Creating Shared Understanding in Statistics and Data Science Collaborations

Eric A. Vance, Jessica L. Alzen, Heather S. Smith

Laboratory for Interdisciplinary Statistical Analysis, Department of Applied Mathematics, University of Colorado Boulder, 1111 Engineering Drive, Boulder, CO 80309, USA

Center for Assessment, Design, Research, and Evaluation, School of Education, University of Colorado Boulder, 249 UCB, Boulder, CO 80309, USA

Department of Statistics, 1 Grand Ave., California Polytechnic State University, San Luis Obispo, CA 93407, USA

Corresponding author email: Eric.Vance@Colorado.EDU

Creating Shared Understanding in Statistics and Data Science Collaborations

Statisticians and data scientists have been called upon to increase the impact they have through their collaborative projects. Statistics and data science practitioners and their educators can achieve and enable greater impact by learning how to create shared understanding with their collaborators as well as teaching this concept to their students, colleagues, and mentees. In this paper, we explore and explain the concepts of common knowledge and shared understanding, which is the basis for action to accomplish greater impacts. We also explore related concepts of misunderstanding and doubtful understanding. We describe a process for teaching oneself and others how to create shared understanding. We conclude that incorporating the concept of shared understanding into one's practice of statistics or data science and following the steps described will result in having more impact on projects and throughout one's career.

Keywords: common knowledge; communication; consulting; doubtful understanding; statistical practice; statistics education

1. Introduction

Throughout the years, many have called for statisticians to develop their communication and collaboration skills so that they can work more closely with clients/domain experts and thereby increase their impact. Ram Gnanadesiken, the 1989 President of the Institute of Mathematical Statistics said, "Cross-disciplinary interactions … are means of justifying the relevance and indeed the reason for existence of statistics," (1990, p. 122) and, "We need a switch turned on, a value established, for impelling statisticians to be challenged intellectually and through a desire to contribute to solving major problems in other fields" (1990, p. 124). Peter Lachenbruch, in his American Statistical Association (ASA) Presidential Address stated: "We need to work with statisticians to improve their skills as ambassadors to non-statisticians and colleagues, so that they can communicate statistical ideas to other statisticians, to nonstatistical professionals

(clients), to the public, and to the media" (2009, p. 1). Ron Wasserstein, the current ASA Executive Director stated, "We must increase the visibility of our profession" (2015, p. 96) and that increasing the visibility of statistics and data science requires speaking "freely and definitively about the power and impact of statistics and its key role in decision-making" (2015, p. 98).

Following this logic, Olubusoye et al. (2021) states that statisticians and data scientists should—in addition to communicating *findings* and *conclusions* based on analyses of data—also collaborate with domain experts to produce *recommendations for action* and create plans for implementing these recommendations to transform evidence into action. Vance and Love (2021) describes lessons learned from the global LISA 2020 Network about transforming evidence into action. We believe that it is through action in collaboration with domain experts that statisticians and data scientists will achieve impact.

How statisticians and data scientists are to impel others to act to achieve positive impact and how to teach this skill are still open questions. Janice Derr motivates her book *Statistical Consulting: A Guide to Effective Communication* by stating: "The extent to which your client understands and accepts your recommendation also depends on your communication skills. This is why we say that skills in communication are *enabling* skills; they enable you to make the best use of your expertise in statistics" (2000, p. 2). Also according to Derr, gathering accurate and complete information about the technical nature of a client's project (i.e., *understanding*) is one of the key tasks in statistical consulting (2000). One way for a statistician to progress along the path from consulting to collaboration to leadership is to elevate the notion of *understanding* the facts of a project to an understanding of their context so as to make an impact on the project or client organization (Love et al. 2017). Vance (2015) calls on academic

consulting centers and collaboration laboratories to focus on educating and training students to increase their impacts, but does not identify specific ways for doing so.

This paper attempts to address this gap. Based on our more than 50 years of experience collaborating with domain experts, we believe that *shared understanding* is the basis for *action* and that statistics and data science educators can teach students and practitioners how to create shared understanding in the context of interdisciplinary collaborations such as those occurring on the job (Marquardt 1979), in a team-based data science course (Vance 2021), in a consulting or collaboration course (Jeske et al. 2007), or a capstone course (Martonosi and Williams 2016) as well as in other interpersonal and cooperative situations arising in the classroom or professional situations such as during group projects (Sisto 2009), peer tutoring (Roseth et al. 2008), peer assessments and feedback (Hall and Vance 2010), and mentoring (Vance et al. 2017a; b).

In the statistics and data science education literature, this concept of *shared understanding* is briefly mentioned by Vance and Smith (2019) as one of two aims for asking great questions and for listening to, paraphrasing, and summarizing the responses. Within their ASCCR Frame for collaboration describing five essential elements of collaboration (Attitude-Structure-Content-Communication-Relationship), they imply that the objectives of communication are to create shared understanding and strengthen the relationship between statistician/data scientist and domain expert. Vance (2020) further elaborates on this concept by making the case that the goal of communication in statistics and data science collaborations is to create shared understanding, which facilitates both making a deep contribution to the project and strengthening the relationship between statistician/data scientist and domain expert.

However, none of the publications cited above actually explain the concept of shared understanding, how we can create it with domain experts and other stakeholders with whom we interact, or how we can teach it. In this paper we will explore and explain the concept of shared understanding in Section 2, beginning with an explanation of the precursory concept of common knowledge. We describe our process in Section 3 for teaching how to create shared understanding. In Section 4 we discuss the relevance of creating shared understanding for interdisciplinary collaboration and statistics and data science education and discuss potential limitations. We conclude in Section 5.

2. Common Knowledge and Shared Understanding

According to our theory of collaboration, the goal of communication in statistics and data science collaborations is to create shared understanding. Furthermore, shared understanding is the basis for action toward making a deep contribution in the domain of application and is a process by which relationships are strengthened.

Shared understanding is the result of a multistep process by which information/facts about a project are exchanged between parties, common knowledge of these facts is established, and the relevance and usefulness of the facts to achieve the goals of the collaboration also becomes common knowledge. In this section we will explain *shared understanding*, beginning with an explanation of the concept of *common knowledge*, upon which our explanation of shared understanding relies.

In our explanations we use the term *statistician* (and the female pronoun) to refer to a data scientist, biostatistician, applied mathematician, statistical collaborator, or any other technical expert. We use the term *domain expert* (and the male pronoun) to refer to a client, customer, colleague, peer, mentee, or anyone else who has expertise in an area other than that of the statistician. A statistics or data science student can assume either role because she often has useful technical expertise as a statistician as well as

expertise as a domain expert in what he knows, what he doesn't know, and how he learns best. In this sense, our concepts can be applied in the classroom to create shared understanding between the educator and student.

