teachers interact, work with, create, and implement answer keys is difficult to ascertain. The similar results which are achieved by (1) applying determinants of representativeness to answer keys and (2) the perceived innate nature of answer keys, in this instance, as shaped by a teacher's interactions with answer keys, make it difficult to attribute the response justifications to one reason and not the other. A key difference in the cases is that although all of these responses seemed to be influenced by participants' perspectives related to the innate structure of answer keys, based on perceived teacher interactions, Marnie and Lydia offered perspectives that align with probabilistic theory, while the others did not. In continuing our analysis of the determinants of representativeness for contextual sequences, we turn our attention to the second of the reflection of randomness determinants: irregularity (i.e., absence or presence of pattern).

Kahneman and Tversky (1972) note that "A major characteristic of apparent randomness is the absence of systematic patterns" (p. 35), and that regularity does not equate with representative sequences. This pattern characteristic was present in the response justifications of seven individuals who took a student's perspective to

their response justification (six of whom deemed AK2 as least likely). Similar to previous perspectives analyzed, their responses provide insight into how they perceive students' use of, and experience with, patterns and answer keys. For example, and as exemplified with the two responses below, references are made to the presence of a pattern for AK2 and the absence of a pattern for AK1, which, according to the irregularity (which is a determinant of the reflection of randomness) determinant of representativeness implies that AK2 was deemed less likely than AK1.

> *Aaron*: Usually there isn't a pattern to the answer key. When students recognize that the answer has been B for the last few questions then they will tend to just pick B for the next ones without reading/fully answering the questions.
>
> *Doug*: I think AK2 is least likely to be the answer key because there is only 2 lines going straight down. AK1 has a zig-zag and it just seems better to have the answers all over rather than a boring pattern. Everyone knows the answers don't follow a pattern, if they did, everyone would get the answers right.

The above responses deviate from previous perspectives because they do not, at least explicitly, reference the small variety of available answers presented and the long run of Bs. Instead, the respondents qualify that patterns are a feature not innate to multiple-choice exams because students will be able to pick up any pattern that exists and 'compromise' the quiz or exam. Therefore, multiple choice quiz answer keys containing a pattern are less likely than answer keys not containing a pattern, because of students' acute ability to pick up on the pattern and, thus, the integrity of the test is compromised. Interestingly, while Andrew also alluded to similar considerations, he came to the opposite conclusion, citing his expectations for a student who follows a consistent choice.

> *Andrew*: AK1 is least likely because the responses are random. In AK2 the responses are more likely to be consistent, because in multiple choice people would pick their first choice which is usually pattern forming.

It is possible that Andrew interpreted the task as a relative comparison between answer sets rather than answer keys and either did not realize, or did not attend to, the intended meaning of an answer key. Thus, while his response contextualized the problem in terms of general pattern-seeking behaviour of individuals, and particularly the answers a student might give, the other respondents made explicit reference to their expectations for answer keys in light of their expectations for students' pattern-seeking behaviour. That is, Aaron and Doug both seem quite certain that students would "just pick B" if they had not learnt to ignore this impulse from prior experiences with answer keys which "don't follow a pattern."

The above responses do not deviate from previous perspectives because they, too, are based on how the respondents' perceive the innate structure of answer keys, which, in this case, is shaped by students' interactions with answer keys. The justifications from Aaron and Doug show that there was a tendency to describe patterning features that demote the likelihood of AK2 relative to that of AK1. More specifically, they reference the presence of a pattern in AK1 and the absence of a pattern in AK2. According to prior research, the presence of a pattern would promote likelihood of AK1, whereas the absence of a pattern would demote the likelihood of AK2. As such, AK2 would be deemed less likely relative to AK1. However, their response justifications indicate that the feature demoting the likelihood of AK2 (i.e., the pattern) are based on a perceived innate structure to answer keys, which is shaped, in this instance, by student involvement, interaction, and experience with answer keys. Consequently, there are difficulties (especially

given the similarities between the irregularity determinant of representativeness applied to answer keys and the perceived innate structure of answer keys shaped by students' perspective) in determining whether or not the determinants of representativeness can account for those choosing AK2 to be least likely. Perhaps the perceived innate structure of answer keys, shaped by taking a student perspective, accounts for those individuals choosing AK2 to be less likely. Perhaps the former and the latter are both in play. Despite the difficulty in discerning between the two, it is possible to assert that, for those individuals taking a students' perspective, the innate structure of answer keys epitomizes, in this instance, the irregularity (which is a determinant of the reflection of randomness) determinant of representativeness.

