Communications of the Association for Information Systems

Volume 52 Article 11

2023

Teaching Awareness of Ambiguity in Data

Earl McKinney
Bowling Green State Univeristy, emckinn@bgsu.edu

Rachel Shaffer
Bowling Green State University

Communications of the Association for Information Systems

Accepted Manuscript

Teaching Awareness of Ambiguity in Data

Earl McKinney Jr.
Department of Accounting and MIS
Bowling Green State University
emckinn@bgsu.edu

Rachel Shaffer
Department of Economics
Bowling Green State University

Please cite this article as: McKinney, Earl Jr.; Shaffer, Rachel: Teaching Awareness of Ambiguity in Data, *Communications of the Association for Information Systems* (forthcoming), In Press.

Communications of the Association for Information Systems

Education Paper ISSN: 1529-3181

Teaching Awareness of Ambiguity in Data

Earl McKinney Jr.
Department of Accounting and MIS
Bowling Green State University
emckinn@bgsu.edu

Rachel Shaffer
Department of Economics
Bowling Green State University

Abstract:

The widespread and growing use of analytics has highlighted the need for more data savvy students. One recently identified aspect of data literacy is the awareness of the ambiguity. This paper outlines a method and initial results for raising awareness of data ambiguity in a short, one lesson, active learning format suitable for business or computer science courses. The paper also includes a summary prior research on ambiguity, data literacy taxonomies, implications, and suggestions for further research.

Keywords: Analytics, Data Literacy, Ambiguity.

[Department statements, if appropriate, will be added by the editors. Teaching cases and panel reports will have a statement, which is also added by the editors.]

[Note: this page has no footnotes.]

This manuscript underwent [editorial/peer] review. It was received xx/xx/20xx and was with the authors for XX months for XX revisions. [firstname lastname] served as Associate Editor.] **or** The Associate Editor chose to remain anonymous.]

1 Introduction

In the current business environment, organizations increasingly rely on analytics (Boy et al., 2014). As a result, data literacy is rapidly becoming a paramount educational objective (Hamilton et al., 2009) that needs more development (Mandinach & Gummer, 2013; Julien, et al., 2018; Flywel & Jorosi, 2018; Pothier & Condon, 2019). Data literacy skills are as important as the ability to read and write text (Borner et al., 2019). Data literacy leads to better communication, problem-solving, and decision-making (Lee et al., 2019; Mandinach & Gummer, 2013; Wolff, et al., 2016).

To better prepare graduates, teaching research is beginning to develop and organize taxonomies for data literacy (Ridsdale, 2015; others). In many of these taxonomies critical thinking about data is a common feature. However, techniques for teaching data literacy critical thinking skills necessary for business and computer science students are inconsistent and underdeveloped. Awareness of ambiguity has only recently been identified as a critical thinking data literacy skill.

This paper presents a way to teach this new aspect of data literacy to business or computer science students. It proceed as follows: the relevant prior research on ambiguity is summarized, the analytics course lesson in which ambiguity is explained and applied is discussed, a preliminary analysis on the impact of the lesson on student awareness of ambiguity is presented, and a discussion of implications and future research concludes the paper.

Presented elsewhere is a method to raise awareness of ambiguity in a semester long case project suitable for an analytics course (McKinney & Ginzinger, 2022). The current paper is offered as a way to raise awareness in one class hour in any business or computer science or analytics course.

2 Ambiguity

Ambiguity is defined as the multiple reasonable interpretations of data, the opposite of certainty. Ambiguity exists in all data –raw data, processed data or results data. When an individual is aware of the ambiguity in data, that individual realizes that the data can mean many things, so that no one can be certain their interpretation is the “right one”, the most valid or the most useful.

The concept of ambiguity and its relevance for analytics is more completely described in McKinney & Bhatia (2022) where a proposed survey used to measure awareness of ambiguity is validated. Ambiguity is a fundamental to the human condition (Kierkegaard, 1983; Prigogine & Stengers, 1997; Gron, 2008). Aristotle attributed the perniciousness of ambiguity to having a finite and limited set of words and sentences while the reality that the words represent is always unlimited (Robinson, 1941). Ambiguity permeates everyday life (Abbott, 1997; Belshaw & Higgins, 2011) and has become an essential phenomenon in many research domains such as language (MacDonald et al., 1994), anxiety (Dibner, 1958; Caplan & Jones, 1975), economics (Shackle, 2010), decision making (Ellsberg, 1961), and machine learning (Peysakhovich & Naecker, 2017). For example, in language, individuals reconcile grammatical, lexical, phonological, and semantic ambiguity to create meaning and take action (MacDonald & Pearlmutter, 1994).