2.1 Explanation of Common Knowledge

A statistician and a domain expert achieve the state of *common knowledge* when the domain expert communicates a message that the statistician registers about a relevant aspect of the project, the statistician paraphrases the message to reveal her interpretation of the information, and the domain expert confirms that the statistician's interpretation matches what he intended to communicate. Both the paraphrasing/interpretation and the confirmation that the interpretation matches the original intent of the message are examples of *feedback*. For common knowledge to occur, the feedback should be explicit so that each party knows what the other party knows about the message being communicated.

For example, in a project to analyze survey data, the statistician may ask the domain expert how he distributed the survey, and the answer could be that the domain expert posted the survey to Twitter and hoped it would spread widely. A fact communicated is that the survey was conducted via Twitter. To create common knowledge (and eventually shared understanding) around this fact, the statistician could provide feedback by paraphrasing, "The survey was posted to Twitter and so we might say that the sample of people who responded was a convenience sample in that there was no intentional randomization of who was or was not asked to respond to the survey, those who did respond were the ones who happened to see the tweet and decided to respond within a specified timeframe, and this sample is not representative of a specific population." At this moment, it would not be known to the statistician whether the domain expert understood the concept of a convenience sample and that his survey

relied on such a sample. To make this known (and thereby finalizing the fact of the survey distribution method to be common knowledge), the statistician would require explicit feedback from the domain expert, which could, in response to the question, "Am I understanding correctly?", come in the form of a reply such as, "Yes, exactly! I did not randomly select who would respond to the survey. Rather, I just put it out there, and whoever happened to see it and respond is who is represented in my data." At this point both parties know that both parties understand how the survey was distributed and we can say that common knowledge about this fact has been achieved.

An analogy for common knowledge is the game of chess, which is described to be a game of perfect information (Schwalbe and Walker 2001). When played at a high enough level, the rules of chess are known to each player and each player knows that the other player knows all of the rules. The current state of the game is known by each player, and the entire history of moves is also known to each player who also knows that the other knows the current state of the game and all previous moves. In chess, all the facts are on the table and each player knows that the other player knows them. In a collaboration, common knowledge occurs when facts about the project are on the table and all parties know that the other parties know these facts.

In a two-person collaboration, common knowledge occurs when both statistician and domain expert have the same interpretation of a concept or idea, they each know that the other knows about this concept, and they know that the other knows that they know the concept, ad infinitum. Ordinary knowledge may be something that one person knows, whereas common knowledge is something both people know that both people know. In other words, common knowledge is a fact that is known to be known by all parties.

Common knowledge is achieved by statements rather than thoughts. Figure 1 illustrates how a mutual declaration that a fact is common knowledge can create common knowledge.
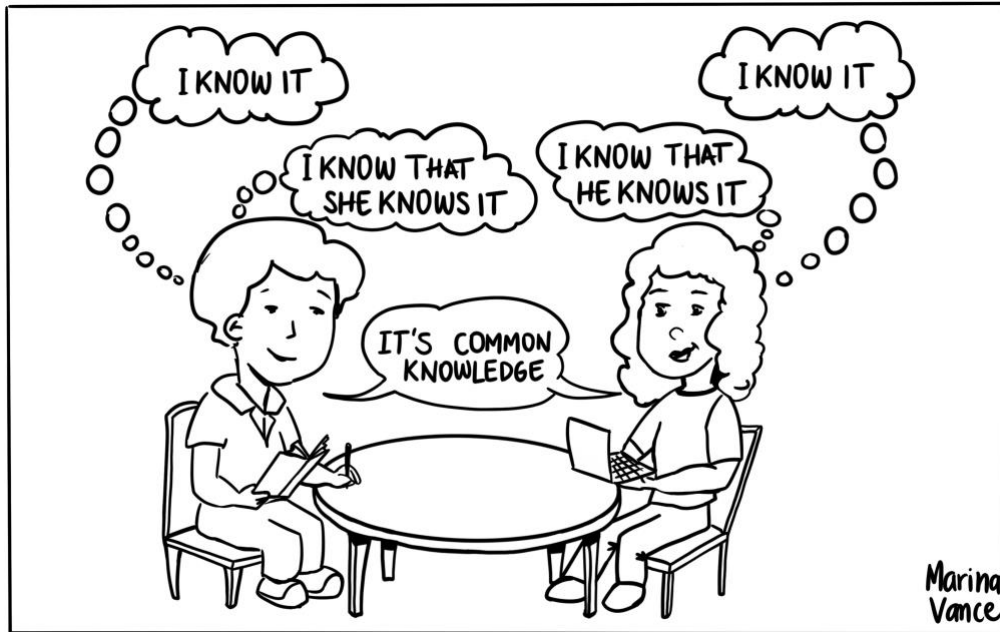

Figure 1: Diagram of individual thoughts of a domain expert and statistician about their knowledge and their joint declaration that actually achieves common knowledge.

Formally, common knowledge is a term used and defined in the literature of game theory (Lewis 1969), sociology (Friedell 1969), philosophy and logic (Schiffer 1972), probability and statistics (Aumann 1976), communication (Harman 1977; Piotr Labenz 2012), and other fields. Our preferred technical definition is the "fixed point" circular definition provided by Barwise (1988) in which a fact f is common knowledge among agents A and B if and only if A and B know *(f and ck)*, where *ck* is the fact that f is common knowledge among A and B.

One way to operationalize this notion of common knowledge in a statistics or data science collaboration is for the statistician to paraphrase a fact stated by the domain expert and to write it on a mutually visible whiteboard or a document of shared notes

(such as a shared Google document). In this way, all parties know both the fact written down and that this fact is common knowledge.

Common knowledge is a powerful concept. In statistics and data science collaborations, common knowledge is the foundation of shared understanding (which will be explained in the next section). An example of the power of common knowledge in ordinary life is that it gives cash/money its value. Common knowledge makes a mere piece of paper (e.g., a $100 bill) valuable. Everyone knows the fact that money can be exchanged for goods and services. However, even if a customer and store owner agree that $100 cash is a fair price for some goods, the store owner must know that other people value money similarly. Otherwise, he risks being stuck with a $100 bill that no one else values. Therefore, it is the common knowledge that everyone knows that a $100 bill has value that bestows value on that particular piece of paper. Figure 2 illustrates this point.

Figure 2: Common knowledge is a powerful concept. It is common knowledge of the value of money that provides money with its value.

2.2 Explanation of Shared Understanding

In a two-person interdisciplinary collaboration, *shared understanding* occurs when the statistician and domain expert have achieved common knowledge about a concept, fact, or idea and its relevance for the project.

In our previous example, the statistician and domain expert established common knowledge about the fact that the project data are from a convenience sample of Twitter users. If statisticians and domain experts were 100% perfectly logical, common knowledge of a fact would be sufficient to appropriately act upon that fact (Parikh 2005). In this example, statistically appropriate action might be to use descriptive statistics to summarize the survey data and to acknowledge in the discussion that the results—while potentially interesting—cannot be generalized to a larger population because they come from a convenience sample.