Thirteen individuals, who also chose AK2 to be least likely, combined many of the perspectives, detailed above, in their response justifications. In doing so, their responses paint a more intricate picture of their perceived innate structure of answer keys.

As seen in the responses from Fred and Igor, while both individuals take multiple perspectives, each individual takes different perspectives into consideration. For example, in the response from Igor, his reason for AK2 being least likely details an interaction between teachers' and students' perspectives. On the other hand, the response from Fred includes teachers' perspectives and answer key structure considerations. Further (from Fred's response), the interaction between teacher and answer key structure presents a notion of a difficulty level to the answer key, which is based on how "spread out" you would make the answers.

> *Igor*: AK2 is least likely because answers to numbers 4 to 10 are all circled B. In multiple choice quizzes, I do not think that there are answers for questions that are in the same category. I find that teachers will not make the answers on the same letter, maybe they will only put it twice like AK1; numbers 7 and 8. I think all teachers like to switch it up just because the students will not answer the same.
>
> *Fred*: AK1 would most likely be the answer key because in AK1 the answers are probably more spread out or you would make the answers too easy to find if they were like AK2. I think I would try to make the answers to a multiple choice quiz a bit confusing to try and make the quiz a little more difficult.

While the underlying motivation for Igor and Fred's perceived innate structure of answer keys (expressed implicitly in their responses) can be inferred from the responses they gave, the responses from 5 individuals, seen below, are rather explicit in describing the underlying motivation for their perceived innate answer key structure, which, again, while focusing on different, multiple perspectives, ends with similar results.

> *Gale*: I've never seen it done that way before. If it's done like AK2 students will second guess themselves because they would assume teachers would make it more like AK1
>
> *Mike*: AK2 is least likely. Answer keys go something like a rhyming scheme: ABBACC etc. A "teacher" would never give so many consecutive correct answers under the same letter. It would both corrupt the integrity of the test and play mind games with the student.
>
> *Nona*: AK2. Teachers usually not put the correct answer on the same letter. But AK1 could also be the least likely because the teacher wants to trick you to think that they couldn't be on the same letter.
>
> *Bertrand*: I think AK2 is least likely because it is too boring the answers are not mixed up enough. This answer key would also be tricky because the student might think "it can't be answer b for six times in a row" and

then they might second guess themselves, and pick different answers. AK1 seems to have a bit more variety.

Jo: AK2 is least likely because teachers are unlikely to make a pattern like this in the answer key. It throws students off.

Despite the diverse perspectives presented in the responses of Gale (who combines personal, student and teacher perspectives), Mike (who combines answer key, teacher and student perspectives), Nona and Jo (who both combine teacher and student perspectives), or Bertrand (who combines answer key and student perspective), all 5 present a similar underlying motivation for their perceived innate structure of answer keys (not represented within AK2) because, as Mike says, it would "play mind games with the students". As such, it appears, at least for these 5 individuals (and for Fred and Igor), and independent of the number and combination of different perspectives (answer key, personal, teacher, and student) taken into account, that answer keys innately: (1) possess a variety of answers, (2) do not have many consecutive identical answers, and (3) are not too regular. Despite the variety and number of perspectives that could be taken into account, all individuals convene on a similar underlying motivation for why an answer key is structured the way it is: to not throw students off, have students second guess themselves, or play mind games with or trick the student.

The above responses do not deviate from previous perspectives because they, too, are based on how the respondents perceive the innate structure of the answer key, which, in this instance, is shaped by combinations of multiple perspectives. As seen above in all seven responses, references are made to a variety of answers, the long run of consecutive answers, and the presence of a pattern, which demote the likelihood of AK2. However, their justifications indicate that the demotion of the likelihood of AK2 is based on their perceived innate structure of answer keys, which, for some, is based upon an underlying motivation of preserving the trust between teacher and student interactions via creation and completion of an answer key. Resultantly, we contend there are difficulties when determining whether it is the determinants of representativeness or the perceived innate structure of answer keys that account for those respondents choosing AK2 as least likely. Perhaps the perceived innate structure of answer keys, shaped by taking combinations of multiple perspectives accounts for those individuals choosing AK2 less likely or, perhaps, determinants of representativeness account for AK2 being chosen least likely. Despite the difficulty in discerning between the two options, it is possible for us to assert that, for those individuals taking a variety of combinations of multiple perspectives, the innate structure of answer keys epitomizes, in this instance, the similarity of sample to population determinant of representativeness, and local representativeness and irregularity, which are determinants of the reflection of randomness, which is a determinant of representativeness.