In decision making literature, ambiguity is often called uncertainty and is associated with unknown probabilities (Gilboa, 2011 p. 8). In social sciences, ambiguity as reasonable interpretation has been used in studies in all domains of social life, action, and communication (McMahan & Evans, 2018). These include such disparate topics such as human play (Sutton-Smith, 2009), institutional change (Mahoney & Thelen, 2009), politics (Page, 1994), medical student selection, training (Geller et al., 1990), virtual team commitment (Galvin & Piccoli, 2006), and sensemaking (Namvar et al., 2021).

In most applications, ambiguity is defined as polysemy or multivocality, the way a symbol, sign, work, document or performance can have multiple possible meanings (Baldassarri, 2018). Positivists and most social science research traditionally has viewed ambiguity as a limitation, a distortion from a model of perfect information, a source of misunderstanding, or a possible cause of unintended consequences. But Levine (1965, 1988) argues social scientists have avoided ambiguity and have as a result failed to realize its constructive possibilities; and Baldassarri (2018) suggests the ambiguity of social actions is necessary for the participation in social relations. His classic example is dating, a romantic relationship that unfolds from the ambiguity that supports the actions and behaviors of the participants.

Ambiguity is often related to vagueness. A vague term is one couched in general terms and not definitely or precisely expressed (Williamson, 2002). A term is considered vague if it includes borderline cases, for example, personhood—when does a fertilized egg become a person. Vagueness applies to terms while ambiguity describes phenomena. The term, heap of sand, is vague as many piles of sand fall into borderline cases. While that particular pile of sand over there is ambiguous--it could be interpreted as a heap of sand, as an input to concrete, or a mess.

The theoretical persistence of ambiguity in analytics is explained by underdeterminism. Underdeterminism holds that for every interpretation or theory that explains data, there are other interpretations or rival theories that explain the data at least as well. The data literally under determines the interpretation of it, in all cases. In contrast, data that fully determines the interpretation would be like a key field a database table, the key field determines all the other fields in the record. Underdeterminism was first applied to phenomena by John Stuart Mill in A System of Logic (1843), and later to all knowledge claims by Quine (1951) in one of the most remarkable developments in epistemology in the 20th century (Stanford, 2017).

The ubiquity of ambiguity is also revealed by the frequent and common use of induction. Most analytics produce results, an analysis on a sample. An interpretation of results often requires the viewer to apply the results of the sample to a population, and multiple reasonable populations are available. For example, when examining a declining sales set of data, does this mean is a problem to be solved, or because these sales totals declined less than sales in the general economy the company has no problem.

Finally, ambiguity can be attributed due to the variety in experiences of the viewers. Different viewers bring different knowledge and experience to the interpretation of a result. When interpretations occur, individuals include their knowledge and experience to provide what is missing; to provide a comparisons for the data, and to provide opportunities. To return to the sales example, a sales manager might interpret the decline to a change in compensation given to sales agents while a supply chain manager might interpret it as evidence that her brilliant suggestion a year ago should have been listened to.

The most frequently studied behavioral construct that involves ambiguity is the tolerance of ambiguity (TA), also called uncertainty avoidance by Hofstede (1983). Tolerance for ambiguity is a behavioral reaction that describes the relationship that individuals have with ambiguous signs or events. Individuals view these stimuli in a neutral and welcome way or as a threat. Intolerance of ambiguity may be defined as the tendency to interpret ambiguous situations as sources of threat; tolerance of ambiguity as the tendency to perceive ambiguous situations as desirable.

Tolerance of ambiguity is a stable, long-term attitude or personality trait that leads to a reaction to a context. However, awareness of ambiguity in data is a skill that can be taught and practiced. In addition, awareness of ambiguity applies to data; whereas tolerance of ambiguity is often about situations an individual finds themselves in or anticipates being in. Data is not personal, and little risk is involved.

2.1 Ambiguity in Data

For this study, data includes raw data, processed data, or results data. In short, any sign that is created to represent something else.

In many analytics settings, there is clearly a best or most favored answer. There is little or no relevant ambiguity about the data that shows the average score on an exam, or the distances between warehouses. The calculation is correct, and there is only one answer that minimizes interwarehouse distance. However, many analytical problems do not have a single best answer. In most business settings, ambiguity, or potential for multiple interpretations (or meanings) is a significant issue. For example, the costs of goods sold (COGS) data is ambiguous. One method of calculating COGS shows that costs have increased 5% per year for the past 3 years; but there are many other ways a cost could have been reasonably calculated and defended. In addition to method differences, interpretation differences also exist.

While modeling uncertainty and unknown probabilities have been the topic of considerable recent analytics research (Hirshleifer & Riles, 1992; Taleb, 2007), ambiguity as multiple interpretations of results in analytics has had little prior work.

Awareness of ambiguity is a skill, it is the ability to perceive multiple reasonable interpretations of data, is the opposite of certainty. When an individual is aware of the ambiguity in data, that individual realizes that the data can mean many things, no one can be certain their interpretation is the “right one”, the most valid, or the most useful.