Unfortunately, none of the statisticians and domain experts we work with—ourselves included—are logically omniscient. Therefore, the extra step of creating common knowledge about the relevance of a fact is necessary to create shared understanding and appropriate action.

An analogy for shared understanding in an interdisciplinary collaboration is a chess master explaining her strategy to a pupil during a game of chess. Between the two it is common knowledge which pieces are on the board and how they got there. But the pupil may not know why the master made her moves or what she intends to do next. **The chess master creates shared understanding by explaining her motivation for each move and her analysis of the implications of potential moves.** With or without the explanation of strategy, the pupil still has all of the relevant information to make his next move. But with the master's explanation, the pupil has a deeper understanding of the game that can guide his future actions. The "master" in this analogy can be both statistician and domain expert because both have expertise the other lacks, and therefore both parties should explain the relevance of their "moves." In a collaboration, shared understanding occurs when relevant facts about the project are on the table, all parties know that the other parties know these facts, and all parties know *why* the facts are on

the table, i.e., they know the relevance of these facts. Such understanding helps guide future actions in the collaboration.

A specific example of creating shared understanding in a statistics or data science collaboration meeting is what Zahn (2019) calls the Time Conversation. The length of a typical one-on-one collaboration meeting may be 15 minutes, 60 minutes, or 90 minutes depending on the organizational norms. Shared understanding of the length of a specific meeting would occur if the statistician were to ask, "I have this meeting scheduled on my calendar for one hour. Does that time work for you? … [Yes, it does.]" and then follow up with explicit confirmation of this understanding and its relevance for the project, "… Great! I know that our last meeting ran long, would it work for you today if we went for 15–30 minutes beyond that? … [No, I have another meeting in 75 minutes.] Got it. I will set an alarm for 50 minutes from now so we can work efficiently to address all of our agenda items and still have 10 minutes to summarize and wrap-up, as that will give you a 15-minute buffer before your next appointment…. [Sounds great, thank you!]"

In that scenario, both people have common knowledge about the length of time for the meeting. In addition, there is common knowledge about *why* the meeting will last 60 minutes, i.e., for them to work toward achieving the goals of the project and so the domain expert can arrive on time to his next meeting. To achieve shared understanding (or common knowledge), there must be no doubt in either party if the other person's interpretation of an idea matches one's own.

An example of incomplete understanding due to possessing different interpretations of the length of the meeting (i.e., a *misunderstanding*) would be if the statistician assumed that, despite scheduling 60 minutes, the domain expert actually had

90 minutes because their previous meeting had lasted for 90 minutes, even though it had also been scheduled for 60 minutes.

An example of incomplete understanding due to doubt about the other's knowledge or assumptions (i.e., a breakdown in common knowledge, which we call *doubtful understanding*) about the length of the meeting would be if both the statistician and domain expert assumed that the meeting would last 60 minutes as scheduled, but neither explicitly addressed this topic. Perhaps midway through the meeting the domain expert might think, 'Should I tell her I have another meeting 15 minutes after this one? We scheduled 60 minutes, but that was true last time when the meeting lasted 90 minutes.' Similarly, the statistician may have doubts and think, 'This meeting is quite productive. Should I start ending it now so we can finish within 60 minutes, or should we try to finish everything on the agenda and go for 90 minutes like last time?' Had the two parties created shared understanding about the length of the meeting and why exactly 60 minutes were required, these doubts would have been removed.

Figure 3 shows a progression of four communication scenarios about the time available for a collaboration meeting. Panel A shows an example of *misunderstanding*, or a disagreement over the facts that occurs when we do not explicitly state our thoughts or assumptions. Panel B shows *doubtful understanding*, which is mutual knowledge of the facts (e.g., they both have 60 minutes scheduled on their calendars) but uncertainty about whether both parties understand and interpret the facts the same way, i.e., uncertainty about how the other will act on this knowledge. Panel C illustrates *common knowledge*, i.e., knowing that the fact is common knowledge but not why the fact is relevant. Panel D illustrates *shared understanding*: agreement on the facts and their relevance that results in certitude that both parties have the same understanding and interpretation of the facts.
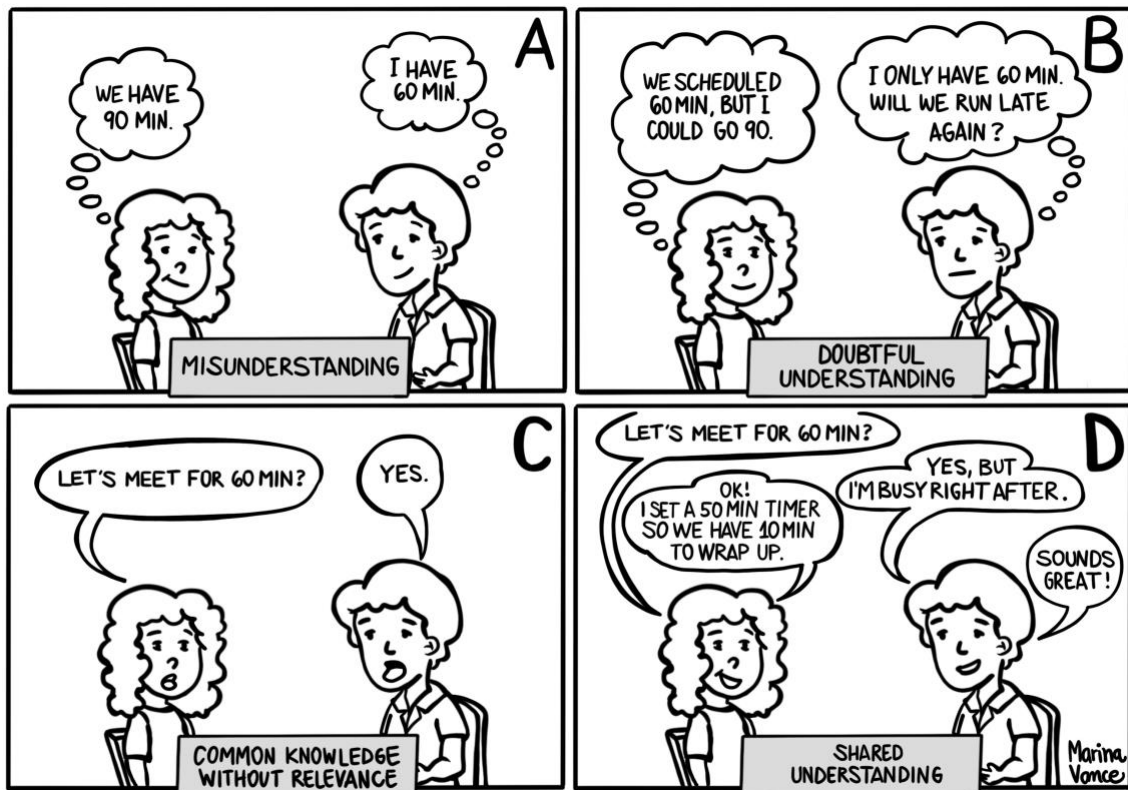
Figure 3: Communication scenarios between statistician and domain expert showing misunderstanding (panel A), doubtful understanding (B), common knowledge (C), and shared understanding (D).