DISCUSSION

As stated earlier, we presented participants with equally likely, yet not equally representative, sequences of outcomes. Equally likely because, theoretically speaking, the answer keys presented to participants are equally likely to occur. In fact, each of the 1,048,576 ($= 4^{10}$) possible answer keys are equally likely to occur. Despite this fact, and although some individuals did identify both answer keys as equally likely, the majority of participants in this study (over 80%), when presented the two multinomial, contextual sequences, indicated that AK2 was least likely to be the answer key for a 10 question multiple-choice exam. The numeric results of our study contrast sharply with what might be expected given the results of Fischbein and Schnarch's (1997) work, and it is worth taking a moment to explore why this might be the case. Whereas Fischbein and Schnarch found that when given the

option, the majority of prospective teachers (78%), grade 11 students (65%) and grade 9 students (65%) were able to correctly identify two equally likely, yet differently representative, sequences of lottery picks, our results are not so encouraging. Considering both the tasks and the methodologies of the two studies, we note a few of the differences: (i) the lottery scenario relies on conditional probability, while the events in the answer keys scenario are independent; (ii) lottery chances are a common context for probability questions in school mathematics, while answer key generation is not; (iii) the lottery scenario provided the option of "equally likely" as one of the possibilities and sought only participants' choice amongst the options without asking for any explanation, while the answer key scenario excluded the option of "equally likely" but required choices to be justified and explained; and (iv) the mathematical backgrounds and preparations of the prospective teachers are different. Regarding the last point, while it may be expected that teachers specializing in mathematics may be better versed in the subject matter than prospective elementary and middle school teachers, we may assume at least as much familiarity with the subject matter as would have a grade 9 student – the numbers, thus are still somewhat surprising. The third point is one for which more research is needed. It is possible that the non-option of "equally likely" influenced participants' responses, and studies which attend to the differences of including or not such an option for a contextualized, multinomial problem can shed further light on the similarities and differences between interpretations of multinomial versus binomial sequences.

Regarding the first and second points, we suggest that it is the familiarity of the context which may more significantly influence perceptions of likelihood – though we acknowledge that a more familiar context could lead to more experience with the content involved (e.g., conditional probability).

We consider the influence of context further in the address of our research questions below, though we draw attention to the work of Boaler (1993) who noted that including context problems in mathematics curriculum is not enough to solve the problem of transferring learning. That is, in our case, even extended experiences with multinomial sequences encountered in familiar school contexts such as lottery tickets and dice rolls may not position individuals to better understand the underlying (theoretical) probability concepts when extrapolated from the usual scenarios. Adding a further complexity in our research with prospective teachers is the contextualization of mathematics in situations of teaching and learning. In resonance with Chernoff and Mamolo (2015), we observed a tendency for participants to interpret the context with respect to what they as teachers might do in a similar situation, though this was not requested of them nor alluded to in the phrasing of the task. Comments that referred to what a teacher "would feel inclined to do" in response to the context seemed, on one hand, to shift focus away from a response to the mathematics. On the other hand, our context may offer a type of authenticity which is absent from (e.g.) lotteries and dice, and which, as such, may provide for researchers a distinctive window into the resilience of the (intuitive) heuristic of representativeness. Kramarsky, Mevarech, and Arami (2002) consider the authenticity of a task to be dependent on the solver, and they suggest that the usual word problems encountered in school are not authentic because of the readily available algorithms that may be applied in the problem solving. For prospective teachers, we suggest that embedding mathematics into the daily routines of teaching and learning may offer such an authenticity, which in turn may account for some of the discrepancies between our results and those of Fischbein and Schnarch (1997). Fischbein (1999) contests that intuitive heuristics are "controlled, in fact, tacitly by a principle, an intellectual attitude, a structural schema… [and if] the schema – elicited by some salient (but not essential) data… is not adequate to the essence of the problem, one should find that the intuition worsens with age" (p.47). While he states

that awareness of the adequate principle for tasks with multinomial sequences increased with age, we found only one participant who alluded to the independence of probabilities in his reply (Zonie) while the others relied on inadequate intuitive heuristics influenced by informal and contextual perceptions of representativeness. This observation brings us back to our original research questions, which we now review.

In addressing our first research question (What determinants of representativeness are present in teachers' considerations of contextualized comparative likelihood task? In particular, how do determinants of representativeness extend to multinomial sequences?) we found that participants, in justifying their choice of AK2 being least likely, acknowledge the determinants of representativeness by expecting an answer key to have no explicit pattern, to include variety of choices and to have each choice appear approximately equal amount of times. Several individuals referenced specific multinomial aspects of the sequences, such as no use of A and D, and the long run of answer B. As such, it appears that certain determinants of representativeness (specifically and respectively: similarity of sample to population and the local representativeness determinant of the reflection of randomness determinant of representativeness) can be extended from binomial to multinomial sequences. Determinants of representativeness were also featured in attention to contextual aspects, as detailed below.