2.2 Data Literacy

Awareness of ambiguity should be a component of critical thinking within data literacy. Data literacy, popularized by Gilster (1997) is the ability to read data, to understand and use data effectively (Mandinach & Gummer, 2013). More specifically, data literacy, or applying data critically, is the ability to select, clean, analyze, visualize, critique, and interpret data as well as communicate stories from data (Mandinach & Gummer, 2013; Wolff et al., 2016). The term information literacy is similar, but distinct; information literacy includes the skill of noticing missing information.

While data literacy is essential, it is lacking in professionals and students (Hamilton et al., 2009; Mandinach, 2009), and needs more research (Mandinach & Gummer, 2013; Julien, et al., 2018; Flywel & Jorosi, 2018; Pothier & Condon, 2019). Schools are still poorly preparing graduates (Ridsdale, 2015), and businesses are recognizing that investments in technology are less essential than investments in skill training (Winterberry Group, 2018; Bean & Davenport, 2019; Pothier & Condon, 2019).

To better prepare graduates, teaching research is beginning to organize taxonomies for data literacy (Carlson et al., 2015; Ridsdale, 2015; Calzada et al., 2017). In many of these taxonomies critical thinking about data is a common feature. However, data literacy critical thinking resources necessary for training is inconsistent and underdeveloped (Ridsdale, 2015; Pothier & Condon, 2019).

2.3 Motivation for study

One critical thinking skill, awareness of ambiguity, is essential for all business and computer science students. Without the awareness that data and results are always somewhat ambiguous, naïve professionals and students will assume data has definitive answers or single best interpretations and overconfidence in their particular interpretation. This often leads to assumptions that data proves a particular result, that alternative explanations of results are unnecessary at best and wrong at worst, and that the data speaks for itself. Without an awareness of ambiguity, students may believe professional analysts produce the right answer to analytics business problems, rather than results that are then interpreted and debated by other professionals.

While the impact of ambiguity is significant, little research has addressed its impact on analytics education. Most generally, awareness of ambiguity would be a critical thinking skill. Several educational models include critical thinking as a skill within data literacy (Ridsdale, 2015), but none identify awareness of ambiguity as an element within either critical thinking or data literacy.

McKinney and Bhatia (2022) reports on the validity of the awareness of ambiguity survey used here to measure student awareness, and provides a more extensive review of ambiguity and its role in analytics. McKinney & Ginzinger (2022), reports on a semester-long case study being developed to raise awareness of ambiguity specifically for an analytics course.

3 Method

This lesson has been taught in a wide variety of programs and classes to undergraduates and graduate students. The results section reports on data collected from students in an analytics course in the first two semesters. In both semesters one group of students were undergrads in their final semester, and the other were students in a course for a one year Masters of Accountancy program, for a total of four sections.

In the course, the teacher makes distinctions between raw data, processed data, and results data and explains to students ambiguity is present in all. However, most of the exercises and discussions presented here are not specific to one type of data, they apply to all three.

This lesson relies on active learning. Active learning involves students in an activity that leads them to reflect on the ideas of the activity and how to employ them in their own lives (Bonwell & Eisen, 1991; Michael, 2006). It emphasizes taking an active part in learning as opposed to passively listening to a lecture and is a proven effective pedagogy (Tharayil et al., 2018). There are many different approaches to active learning in higher education (Johnson, 1991); the exercise presented here is an example of collaborative learning (De Hei, Strijbos, & Admiraal, 2015), and self-regulated learning (Pintrich, 1995).

This lesson on the ambiguity of data supports two learning objectives. First, for students to be able to recognize that there are multiple valid interpretations of all sets of data. Second, for students to be able to explain skills needed by analysts to interpret ambiguous data. These objectives were included in the syllabus.

The class lesson has two phases. First, the teacher attempts to persuade students that ambiguity is common in the data of their life, this phase takes 25 to 30 minutes. Then the teacher leads a discussion on skills necessary for analysts in such an environment which requires about 20 minutes.

3.1 Phase 1.

To persuade students about the ubiquity of ambiguity the class begins with a series of questions and follow up discussion. Students are shown a picture of the top row of yellow and red tiles as shown in Figure 1. They are asked to determine the next color in the sequence. When a student nominates red or yellow, they are asked why, so that other students can hear the reasonable interpretation of the tile data. Next they are shown each of the other sequences in Figure 1. After a couple of these, some students begin to nominate unusual, but reasonable conclusions, such as, the next color is neither red or yellow, or perhaps both. These unusual, but reasonable responses are appreciated by both instructors and students.