The concept of shared understanding has been defined or used in engineering systems science (Smart et al. 2009), design science (Piirainen et al. 2012; Walthall et al. 2011), and collaboration engineering (Bittner and Leimeister 2014). Our conception of shared understanding was developed independently of this literature based on our experience teaching collaboration to statisticians. Formally, we define a fact f of a project to be an element of *shared understanding* among agents A and B if and only if A and B know *(g, f, r,* and *ck)*, where g are the goals of the project, r is the relevance of f toward achieving g, and ck is the fact that *g, f,* and r are common knowledge among A and B. In other words, a fact is an element of the shared understanding between statistician and domain expert if and only if the fact is common knowledge between both parties and the relevance or usefulness of the fact toward achieving the project's

goals is also common knowledge. Figure 4 shows a diagram of this conceptual model of shared understanding.
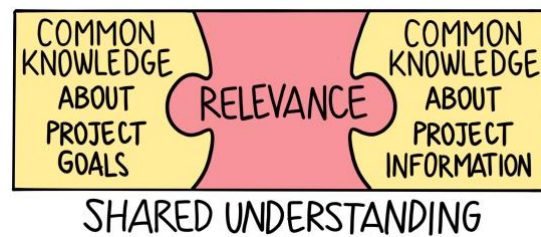

Figure 4: Shared understanding is common knowledge of the relevance of facts to the goals of a project.

   A fact useful for achieving a goal of a collaboration might not be known to be useful, in which case it is not a useful fact. For example, in analyzing the results of a survey, knowing in what ways the sample is or is not representative of the population is useful information for achieving a goal of making inferences about the population from the sample. For a domain expert to effectively engage in discussion about the methods of sampling and their implications for the representativeness of the sample, he must understand the relevance of this issue for achieving his goals. It is not the sole responsibility of the statistician to dig for this information, and it should not be up to her to unilaterally decide whether a sample is sufficiently representative. The domain expert must know why the statistician is digging and must collaborate with her to unearth the useful facts of the project.

3. A 5-Step Process for Creating and Teaching Shared Understanding About Domain Issues

Our recommended process for achieving common knowledge and creating shared understanding of project domain issues begins by asking great questions (Vance and Smith 2021); listening to, paraphrasing, and summarizing the responses; and iterating as necessary. Below we describe five steps we use for teaching how to create shared understanding about project domain issues. This process can be taught by teachers of

collaboration and implemented by beginning and experienced statisticians to increase the effectiveness of their collaborations. Providing helpful feedback (Michaelsen and Schultheiss 1989) and acting upon feedback received from the domain expert is interspersed within every step.

*Step 1: Make the **goals** of the collaboration common knowledge*

The goals of the project determine which facts about the project are relevant. Making these goals common knowledge is therefore the initial step toward creating shared understanding in a statistics or data science collaboration. We recommend making the discussion of goals an explicit agenda item during the initial collaboration meeting. Vance (2020) describes three stages of a conversation about goals:

(1) Prefaced by her intent for initiating this conversation, the statistician states her goals for the collaboration

(2) The statistician asks the domain expert about his overall goals for the project and specific goals for the current meeting

(3) The statistician listens, paraphrases, and summarizes the domain expert's goals and how they overlap with hers.

An example of a minimal conversation about goals is: "Agreeing on our goals for the project and this meeting will help me be a better statistician. My goals are to help you achieve your goals, help make an impact, and create a strong relationship. What are your goals for the project? … Considering that, what would you like to accomplish in the meeting today?" … followed by the statistician listening, paraphrasing, and summarizing the domain expert's goals and how they overlap with hers.

*Step 2: Elicit information about the project by **asking** great questions and **listening** to the responses*

Derr (2000, chap. 5) provides examples of good questions to elicit relevant information about the domain of application and the statistical issues of a project. Vance and Smith (2021) defines a great question as one that elicits information useful for answering the research/business/policy questions of the project and is asked in a way that strengthens the relationship with the domain expert. They provide examples of great questions.

Listening to the responses to register the content and collect information/facts about the project is an essential communication skill. A statistician can improve her listening skills through study, practice, and reflection. We recommend familiarizing oneself with conventional active listening tips such as positioning oneself to facilitate eye contact and note taking; ensuring that one's posture and body movements communicate a connection to what one is hearing (e.g., nodding, leaning in); and keeping one's eyes, ears, and mind open to register without evaluating what is being communicated verbally and nonverbally. Dunkel (1991) and Kök (2018) review barriers to listening and categorize them into internal factors (e.g., distractions, disinterest, inattentiveness, detouring, and emotions) and external factors (e.g., rate of delivery of speech, linguistic complexity, and organization).

From our more than 50 years of experience collaborating with domain experts we believe that there are three primary categories of barriers to listening: physical (e.g., a noisy room is distracting, visual distractions, being overly tired), mental (e.g., thinking about something else, thinking about what one wants to say, thinking about implications of what the domain expert just said, hard to understand language or accents), and emotional (e.g., lack of interest, anxiety). These barriers contribute to three common reasons statisticians may fail to listen:

(1) Too busy thinking about something else

(2) Difficulty understanding the domain expert

(3) Limited opportunities, i.e., the statistician is talking the whole time or the domain expert is especially reticent.

Four tips that we have found useful to help ourselves, our students, and our colleagues listen better are:

(1) Prioritize the *Fundamental Law of Statistical Collaboration*, "Seek first to understand, then to be understood" (Covey 1989; Vance and Smith 2019). Provide opportunities for domain experts to talk and for oneself to listen.

(2) Be patient. "Be in the now." Focus on the present rather than the past (i.e., what the domain expert said a minute ago) or the future (i.e., what the implications may be of what the domain expert said). Acknowledge what is said now and evaluate it later.

(3) Manage one's distractions. Preemptively eliminate common distractions. When distracted, one should be honest with the domain expert and ask him to repeat what one missed.

(4) Listen to what the domain expert says you said. The mantra of 2017 ASA President Barry Nussbaum is, "It's not what we [statisticians] said, it's not what they [domain experts] heard, it's what they say they heard" (2018, p. 491). Nussbaum states: "The statistician has an obligation to lend insight and try to ascertain if the message is getting through" (2017, p. 3). A good way to understand what message the domain expert will be communicating to others is to ask him, "How would you say these statistical results fit in with the big picture of your project?" or "How will you be explaining what we discussed to your advisor/supervisor/the media?"