The majority of justifications provided by the participants addressed our second research question (What impact does contextualizing of sequences have on participants' responses?). In the presented context most respondents implied that the teacher is the creator of the answer key. As such, the independence of events is not in accord with this assumption. Furthermore, where some participants expected an "approximately equal" appearance of each answer, others referred to an informal belief of "when in doubt choose C," which may support the application of the similarity of sample to population determinant of representativeness. We itemized the references to the context in which the task was presented, illustrating the ways in which the answer key perspective, student and teacher perspectives, and personal experience featured in the participants' justifications. We note that these perspectives were present in participants' justifications, regardless of their choice of answer key.

Individuals who referenced the contextual aspects of the sequences they were presented referenced a variety of different perspectives within their response justifications. Moreover, independent of whether the individuals referenced the answer key, personal, student, teacher, multiple or combined perspectives in their responses, the lack of variety, consecutive answers of letter B, and presence of pattern in AK2 all demoted the likelihood of AK2. As such, while it can be argued that the similarity of sample to population determinant of representativeness and both determinants of the reflection of randomness determinant of representativeness (i.e., local representativeness and irregularity) extend from platonic, binomial sequences to contextual, multinomial sequences; response justifications, from all perspectives, referenced in one form or another, an innate structure for answer keys. As such, we are reticent to declare with certainty whether (1) the determinants of representativeness or (2) an individual's perceived innate structure of an answer key is the root cause for declaring AK2 as less likely than AK1. That being said, what we are able to assert is that the novel task we chose for our study characterizes, even epitomizes, the determinants of representativeness.

CONCLUDING REMARKS

As mentioned earlier, ignoring the context and taking a purely theoretical perspective, there are 410 possible answer keys that are equally likely to occur, if constructed by an independent random choice for each specific answer. None of the respondents (with the possible exception of Zonie) took a 'purely mathematical' approach to the task, that is, an approach that involves independence of choices. It can be argued that one of the reasons for this was the similarity to the early CLT used by Tversky and Kahneman (1974) where the "equally likely" option was not available. Another possible explanation can be sought in that prospective teachers responding to the questionnaire, though enrolled in courses addressing the teaching of mathematics, did not have a "strong" mathematics background, such as a degree in the subject. As a counter-argument to that we refer to the numerous studies (e.g., Chernoff, 2009; Cox & Mouw, 1992; Hirsch & O'Donnell, 2001; Kahneman & Tverksy, 1972; Konold et al., 1993; Rubel, 2006, 2007; Shaughnessy, 1977, 1981) where platonic, binomial sequences were presented, in terms of a CLT (e.g., flips of a coin), to a population with a similar mathematics background; and the majority of respondents identified the sequences as equally likely. Therefore, we find that avoidance of theoretical probabilistic considerations was, in this instance, influenced by the two novel characteristics of the task: multinomial choices and contextual setting. However, our study does not separate the relative influence of each of these features in this regard.

According to the representativeness heuristic, certain individuals (namely those who employ the heuristic during judgments of relative likelihood of sequences of outcomes) expect sequences of outcomes to (1) accurately reflect the population ratio even in small samples, (2) have frequent switches and short runs, and (3) reflect, in the outcomes, randomness. For example, for coin flip sequences, individuals employing the representative heuristic would expect a near even ratio of head to tails, frequent switches between heads and tails, short runs of heads (or tails), and the (perceived) absence of a pattern in the results. When the determinants of representativeness are employed in answer key situations, one would expect the quiz to (1) accurately reflect the ratio of possible answers (i.e., if four answers are available, then each answer should appear approximately 25% of the time), (2) not have a long, consecutive run of one answer, but, rather, frequent switches between each answer and short consecutive runs of one answer, and (3) to not have a pattern, which could be picked up upon by, say, the student. However, for some individuals, including certain individuals in this study, having a variety of answers, frequent switches and short runs, and the absence of a pattern all represent innate structural features inherent to answer keys. As such, it becomes difficult to declare whether, for certain contextual sequences, like in the answer key task, participants are employing determinants of representativeness or if the representativeness lens is rendered moot due to contextual considerations. In order to parse the 'chicken and egg' scenario brought forth, further research in the area of contextual sequences (e.g., the answer key to a True of False quiz or rolling a regular tetrahedron) will need to be conducted.