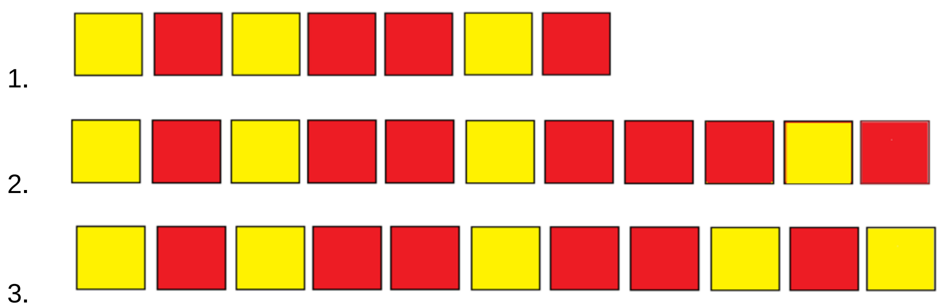

Figure 1. Color Tile Sequences

In the next exercise the teacher asks students to identify the signs (the data) that accompany the activity of dating. The teacher suggests that each individual in a potentially romantic situation is providing signs to the other. Students nominate different signs, and enjoy considering the common malady that each partner has interpreted the sign a different way.

Next the instructor introduces the concept of polysemy, that words can have many distinct meanings. The teacher asks them to estimate the number of uses of common terms such as brake, put, stand, and get. They typically respond in a range from 10 to 30. Then the teacher reveals these words can be used in over 250 different ways, and include slides with examples of some of these uses.

The teacher then refers the students to puns, and the need for words to have multiple interpretations for puns to work. The instructor asks them to find a pun on the internet, or via their voice assistant, share it with the class, and suggest the term or phrase that has multiple interpretations.

The teacher shows the students a slide with pictures of objects as shown in Figure 2. They are asked which animal, the cat or moth should follow the first three—the dog, bird, and fish sequence. Both the cat and moth are often suggested and defended reasonably. The teacher then suggests that another reasonable interpretation of this data is may occur to Spanish speaking students who might label the five as perra, pajaro, pascado, gata, and polilla; and defend the moth as the next animal as it begins with "p" in Spanish. This highlights the role of experience in interpretation; each individual in the class has different experiences which lead to different reasonable interpretations.

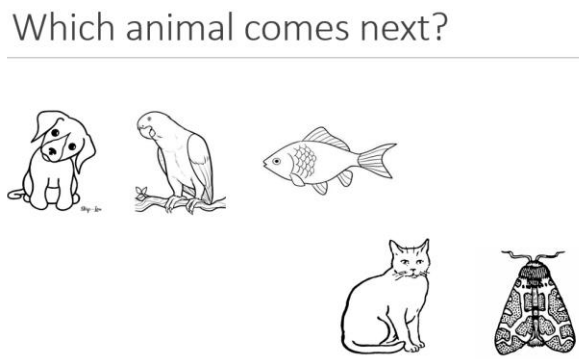

Figure 2. Animal Sequence

Next, the teacher displays a simple visualization as shown in Figure 3, explain that the data comes from a firm in the region, and ask for interpretations of the visualization. After a variety of responses the teacher adds four or five of their own to demonstrate the variety of possible interpretations. The teacher suggests that all visualizations are similar, that each has many reasonable interpretations.

Figure 3. Business Data Visualization

During this discussion of the viz, the teacher emphasizes how the experience of the individual can impact interpretation. The instructor asks students to imagine they are a professional in their favorite hobby or pastime—a setting for which they have considerable knowledge and experience. The teacher suggests to them that in their highly experienced domain, they will be able to nominate many interpretations. For example, the viz used in the class shows sales data, and the teacher asks the students to imagine how professionals attending a meeting might discuss the sales data—some will interpret the data with a supply chain explanation, others marketing, others personnel changes, and still others with decision making explanations. Further, the instructor suggests it is reasonable to assume that most of us underestimate the variety of different experiences people will apply to an interpretation.

Finally the teacher provides a short explanation of underdeterminism and induction for theoretical support. The instructor explains underdeterminism and provide them some financial data on a company's annual report. The teacher asks them to consider if they generated an interpretation of the data, how likely is that interpretation the best one, or would an individual working for a million years, or a million other analysts generate a better interpretation.

The instructor then reminds them of the principles of induction. The teacher asks them how many pennies must an individual pull out of a bag of coins to conclude it is a bag of pennies? In life, individuals conclude a thing fits a category after a number of trials, individuals sample then generalize to a population. The teacher does not discourage or disparage induction, as it clearly works most of the time, and individuals rely on it perhaps hundreds of times a day. The instructor then suggests to the students that different individuals generalize the results of a sample to different populations based on their experiences, and that most individuals underestimate how many different populations other individuals might generalize to.

3.2 Phase 2

Next the teacher shifts to a discussion of what professional skills would be helpful in an environment where all data are known to be ambiguous. The instructor contrasts this environment with a mathematical environment where there is a single best interpretation.