*Step 3: **Paraphrase** the information and seek feedback to create common knowledge*

Paraphrase (i.e., restate in one's own words) the information the domain expert provided about the project and ask for feedback about whether one's understanding is correct (i.e., whether your understanding matches his understanding). An example of this is, "I want to make sure I understand how the respondents were selected. They were contacted by direct messages on Twitter? Is there anything I am missing?"

We find that more often than not, our initial paraphrasing provides the time and space for the domain expert to provide additional clarifying (and useful!) information. We then paraphrase this additional information and ask for feedback again. Our goal is to repeat this process until the domain expert responds, "Exactly!" to our paraphrasing, which indicates to us that common knowledge has been achieved. With practice, two rounds of paraphrasing are often sufficient to create common knowledge.

Tips for paraphrasing we have found useful include:

(1) State your intent. Tell the domain expert why you are paraphrasing (i.e., to clarify your understanding and establish common knowledge of the facts of the project).

(2) Use some of the domain expert's nouns and verbs. To test whether you understand them, practice using terms from the domain new to you. Surround these new terms with your own words, analogies, diagrams, and examples.

(3) Paraphrasing is an iterative process: check your understanding, revise, check again.

(4) During in-person meetings, writing key information on a whiteboard enables the domain expert to see what the statistician understood to be important and immediately correct any misunderstandings. During remote, online meetings we

recommend creating a shared, mutually editable document (see

bit.ly/gdoccollabtemplate for an example) for all parties to record notes. Both

the whiteboard and shared notes facilitate the creation of common knowledge!

A frequently asked question about paraphrasing, is "When and how often should

I do it?" We recommend paraphrasing in stages rather than all at once. When a domain

expert introduces what may be an important piece of information, let him finish his

thought and then paraphrase the new information. Whenever a new idea shifts your

understanding or mental model of the project, paraphrase to clarify the information and

check your understanding. Analogous questions to "How often should I paraphrase?"

are "How often should I commit and push changes to GitHub?", "How often should I

save a Word document?", and "How often should I say, 'I love you' to my significant

other?" Our answers are that it depends on your environment, your preferences, and the

preferences of those with whom you are working. Generally, we recommend

paraphrasing more often than one may be accustomed to doing.

*Step 4: **Summarize** the information and its **relevance** to achieving the goals of the project*

To convert common knowledge of facts and information about the project into shared

understanding useful for achieving the goals of the project, the statistician and domain

expert must both understand the *relevance* of the facts and how they fit together to

inform a solution and a successful implementation of the solution.

Doug Zahn (2019) created the POWER process, which is an acronym for five

structural aspects of effective collaboration meetings: Prepare, Open, Work, End, and

Reflect. Statisticians in the Laboratory for Interdisciplinary Statistical Analysis (LISA)

and at Cal Poly have used POWER to structure meetings with domain experts since

2010 and 2013 respectively. Below we highlight five key aspects of a project and when—using POWER—they should be summarized to create shared understanding.

(1) The domain expert's **goals and** what he **wants** to achieve during a particular meeting. These should be summarized during the Opening of a meeting, with the goals discussed at the initial meeting and revisited in subsequent meetings.

(2) The domain expert's **timeline** for both near-term deadlines and any longer-term deadlines should also be summarized during the Opening of a meeting. These should also be summarized at the End of a meeting.

(3) The domain expert's **research, business, or policy questions** and why they are important should be summarized during the Open or Work phases of a meeting *before* moving on to the quantitative aspects of the project in the Work phase.

(4) Whenever **statistics or data science issues** are addressed during the Work phase of a meeting, they should be summarized to create shared understanding before moving on to the next topic.

(5) What was decided, who will do what by when, and what the specific **next step action items** are should be summarized during the Ending of the meeting, preferably in writing via a shared notes document or an email sent shortly after the meeting.

Generally, a collaborative statistician should summarize to complete a conversation before moving on to the next topic. An analogy for summarizing is that it moves information from working (short-term) memory into long-term memory. Sending an email or a report to the domain expert summarizing the main points of shared understanding immediately after a meeting will save the long-term memory of the meeting in a form that can be searched and will be a useful reference for the remainder of the project. The summary will be the permanent record of the meeting, whereas

whatever was written on the whiteboard is fleeting and will be washed away. One can take a picture of the whiteboard and include that in the written summary of the meeting.

In addition to the five aspects above, the **roles** of all parties in the collaboration should be summarized, preferably in writing, and preferably before the statistician engages in significant effort (e.g., before cleaning or analyzing data). Statisticians can play many roles on a project, and the role of the statistician is an important decision to agree upon (Halvorsen et al. 2020). For example, she can be an advisor, consultant, collaborator, mentor. She can help design a study or experiment, collect data, analyze and model the data, interpret the results, make decisions, and help the domain expert take action (Vance and Pruitt in press). In our experience, a statistician's role on a project often evolves through the course of the project. Shared understanding of roles is essential for an equitable distribution of the efforts and the outcomes of the project (e.g., co-authorship).

Tips for summarizing to create shared understanding include stating your intent before summarizing, using visual aids such as sketching a casual diagram to indicate potential relationships between variables (see Pearl 1995), **focusing on how the information will be useful for achieving the project's goals**, and reserving adequate time (10–20% of the meeting time) to complete an effective, final summary.

*Step 5: Apply the shared understanding to accomplish meaningful **action***
A statistician can increase the impact of her work by being mindful of this final step and helping the domain expert develop and implement a plan for action based on the generated findings, conclusions, and recommendations. One way to do this is to have a complete conversation at the beginning of the project about what potential actions would achieve the desired impacts of the project. Such a conversation could reframe the initial goals of the project to include meaningful actions beyond the completion of an

analysis or the presentation of a report or manuscript. An example for what the statistician might ask during the initial meeting is, "After the analyses are completed and these questions answered, what do you want to accomplish? How will this impact your domain? What actions would you like to see taken? Knowing this will help me better understand the context of the project and how I can help you achieve the desired impacts."

Using this process to teach how to create shared understanding

Classroom instructors and mentors of statisticians can use this five-step process to teach shared understanding. For the past year, we have been teaching our undergraduate and graduate students how to create shared understanding in our capstone statistical consulting and collaboration courses via an education and training program comprising preparation, practice, doing, reflecting, and mentoring. Our students read this paper as part of their preparation for learning how to create shared understanding. They practice during in-class exercises (available at osf.io/wya7g), and their performance (doing) is assessed using a rubric (also available at osf.io/wya7g). Collaboration meetings are recorded so that students and faculty are able to engage in reflection and mentoring during Video Coaching and Feedback Sessions, during which we review a few short clips (1–5 minutes) from video-recorded collaboration meetings (McCulloch et al. 1985; Vance 2014). Students watch the clips for specific aspects of collaboration, including opportunities seized or missed to achieve common knowledge and create shared understanding.