Despite the difficulty presented above, we suggest that context had an overwhelming impact on the presented justifications, where, in some cases, the probability consideration was replaced with a possibility and reflection based upon personal experience. While the tasks involving flipping a coin or a dice refer to an experiment, it is a thought experiment rather than a conducted one. That is, in considering the task presented in Figure 2, it is unlikely that respondents would engage in repeating five flips of a coin. It is similarly unlikely that respondents have in their memory a large repertoire of such experiments carried out in the past.

However, when a task refers to a multiple choice test, past experience, especially in the case of future teachers of mathematics, is relied upon and featured in justifications. To study the extent of the impacts of such experience and its context, future research should attend to two isomorphic tasks, where only one of them is presented in a contextualized setting, which would contribute to needed research investigating teachers' conceptions of probability.

REFERENCES

6th human foot found on B.C. south coast. (2008). CBC news – British Columbia. Retrieved from http://www.cbc.ca/canada/british-columbia/story/2008/06/18/bc-sixth-foot-found.html

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.

Batanero, C., Green, D. R., & Serrano, L. R. (1998). Randomness, its meaning and educational implications. *International Journal of Mathematical Education in Science and Technology, 29*(1), 113-123.

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

Beswick, K. (2011). Putting context in context: An examination of the evidence for the benefits of 'contextualised' tasks. *International Journal of Science and Mathematics Education, 9*, 367-390.

Boaler, J. (1993). The role of contexts in the mathematics classroom: Do they make mathematics more real? *For the Learning of Mathematics, 13*(2), 12-17.

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. (2011). Investigating relative likelihood comparisons of multinomial, contextual sequences. Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education. [Online: http://www.cerme7.univ.rzeszow.pl/index.php?id=wg5].

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

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

Chernoff, E. J., Zazkis, R. (2011). From personal to conventional probabilities: from sample set to sample space. *Educational Studies in Mathematics, 77*(1), 15-33.

Cohen, J., & Hansel, C. E. M. (1958). The nature of decisions in gambling. *Acta Psychologica, 13*(24), 357-370.

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.

Falk, R. (1981). The perception of randomness. In *Proceedings of the fifth conference of the International Group for the Psychology of Mathematics Education* (pp. 222-229). Grenoble, France: University of Grenoble.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgement. *Psychological Review, 104*(2), 310-318.

Fischbein, E. (1999). Intuitions and schemata in mathematical reasoning. *Educational Studies in Mathematics, 38*, 11-50.

Fischbein, E. & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*(1), 96-105.

Green, D. R. (1983). A survey of probability concepts in 3000 pupils aged 11-16 years. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.), *Proceedings of the first international conference on teaching statistics* (pp. 766-783). Sheffield, UK: Teaching Statistics Trust.

Green, D. R. (1988). Children's understanding of randomness: Report of a survey of 1600 children aged 7-11 years. In R. Davidson & J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 287-291). Victoria, BC: University of Victoria.

Grouws, D. A. (1992). *Handbook of research on mathematics teaching and learning*. New York: Macmillan Publishing Company.

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.

Jones, G. A., & Thornton, C. A. (2005). An overview of research into the learning and teaching of probability. In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 65-92). New York: Springer.

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

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. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education, 3*(1). Retrieved from: http://www.amstat.org/publications/jse/v3n1/konold.html

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.

Kramarski, B., Mevarech, Z., & Arami, M. (2002). The effects of metacognitive instruction on solving mathematical authentic tasks. *Educational Studies in Mathematics, 49*(2), 225-250.

Mamolo, A. & Zazkis, R. (2014). Contextual considerations in probabilistic situations: an aid or a hindrance? In (Eds.) E. Chernoff & B. Srirman, *Probabilistic thinking: presenting plural perspectives (PT: PPP)*, (pp.641-656). Dordrechet: Springer

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.

Rubel, L. H. (2007). Middle school and high school students' probabilistic reasoning on coin tasks. *Journal for Research in Mathematics Education, 38*(5), 531-556.

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.

Shaughnessy, J. M. (1981). Misconceptions of probability: From systematic errors to systematic experiments and decisions. In A. Schulte (Ed.), *Teaching Statistics and Probability: Yearbook of the National Council of Teachers of Mathematics* (pp. 90-100). Reston, VA: NCTM.

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.

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 216-226). Reston, VA: National Council of Teachers of Mathematics.

Schilling, M. F. (1990). The longest run of heads. *The College Mathematics Journal, 21*, 196-207.

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.

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable*. New York: Random House.

Toohey, P. G. (1995). Adolescent perceptions of the concept of randomness. *Unpublished master's thesis*, The University of Waikato, New Zealand.

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

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

Watson, J. M., & Kelly, B. A. (2009). Development of student understanding of outcomes involving two or more dice. *International Journal of Science and Mathematics Education, 7*, 25-54.