In the first two semesters using this lesson, students created the list in Figure 4.

a) desire to learn on your own

b) productive skepticism

c) curiosity

d) openness to contrary feedback

e) sensitivity to different experience of individuals

f) anticipate and rule out alternative explanations

g) nominate and discuss criteria for evaluating the various reasonable interpretations

h) creativity

i) humility

Figure 4. Skills needed for ambiguous data nominated by students

The teacher then leads a discussion of which of these skills identified in phase 2 they lack. The teacher suggests that although each student is limited in their ability to know their own limitations, they are the best judge of their own shortcomings. The teacher requires that they turn in a self-assessment of their skills at that point in time, along with a description of how the student plans to improve. This self-assessment is returned at the end of the semester and students are asked to reflect on their improvements. The teacher also suggests students maintain a place to record notes about their experiences as evidence of skill improvement.

4 Results

To assess whether these exercises were effective, the survey instrument developed to assess awareness of ambiguity in visualizations (McKinney & Bhatia, 2022) was used and the results compared the students who received this lesson with students not in the class. The scale ranges from 0 to 31 based on 6 questions. This survey was given to both students in the class and a group of over 300 students at the university who were not in the course and did not experience this lesson. Students not in the class averaged 19.2. The differences are shown in Table 1 and are significant (alpha < .05).

Table 1. Students in class compared to those not in class

Student Group	Awareness of Ambiguity Survey Score After Lesson
Not in the Course	19.2
In Class (Spring 2021)	22.5
In Class (Fall 2021)	24.2

Student survey results after the lesson were compared to their score before the lesson. The students took the survey twice, once two weeks before the lesson on ambiguity and once a month later, two weeks after the lesson on ambiguity. Students were not rewarded for participation in the survey. With a 70% participation rate for each survey and approximately 100 students in each semester, only 58 and 61 matched pairs of student records are available from the spring and fall semesters respectively. The initial results shown in Table 2 are statistically significant for each semester using a .01 level of significance.

Table 2. Students Increasing or Decreasing Scores on Awareness of Ambiguity on Pre and Post Survey

Semester	Increase	Decrease	Same Score	Total
SPRING 2021 1	--	--		
Number of Students	47	6	5	58
FALL 2021	--	--		
Number of Students	54	4	3	61

In addition to the matched pair analysis, the overall averages on the survey by the students before and after the lesson were compared. During the spring 2021 semester, students had the option of attending class or taking it remotely synchronously. Table 3 shows the Spring students in 3 groups—students who attended the class in person, those who participated and watched synchronously, and those who did neither. All students in the Fall semester were in class.

Table 3. Student Awareness of Ambiguity Before and After Lesson

Semester	Number of Students	Awareness of Ambiguity Before	Awareness of Ambiguity After	Change
SPRING 2021				
Absent	16	19.6	21.6	2.0
Zoom	25	20.2	21.8	1.6
In Classroom	17	19.8	24.6	4.8*
Total	58	19.9	22.5	2.6
FALL 2021				
In Classroom	61	19.3	24.2	4.9*
* Note: < .01 significance				

As shown in Table 3, the set of exercises described earlier led to a significant change in Awareness of Ambiguity for the students who attended class in the spring semester. For those 17 students, their awareness increased to 24.6, a significant difference of 4.8 from 19.8 before the exercises. For students not in class, the zoom or absent students, the increase in Awareness of Ambiguity were 1.6 and 2.0 respectively. Those students were assigned a reading about ambiguity in visualizations, but no assignment on the reading nor responses to questions was required.

Finally the impact of the lesson was assessed using an end of course assigned project. This assignment, called a journal, asked students to reflect on the key lessons over the entire course, and apply these lessons to their own lives at work and away from work. The journal was written three weeks after the ambiguity lesson/exercises. The assignment was ambiguous by design and was not created for this research. Students were not rewarded for identifying particular topics, but rather if they could construct course long themes and apply lessons to their lives.

Using a text mining tool, the frequency of use of the term “ambiguity” was counted for groups of students, and the results are summarized in Table 4. The top row of Table 4 shows data from the previous semester, a control group, when the ambiguity exercises were not an essential topic in the course. The same teacher and same journal instructions were used in both classes.

Table 4. Word counts of Ambiguity and other Terms This Semester and Previous Semester

Semester	Rank of Ambiguity in Most Common Words	Percentage of Words that are Ambiguity
SPRING 2020 (no lesson)	48th	.30%
SPRING 2021		
Absent	39th	.51%
Zoom	42nd	.46%
In Classroom	33rd	.52%
Total	38th	.50%
FALL 2021	32nd	.53%

These results imply that ambiguity had become an important construct to students during the semester. Compared to the most recent previous version of the class, the use of the term grew from .30% to .50% and .53%. The students who attended the class seemed again to be the most affected.