Summary of how to create shared understanding of domain issues

We will summarize this section by extending our chess analogy. The initial step in our process for creating shared understanding (agreeing on the **goals**) is like agreeing on the

rules of the chess game (e.g., time limit for the game) and ensuring that the rules are common knowledge. The second step (getting the information about the project via **listening**) is like collecting the chess pieces in preparation for play. The next step (**paraphrasing** to make the information common knowledge) is like arranging the chess pieces on the board. The fourth step (putting the information into context and **summarizing its relevance**) is like a chess coach explaining her strategy to a pupil. The final step (putting the shared understanding into **action** for impact) is like making a move to win the game. In chess, who plays white and moves first is decided before play begins. In a collaboration, **roles** should be discussed early and often as circumstances evolve.

We suggest beginners start learning how to create shared understanding about the goals of the project and the structure of a collaboration meeting while experts focus on learning how to create shared understanding about the impacts of the project and how the ultimate solution(s) will be implemented.

4. Discussion

4.1 Relevance of Shared Understanding

Barwise (1988, p. 368) wrote about common knowledge: "Information travels at the speed of logic, genuine knowledge only travels at the speed of cognition and inference." A database has information; a robot needs knowledge about the information it has to be able to do useful things with the information.

We believe that *collaboration travels at the speed of shared understanding.* The more shared understanding there is, the more effective the collaboration will be. According to Vance (2020), one of the terminal goals of a collaboration is to make a deep contribution that will make an impact in the domain expert's domain or within the

fields of statistics and data science. Shared understanding enables statisticians to make impacts because *useful action travels at the speed of collaboration* and *shared understanding is the basis for action*.

Shared understanding guides the actions of both statistician and domain expert and helps each party make appropriate decisions. When operating from a basis of shared understanding, the domain expert does not have to guess about what context/background is relevant for the statistician because he learns through the process of collaboration what is relevant. For the statistician, she does not have to guess at the background when analyzing data. She understands the domain context, which is important for making statistical decisions.

An example of a lack of shared understanding and its consequences is from an interdisciplinary research project to determine whether investments in constructing water supply infrastructure (e.g., pumps, holding tanks, distribution pipes) in Senegal would cause enough economic activity to pay for the initial investment (Hall et al. 2015). The first author was contracted to analyze the data and found that an important source of water—surface water—was not queried due to a lack of shared understanding between the research team and the survey enumerators. Both parties knew that they wanted to measure all sources of water households used. Yet, because it was difficult to ascertain how much surface water was used daily and to translate that concept into the local language of the households, and because the survey enumeration team did not understand how the data would ultimately be used, they dropped this concept from the survey, and the research team was not aware of this omission. The lack of this data due to lack of shared understanding made the statistical analyses much more complicated and time consuming, necessitated that new research questions be devised that *could* be

answered with the available data, and resulted in the project having less impact on policy than desired (Hall et al. 2014b).

This experience prompted the first author—on a new project—to embed an "on-the-ground statistician" within the survey enumeration team to ensure that no relevant questions were accidentally dropped from the survey, flag suspicious data points, clean the data in real time, and create shared understanding with the enumerators about the goals of data collection. The result was a high-quality dataset that could be easily modeled and analyzed by the same on-the-ground statisticians who now had a shared understanding of the local context of the data production (Seiss et al. 2014; Van Houweling et al. 2017) and a much greater potential to achieve impact (Hall et al. 2014a).

4.2 Explaining Statistics and Data Science Concepts

Section 3 focused on creating shared understanding of project domain issues. Equally important is for the domain expert to understand the statistics and data science issues of his project, which we intend to be the subject of a future manuscript. To create shared understanding of technical concepts, we use and teach the ADEPT method, which was developed by Azad (2015) to explain mathematical concepts using Analogies, Diagrams, Examples, Plain language, and Technical definitions (ADEPT). Often a plain language explanation of the concept paired with a diagram or an example is sufficient for the domain expert to understand the concept. Verifying that the domain expert does understand the concept and its relevance to the project creates shared understanding. If it is not clear that the domain expert understands the concept or its relevance, we might share an analogy that relates the unknown concept to something familiar, provide another example or diagram that directly relates to his work, and/or ask how much technical detail he desires to complete his understanding.

4.3 Shared Understanding Throughout the ASCCR Frame

Vance and Smith (2019) mentioned shared understanding within the ASCCR (Attitude-Structure-Content-Communication-Relationship) Frame as one of the objectives of Communication. We believe that shared understanding is a more comprehensive and influential concept for teaching interdisciplinary collaboration than Vance and Smith (2019) suggested, and therefore, a statistician should aim to explicitly create shared understanding with the domain expert in all five components of ASCCR. For Attitude, both statistician and domain expert should agree on the "roles and goals" for the project. The statistician should propose a Structure for facilitating meetings and for working on the project outside of meetings, including a proposed communication plan and a timeline for deliverables. The domain expert should be empowered to propose alternative structures, plans, or timelines. The understanding of whatever is agreed upon should be explicitly shared by both statistician and domain expert. There should be shared understanding of the Content of the project in all three aspects of the $Q_1Q_2Q_3$ process (Leman et al. 2015; Vance 2019). Communication methods—including asking great questions; listening, paraphrasing, and summarizing; explaining statistics using the ADEPT method; and providing and receiving feedback—are the means for creating shared understanding. Finally, we recommend the statistician create shared understanding around the fact that creating a strong Relationship with the domain expert is an explicit goal of the collaboration. In our experience, the simple act of explicitly setting a goal to create a strong relationship leads to stronger relationships.

4.4 Potential Limitations

We believe that achieving common knowledge and creating shared understanding throughout all aspects of a collaboration is an optimal strategy, something toward which

statisticians (and domain experts) should aspire. In practice, shared understanding lies on a continuum as something that can exist to a greater or lesser extent (Smart et al. 2009). Creating it may be limited by three interrelated factors: willingness, ability, and time.

Willingness. To create shared understanding, the statistician must be willing to expend the effort, which usually entails taking extra steps in every conversation with a domain expert to paraphrase and summarize the relevance of the information presented, to verify her own understanding, to check for his understanding, and to help him develop and implement a plan for action that will lead to impact. A statistician must decide for herself how much impact she wants to have and how willing she is to create the shared understanding necessary to achieve it. A domain expert must also be willing to engage in these conversations, and the statistician can influence his willingness to do so by emphasizing how this process will lead to the accomplishment of his goals.

Ability. Is it possible to be 100% certain about what someone else knows? In practice, statisticians and domain experts are not logically omniscient, and therefore perfect common knowledge (and thus shared understanding) may be difficult or impossible to obtain. We believe that shared understanding exists along a continuum and that complete shared understanding is an aspirational goal of communication. If we substitute "confidence" instead of "certainty" that common knowledge of the relevance of the project information to the project's goals has been achieved, we can be confident that we have created sufficient shared understanding within a collaboration.

This paper has focused on creating shared understanding between two parties. When three or more parties are involved, the process is much more complicated, especially if one or more parties is missing from the conversations. In either case,

written documentation of project goals, facts, and the relevance of the facts toward achieving the goals will help in creating more shared understanding.

Time. In our experience as statisticians who have collaborated on more than 1000 projects, the only practical limitation to how much shared understanding we can create with domain experts is time. Even with our experience, we feel we still spend too little time creating shared understanding. The more time we spend engaged with a domain expert on a topic, the more shared understanding we can create and the smoother the project will proceed. Devoting too little time risks advancing to the next stage of the project on the basis of a misunderstanding or doubtful understanding, which can result in providing bad advice, using the wrong data, creating inappropriate models, conducting incorrect analyses, damaging the relationship, and nil or negative impact.

Time constraints exist, however, for every project. How much time is "enough" to spend on creating shared understanding depends on one's environment, preferences, and the preferences of those with whom one is working. Generally, we recommend spending more time creating shared understanding than one may be accustomed to doing. We believe that a statistician can move on to the next topic when she feels that she has achieved common knowledge on the topic and understands the relevance of that topic to achieving the project goals.

It may be tempting to use the pressures of time to skip or rush through steps necessary for creating shared understanding. In our experience, this merely results in our spending more time later because of our ignorance and need to make guesses about the best path forward. Analyzing data requires making many decisions, and without shared understanding, doing so can be difficult, frustrating, and mentally exhausting. When we have created shared understanding, the analyses tend to be easy, enjoyable, and invigorating. We believe that creating shared understanding saves us time and

aggravation and helps us appreciate the many benefits of being collaborative statisticians.

5. Conclusion

In this paper our goal has been to create shared understanding about shared understanding, which is a powerful concept relevant throughout statistics and data science collaborations. We believe that incorporating this concept into one's practice of statistics or data science and following the steps outlined above will result in statisticians having more impact on projects and throughout their careers. We are hopeful that this paper will be a useful starting point for other educators intent on helping their students, colleagues, or mentees learn to create shared understanding in their interdisciplinary collaborations. While developed and explored in the context of statistics and data science collaborations, we believe that the concepts and techniques presented here are useful for all who collaborate and all who want to teach others to collaborate better.

References

Aumann, R. J. (1976), "Agreeing to Disagree," *The Annals of Statistics*, Institute of Mathematical Statistics, 4, 1236–1239.

Azad, K. (2015), *Math, Better Explained: Learn to Unlock Your Math Intuition*, Scotts Valley, CA: CreateSpace Independent Publishing Platform.

Barwise, J. (1988), "Three views of common knowledge," in *Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge*, TARK '88, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 365–379.

Bittner, E. A. C., and Leimeister, J. M. (2014), "Creating Shared Understanding in Heterogeneous Work Groups: Why It Matters and How to Achieve It," *Journal of Management Information Systems*, 31, 111–144. https://doi.org/10.2753/MIS0742-1222310106.

Covey, S. R. (1989), *The Seven Habits of Highly Effective People: Restoring the Character Ethic*, New York: Simon and Schuster.

Derr, J. (2000), *Statistical Consulting: A Guide to Effective Communication*, Pacific Grove, CA: Duxbury Press.

Dunkel, P. (1991), "Listening in the Native and Second/Foreign Language: Toward an Integration of Research and Practice," *TESOL Quarterly*, 25, 431–457. https://doi.org/10.2307/3586979.

Friedell, M. F. (1969), "On the structure of shared awareness," *Behavioral Science*, 14, 28–39. https://doi.org/10.1002/bs.3830140105.

Gnanadesikan, R. (1990), "Looking Ahead: Cross-Disciplinary Opportunities for Statistics," *The American Statistician*, [American Statistical Association, Taylor & Francis, Ltd.], 44, 121–125. https://doi.org/10.2307/2684145.

Hall, R. P., Davis, J., van Houweling, E., Vance, E. A., Seiss, M., and Russel, K. (2014a), *Impact Evaluation of the Mozambique Rural Water Supply Activity*, Blacksburg, VA: Virginia Tech School of Public and International Affairs.

Hall, R. P., Vance, E. A., and van Houweling, E. (2014b), "The Productive Use of Rural Piped Water in Senegal," *Water Alternatives*, 7, 480–498.

Hall, R. P., Vance, E. A., and van Houweling, E. (2015), "Upgrading Domestic-plus Systems in Rural Senegal: An Incremental Income-cost (IC) Analysis," *Water Alternatives*, Water Alternatives Association, 8.

Hall, S., and Vance, E. A. (2010), "Improving Self-Efficacy in Statistics: Role of Self-Explanation & Feedback," *Journal of Statistics Education*, 18, 1–22. https://doi.org/10.1080/10691898.2010.11889583.

Halvorsen, K. T., Hanford, K. J., Vance, E. A., Wilson, J., and Zahn, D. (2020), "Transforming Your Stumbling Blocks into Stepping Stones," in *JSM Proceedings*, Alexandria, VA: American Statistical Association: Statistical Consulting Section, pp. 2523–2541.

Harman, G. (1977), "Review of Jonathan Bennett's Linguistic Behaviour," *Language*, 53, 417–424.

Jeske, D. R., Lesch, S. M., and Deng, H. (2007), "The Merging of Statistics Education, Consulting and Research: A Case Study," *Journal of Statistics Education*, 15, 1–19. https://doi.org/10.1080/10691898.2007.11889544.

Kök, İ. (2018), "Relationship between Listening Comprehension Strategy Use and Listening Comprehension Proficiency," *International Journal of Listening*, Routledge, 32, 163–179. https://doi.org/10.1080/10904018.2016.1276457.

Lachenbruch, P. A. (2009), "Communicating Statistics and Developing Professionals: The 2008 ASA Presidential Address," *Journal of the American Statistical Association*, Taylor & Francis, 104, 1–4. https://doi.org/10.1198/jasa.2009.0033.

Leman, S., House, L., and Hoegh, A. (2015), "Developing a New Interdisciplinary Computational Analytics Undergraduate Program: A Qualitative-Quantitative-Qualitative Approach," *The American Statistician*, 69, 397–408. https://doi.org/10.1080/00031305.2015.1090337.

Lewis, D. K. (1969), *Convention: A Philosophical Study*, Wiley-Blackwell.

Love, K., Vance, E. A., Harrell, F. E., Johnson, D. E., Kutner, M. H., Snee, R. D., and Zahn, D. (2017), "Developing a Career in the Practice of Statistics: The Mentor's Perspective," *The American Statistician*, 71, 38–46. https://doi.org/10.1080/00031305.2016.1255257.

Marquardt, D. W. (1979), "Statistical Consulting in Industry," *The American Statistician*, 33, 102–107. https://doi.org/10.2307/2683809.

Martonosi, S. E., and Williams, T. D. (2016), "A Survey of Statistical Capstone Projects," *Journal of Statistics Education*, Taylor & Francis, 24, 127–135. https://doi.org/10.1080/10691898.2016.1257927.