5 Discussion

This study attempted to raise student awareness of data ambiguity. Initial results were encouraging as a significant number of students increased their awareness of ambiguity. Further, those who attended the class and participated in the ambiguity exercises displayed a significant gain in their awareness of ambiguity in a data visualization compared to students in the same course who did not actively participate in the ambiguity exercise. Also, these active students displayed a greater awareness of ambiguity when describing their impressions of the important lessons of the course compared to students who did not actively engage and to students in the most recent ambiguity-free version of the course. Although these effects were statistically significant, the entire lesson was completed in 45 to 50 minutes of class time.

While this particular set of exercises was used, alternative exercises and exercise domains based on instructor experiences and knowledge may be just as effective. The key criteria are that students recognize no authoritative solution exists in the data in the exercise domain, that the collection of exercises vary widely to suggest ambiguity is present in all forms of data, and the exercises involve commonly occurring experiences in the lives of students.

Helping students become more aware of ambiguity should become an essential critical thinking skill in data literacy. Without an awareness of ambiguity, many students fail to appreciate the variety of interpretations of a set of data. Without the awareness that data is always ambiguous, students might assume data has absolute answers or a single conclusive interpretation and be overconfident that their particular interpretation is the one answer. Without ambiguity awareness, student might hold the naïve view that data proves results, that alternative explanations are not needed, and that the data can speak for itself.

Being able to perceive ambiguity alters expectations of analysts. Instead of assuming the false dichotomy that results are either truth or false and conducting an analysis to find the one true answer, students can view the analytics process as uncovering a variety of alternative interpretations and explanations, seeking to establish the criteria on which the interpretations should be assessed, and participating in the collaboration that discusses alternatives. In a lesson after the ambiguity lesson the teacher explains that ambiguity does not need to result in relativism, that all interpretations are equal. The teacher explains that the role of decision makers is to nominate criteria (McKinney and Yoos, 2019) for decisions and chose among alternatives.

When viewed as developing answers rather than "the" answer, a different configuration of desirable skills for students becomes evident. As nominated by the students, these skills include curiosity, imagination, communication, independent learning, skepticism, openness to contrary feedback, sensitivity to different experiences of individuals, and ability to nominate and discuss criteria for evaluating interpretations. Many of these skills have long been the goals of a liberal education, so using them in an analytics course reinforces and applies skills developed elsewhere.

In an age of AI, an awareness of ambiguity of data helps students realize the limitations of AI and the role of human interpretation. AI algorithms will continue to improve and support the continued explosion in data and analytics; however, the final step of interpretation of results for novel analysis, or an analysis done for the first time by a firm will require human interpretation. Student analysts should be prepared to recognize where the human element will always be required. Students can recognize that although ambiguity can be off-putting, it creates positive possibilities.

At the end of the discussion students are asked to notice that the discussion has presented them with data in the form of words. These words, the discussion, is clearly ambiguous, different students in the room will interpret the skills discussion in different ways. This recursion of ambiguity typically surprises students, which allows us to suggest to them that they are still not as aware of ambiguity as they might think.

Some students interpret this discussion as a promotion of relativism; however teachers attempt to convey that ambiguity does not lead to relativism, rather to learning how to inform in ambiguity with criteria. While many interpretations are always reasonable, some are more useful, simple, and predictive to name a few criteria. In a previous report**,** McKinney & Yoos (2016**),** suggested ways for students to flourish in ambiguity by informing themselves (McKinney & Yoos, 2019) by using criteria.

5.1 Suggestions for Future Work and Limitations

While this lesson has been taught to other faculty, and in many other student settings, the current work only reports on the impact on students in two semesters. Clearly a next step would be to test these ambiguity awareness exercises on other groups of students. Further, this study used an ambiguity survey that measures the awareness of ambiguity in a visualization. Other studies should also endeavor to assess if students become more aware of the ambiguity in a variety of datasets. For example, students aware of ambiguity may still believe that data that has been processed will be less ambiguous than raw data. Finally, it is unclear if this awareness of ambiguity is lasting or short term.

In conclusion, this study attempted to raise student awareness of data ambiguity. This paper suggests this skill be added to the growing list of critical thinking skills for data literacy.

Acknowledgments

Permission to evaluate human subjects for this study was obtained from Bowling Green State University Institutional Review Board (Exemption #1645509).

Funding for this research was provided in part by Fulbright Grant 11506-AU.

References

Abbott, A. (1997), Seven types of ambiguity. *Theory and Society, 26(2-3)*, 357-391.

Baldassarri, D. (2018). Keeping one's distance: Truth and ambiguity in social research. *Sociologica, 12(1)*, 5-11.

Bean, R., & Davenport, T. H. (2019). Companies are failing in their efforts to become data driven. *Harvard Business Review,* 5-8.

Belshaw, D., & Higgins, S. (2011). *Digital literacy, digital natives, and the continuum of ambiguity*. Dougbelshaw.com.