McCulloch, C. E., Boroto, D. R., Meeter, D., Polland, R., and Zahn, D. A. (1985), "An Expanded Approach to Educating Statistical Consultants," *The American Statistician*, 39, 159–167. https://doi.org/10.2307/2683923.

Michaelsen, L. K., and Schultheiss, E. E. (1989), "Making feedback helpful," *Organizational Behavior Teaching Review*, 13, 109–113.

Nussbaum, B. D. (2017), "It's What They Say They Heard," *Amstat News*, 3–4.

Nussbaum, B. D. (2018), "Statistics: Essential Now More Than Ever," *Journal of the American Statistical Association*, 113, 489–493. https://doi.org/10.1080/01621459.2018.1463486.

Olubusoye, O. E., Akintande, O. J., and Vance, E. A. (2021), "Transforming Evidence to Action: The Case of Election Participation in Nigeria," *CHANCE*, Taylor & Francis, 34, 13–23. https://doi.org/10.1080/09332480.2021.1979807.

Parikh, R. (2005), "Logical omniscience and common knowledge: WHAT do we know and what do WE know?," in *Proceedings of the 10th conference on Theoretical aspects of rationality and knowledge*, TARK '05, SGP: National University of Singapore, pp. 62–77.

Pearl, J. (1995), "Causal Diagrams for Empirical Research," *Biometrika*, 82, 669–688. https://doi.org/10.2307/2337329.

Piirainen, K. A., Kolfschoten, G. L., and Lukosch, S. (2012), "The joint struggle of complex engineering: a study of the challenges of collaborative design," *International Journal of Information Technology & Decision Making*, World

Scientific Publishing Co., 11, 1087–1125.
https://doi.org/10.1142/S0219622012400160.

Piotr Labenz (2012), "On common knowledge in conversation," *Semiotica*, De Gruyter Mouton, 2012, 121–137. https://doi.org/10.1515/sem-2012-0009.

Roseth, C. J., Garfield, J. B., and Ben-Zvi, D. (2008), "Collaboration in Learning and Teaching Statistics," *Journal of Statistics Education*, Taylor & Francis, 16, null. https://doi.org/10.1080/10691898.2008.11889557.

Schiffer, S. (1972), *Meaning*, Oxford, Clarendon Press.

Schwalbe, U., and Walker, P. (2001), "Zermelo and the Early History of Game Theory," *Games and Economic Behavior*, 34, 123–137. https://doi.org/10.1006/game.2000.0794.

Seiss, M., Vance, E., and Hall, R. (2014), "The Importance of Cleaning Data During Fieldwork: Evidence from Mozambique," *Survey Practice*, 7.

Sisto, M. (2009), "Can You Explain That in Plain English? Making Statistics Group Projects Work in a Multicultural Setting," *Journal of Statistics Education*, Taylor & Francis, 17, null-null. https://doi.org/10.1080/10691898.2009.11889522.

Smart, P. R., Mott, D., Sycara, K., Braines, D., Strub, M., and Shadbolt, N. R. (2009), "Shared understanding within military coalitions: A definition and review of research challenges," in *Proceedings of Knowledge Systems for Coalition Operations*, Southampton, UK, pp. 1–9.

Van Houweling, E., Hall, R., Carzolio, M., and Vance, E. (2017), "'My Neighbour Drinks Clean Water, While I Continue To Suffer': An Analysis of the Intra-Community Impacts of a Rural Water Supply Project in Mozambique," *The Journal of Development Studies*, 53, 1147–1162. https://doi.org/10.1080/00220388.2016.1224852.

Vance, E. A. (2014), "LISA Video Coaching and Feedback Sessions," *Collaboration in a Bag*, 1–1. https://doi.org/10.17605/OSF.IO/XMTCE.

Vance, E. A. (2015), "Recent Developments and Their Implications for the Future of Academic Statistical Consulting Centers," *The American Statistician*, 69, 127–137. https://doi.org/10.1080/00031305.2015.1033990.

Vance, E. A. (2019), "Content of Collaborations QQQ," *Collaboration in a Bag*, 1–4. https://doi.org/10.17605/OSF.IO/XMTCE.

Vance, E. A. (2020), "Goals for Statistics and Data Science Collaborations," in *JSM Proceedings*, Alexandria, VA: American Statistical Association: Statistical Consulting Section, pp. 2198–2209.

Vance, E. A. (2021), "Using Team-Based Learning to Teach Data Science," *Journal of Statistics and Data Science Education*, Taylor & Francis, 0, 1–27. https://doi.org/10.1080/26939169.2021.1971587.

Vance, E. A., LaLonde, D. E., and Zhang, L. (2017a), "The Big Tent for Statistics: Mentoring Required," *The American Statistician*, 71, 15–22. https://doi.org/10.1080/00031305.2016.1247016.

Vance, E. A., and Love, K. (2021), "Building Statistics and Data Science Capacity for Development," *CHANCE*, Taylor & Francis, 34, 38–46. https://doi.org/10.1080/09332480.2021.1979810.

Vance, E. A., and Pruitt, T. R. (in press), "Statistics and Data Science Collaboration Laboratories: Engines for Development," in *Promoting Statistical Practice and Collaboration in Developing Countries*, Chapman & Hall.

Vance, E. A., and Smith, H. S. (2019), "The ASCCR Frame for Learning Essential Collaboration Skills," *Journal of Statistics Education*, Taylor & Francis, 27, 265–274. https://doi.org/10.1080/10691898.2019.1687370.

Vance, E. A., and Smith, H. S. (2021), "Asking Great Questions: Part of a Theory of Communication in Interdisciplinary Collaborations," in *JSM Proceedings*, Alexandria, VA: American Statistical Association: Statistical Consulting Section, pp. 995–1008.

Vance, E. A., Tanenbaum, E., Kaur, A., Otto, M. C., and Morris, R. (2017b), "An Eight-Step Guide to Creating and Sustaining a Mentoring Program," *The American Statistician*, 71, 23–29. https://doi.org/10.1080/00031305.2016.1251493.

Walthall, C. J., Devanathan, S., Kisselburgh, L. G., Ramani, K., Hirleman, E. D., and Yang, M. C. (2011), "Evaluating Wikis as a Communicative Medium for Collaboration Within Colocated and Distributed Engineering Design Teams," *Journal of Mechanical Design*, 133, 071001. https://doi.org/10.1115/1.4004115.

Wasserstein, R. (2015), "Communicating the Power and Impact of Our Profession: A Heads Up for the Next Executive Directors of the ASA," *The American Statistician*, Taylor & Francis, 69, 96–99. https://doi.org/10.1080/00031305.2015.1031283.

Zahn, D. (2019), *Stumbling Blocks to Stepping Stones*, Bloomington, Indiana: iUniverse.