Bonwell, C.C. & Eison, J.A. (1991). Active learning: Creating excitement in the classroom. In: 1991 ASHE-ERIC Higher Education Reports. ERIC Clearinghouse on Higher Education. The George Washington University, One Dupont Circle, Suite 630, Washington, DC 20036-1183.Prince, 2013

Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences 116.6* 1857-1864.

Boy, R. A., Rensink, E., & Fekete, J. (2014). A principled way of assessing visualization literacy. IEEE Transactions on Visualization and Computer Graphics, Institute of Electrical and Electronics Engineers, 2014, 20 (12), pp.10. ff10.1109/TVCG.2014.2346984ff. ffhal-01027582f

Caplan, R. D., & Jones, K. W. (1975). Effects of workload, role ambiguity, and type A personality on anxiety, depression, and heart rate. *Journal of Applied Psychology, 60(6),* 713.

Calzada, I. (2017). Problematizing and politicizing smart city-regions: is devolution smart?. *Problematizing and Politicizing Smart City-Regions: Is Devolution Smart?*, 37-47.

Carlson, J., Nelson, M. S., Johnston, L. R., & Koshoffer, A. (2015). Developing data literacy programs: Working with faculty, graduate students and undergraduates. *Bulletin of the Association for Information Science and Technology*, 41(6), 14-17.

De Hei, M.S.A., Strijbos, J.W., Sjoer, E., & Admiraal, W. (2015). Collaborative learning in higher education: Lecturers' practices and beliefs. *Research Papers in Education, 30(2),* 232-247.

Dibner, A. S. (1958). Ambiguity and anxiety. *The Journal of Abnormal and Social Psychology, 56(2),* 165.

Ellsberg, D. (1961) Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 643-669.

Flywel, D., & Jorosi, B. N. (2018). Information literacy skills among the undergraduate students at the University of Livingstonia, Malawi. International Journal of Library and Information Services (IJLIS), 7(2), 43-56.

Galvin, A., & Piccoli, J. (2006). Antecedents to team member commitment from near and far. *Information Technology & People,* 19, 299-322.

Geller, G., Tambor, G., Chase, G., &Holtzman, N. (1993). Measuring physicians' tolerance for ambiguity and its relationship to their reported practices regarding genetic testing. *Medical care*, 989-1001.

Gilboa, I. (2011). *Problems in decision theory*. New York: Rutledge.

Gilster, P. (1997). *Digital literacy*. New York: John Wiley & Sons, Inc.

Grøn, A. (2008). *The concept of anxiety in Søren Kierkegaard*. New York: Mercer University Press.

Hamilton, L., Halverson, R., Jackson, S. S., Mandinach, E., Supovitz, J. A., Wayman, J. C., & Steele, J. L. (2009). Using student achievement data to support instructional decision making.

Hirshleifer, J., Hirshleifer J., & Riley, J. (1992). *The analytics of uncertainty and information*. Cambridge: Cambridge University Press.

Hofstede, G. (1983). National cultures in four dimensions: A research-based theory of cultural differences among nations. *International Studies of Management & Organization, 13(1-2),* 46-74.

Johnson, D.W., Johnson, R.T., & Smith, K.A. (1991). *Active Learning: Cooperation in the College Classroom. Edina*, MN: Interaction Book.l

Julien, H., Gross, M., & Latham, D. (2018). Survey of information literacy instructional practices in US academic libraries. *College & Research Libraries, 79(2),* 179.

Kierkegaard, S. A. (1983). *Fear and trembling, and repetition*, (H. Hong & E. Hong, Eds. & Trans.). Princeton: Princeton University Press. (Original work published 1843)

Lee, S., Kwon, B. C., Yang, J., Lee, B. C., & Kim, S. H. (2019). The correlation between users' cognitive characteristics and visualization literacy. *Applied Sciences, 9(3)*, 488.

Levine, D. N. (1988). *The flight from ambiguity: Essays in social and cultural theory*. Chicago: University of Chicago Press.

Levine, M. M., & Chernick, J. (1965). A numerical model of random packing of spheres. *Nature, 208(5005),* 68-69.

Mac Donald, A. (1970). Revised scale for ambiguity tolerance: Reliability and validity," *Psychological Reports 26(3),* 791–798.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution*. Psychological Review, 101(4),* 676.

Mahoney, J., & Thelen, K. (Eds.), (2009). *Explaining institutional change: Ambiguity, agency, and power*. Cambridge: Cambridge University Press.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42(1)*, 30-37.

McKinney, E. & Bhatia, M. (2022). Measuring awareness of ambiguity in a visualization. Submitted to Inf*ormation Technology & People*.…

McKinney, E. & Ginzinger, S. (2022). Teaching about ambiguity in analytics: A student centered case study to improve data literacy. Submitted to *Journal of Information Systems Education*.

McKinney, E. & Yoos, C. J. (2016). Informing students about information: Seven semantic exercises. *Communications of the Association for Information Systems*, *39(1)*, 16-32.

McKinney, E. & Yoos, C. J. (2019). Information as a difference: Toward a subjective theory of information. *European Journal of Information Systems, 28(4)*, 355-369.

McMahan, P., & Evans, J. (2018). Ambiguity and engagement. *American Journal of Sociology, 124(3),* 860-912.

Michael, J. (2006). Where's the evidence that active learning works? *Advances in physiology education*.

Mill, J. S. (1843). *A System of logic, Ratiocinative and inductive:being a connected view of the principles of evidence and the methods of scientific investigation*. London: John W. Parker.

Namvar, M., Intezari, A. & Im, G. (2021). Sensegiving in organizations via the use of business analytics. *Information Technology & People*.

Page, B. I. (1994). Democratic responsiveness? Untangling the links between public opinion and policy. PS: *Political Science & Politics, 27(1),* 25-29.

Peysakhovich, A., & Naecker, J. (2017). Using methods from machine learning to evaluate behavioral models of choice under risk and ambiguity. *Journal of Economic Behavior & Organization*, 133, 373-384.

Pintrich, P.R. (1995). Understanding self-regulated learning. *New Directions for Teaching and Learning,* 63, 3-12.

Pothier, W. G., & Condon, P. B. (2019). Towards data literacy competencies: Business students, workforce needs, and the role of the librarian. *Journal of Business & Finance Librarianship, 25(3-4),* 123-146.

Prigogine, I., & Stengers, I. (1997). *The end of certainty*. New York: Simon and Schuster.

Quine, W. V. (1951). On Carnap's views on ontology, *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 2(5),* 65-72.

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., & Wuetherick, B. (2015). Strategies and best practices for data literacy education: Knowledge synthesis report.

Robinson, R. (1941). Ambiguity. *Mind, 50(198),* 140-155.

Shackle, G. L. S. (2010). *Uncertainty in economics and other reflections.* Cambridge: Cambridge University Press.

Stanford, K. (2017). Underdetermination of Scientific Theory. The Stanford Encyclopedia of Philosophy (Winter 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2017/entries/scientific-underdetermination/>.

Sutton-Smith, B. (2009). *The ambiguity of play*. Boston: Harvard University Press.

Taleb, Nassim N. (2007). *The black swan: The impact of the highly improbable*. Vol. 2. Random house, 2007.

Tharayil, S., Borrego, M., Prince, M., Nguyen, K. A., Shekhar, P., Finelli, C. J., & Waters, C. (2018). Strategies to mitigate student resistance to active learning. *International Journal of STEM Education*, *5*(1), 1-16.

Williamson, T (2002). *Vagueness*. Routledge.

Winterberry Group (2018). The Data-Centric Organization 2018. (A Winterberry Group Research Report), Retrieved from https://www.iab.com/wp-content/uploads/2018/02/DMA-IAB-Winterberry-Group-The-Data-Centric-Org-2018-February-2018.pdf

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. The Journal of Community Informatics, 12(3).

Appendix: Summary of Ambiguity Awareness Survey

The awareness of ambiguity survey in its complete form can be found in McKinney and Bhatia (2022). Below are the three example questions that form the basis of the instrument. For question 1 a visualization of sales totals over the span of 5 years is presented. The other questions are answers strongly agree to strongly disagree.

1. In visualization 1, how many reasonable interpretations can you identify?
   a. 1 or 2
   b. About 5
   c. About 10
   d. About 15
   e. About 20
   f. More than 25
2. When I create a data visualization at work, I expect others to interpret my visualization differently than me.
3. When I look at a data visualization, I try to see if there are several good ideas.

About the Authors

Earl H. McKinney Jr. is the Timothy and Sue Anne Ross Professor of IS at Bowling Green State University. He previously was a professor at the US Air Force Academy and holds a PhD in MIS from the University of Texas and a Masters of Engineering at Cornell University. His previous works on information, crew communication in aviation, and e-commerce adoption have appeared in *MIS Quarterly, the European Journal of Information Systems, Information and Management,* and *Human Factors*. He recently completed a Fulbright Grant at the Fachhochschule Salzburg Austria, and has recently taught at Tiangong University Tianjin China. He teaches database, analytics, and cybersecurity.

Rachel Shaffer teaches microeconomics at the principles, intermediate, and graduate level at Bowling Green State University. She teaches advanced undergraduate courses in auctions and in industrial organization. Her dissertation work considers traders in double auctions who are regret-minimizers, rather than expected utility maximizers. Other research interests include game theory, behavioral economics, and decision theory. Her PhD in Economics is from the University of Illinois, and her undergraduate degree in Mathematics is from the University of Tulsa.