2017

Contextual information, answerability, and the logical construction of social how-to questions

Matthew John Baker
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Other Communication Commons, and the Rhetoric Commons

Recommended Citation
Baker, Matthew John, "Contextual information, answerability, and the logical construction of social how-to questions" (2017).
Graduate Theses and Dissertations. 15485.
https://lib.dr.iastate.edu/etd/15485

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Contextual information, answerability, and the logical construction of social how-to questions

by

Matthew J. Baker

A dissertation submitted to the graduate faculty in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY

Major: Rhetoric and Professional Communication

Program of Study Committee:
Jo Mackiewicz, Major Professor
David R. Russell
Stacy Tye-Williams
Bethany Gray
Zhu Zhang

Iowa State University
Ames, Iowa

2017

Copyright © Matthew J. Baker, 2017. All rights reserved.
# TABLE OF CONTENTS

LIST OF TABLES .................................................................................................................. vi

ACKNOWLEDGEMENTS........................................................................................................ viii

ABSTRACT ............................................................................................................................... ix

CHAPTER I: INTRODUCTION ................................................................................................. 1
   SQA Communication as Documentation: A Theoretical Framework ................. 6
   SQA communication as reaction and interaction ............................................... 7
   Contextual information and the logical construction of questions ................. 10
   The Focus of this Study ............................................................................................ 13
   The Goal and Contribution of this Study .............................................................. 15
   Conclusion and Overview of this Study ................................................................. 16

CHAPTER II: LITERATURE REVIEW ...................................................................................... 18
   Characteristics and Predictors of Answerability ................................................... 18
   The Relationship Between Contextual Information and Answerability (RQ1 and RQ2) ............................................................... 23
   The Relationship Between Logical Coherence and Answerability (RQ 3) ....... 28
   The Relationship Between Audience Expectations and Contextual Information (RQ 4) .................................................................................. 37

CHAPTER III: METHODS ....................................................................................................... 40
   Research Site and Participants ............................................................................. 41
   Data and Data Collection Procedures .................................................................. 44
Research Methodology and Data Analysis ........................................... 52
Analyzing contextual-information types ........................................... 53
Analyzing differences in contextual-information types of answered and unanswered questions ........................................... 54
Analyzing presence and absence of current- and desired-state information ........................................... 58
Analyzing comments ................................................................. 58
Codebook Development ............................................................... 59
Phase 1: How-to questions ........................................................... 59
Phase 2: Contextual information ...................................................... 66
Phase 3: Comments ................................................................. 71
Conclusion ..................................................................................... 73

CHAPTER IV: CONTEXTUAL INFORMATION TYPES ......................... 75
RQ1: Overview of Contextual Information Coding Results ............... 75
Current state information ................................................................. 77
  Situation ....................................................................................... 79
  Specifications .............................................................................. 83
  Examples ....................................................................................... 84
  Previous attempts ........................................................................ 87
  Thought ......................................................................................... 88
  Knowledge ..................................................................................... 89
  Sources ......................................................................................... 91
  Error ............................................................................................ 91
  Frustration ..................................................................................... 92
Desired state information ................................................................. 93
  Task ............................................................................................. 94
  Limit ............................................................................................. 97
  Specifications .............................................................................. 98
  Examples ....................................................................................... 100
  Thought ......................................................................................... 102
CHAPTER VI: CONCLUSION...................................................................................... 149

RQ1: Multiple Types of Contextual Information................................................. 149

RQ2: Question Content and Answerability....................................................... 152
   Complex ........................................................................................................ 154
   Gratitude filled .............................................................................................. 157
   Example filled ............................................................................................... 159

RQ3: Logical Coherence and Answerability ..................................................... 160

RQ4: Comments.................................................................................................. 164

Limitations ......................................................................................................... 166

Closing Thoughts .............................................................................................. 168

REFERENCES..................................................................................................... 170

APPENDIX A: SQL CODE FOR DATA DOWNLOAD ........................................ 182

APPENDIX B: RANDOMIZATION UNDER CONDITIONS OF
HETEROGENEOUS VARIANCES ........................................................................ 183
LIST OF TABLES

Table 2.1   Contextual-information types in forum and social questions.............. 25
Table 2.2   The states and actions of logical documentation aligned with observed contextual information in forums and social questions ........ 34
Table 3.1   Tag frequency on Super User and in the present study’s data set.......... 45
Table 3.2   Number of t-units and t-units per question within pilot studies and final study .................................................................................................................. 51
Table 3.3   Number of t-units and t-units per comment within pilot studies and final study .................................................................................................................. 52
Table 3.4   How-to question codebook..................................................................... 63
Table 3.5   How-to question codebook with coding results .................................... 65
Table 3.6   Contextual-information types............................................................... 67
Table 3.7   Contextual-information types requested by askers .............................. 72
Table 4.1   Contextual-information t-unit totals by code ....................................... 76
Table 4.2   Contextual-information t-unit presence or absence by total questions .... 77
Table 4.3   Examples of current-state contextual information ................................ 78
Table 4.4   Examples of desired-state contextual information ............................... 94
Table 5.1   Chi-square test results for answerability and contextual-information type .................................................................................................................. 113
Table 5.2   Contextual-information token totals by code and by answered and unanswered questions ............................................................... 117
Table 5.3   Results of randomization tests comparing the number of contextual-information types in answered and unanswered questions .............. 118
Table 5.4   Type-token ratio for contextual information within questions.............. 123
Table 5.5   Proportion test results for distinctness and contextual-information type .................................................................................................................. 124
Table 5.6   Results of randomization tests comparing the average number of words in answered and unanswered questions by contextual-information type .................................................................................................................. 126
Table 5.7  Results of randomization tests comparing the number of words in contextual-information types of answered and unanswered questions ... 129
Table 5.8  Contingency table of questions not including and including contextual information related to the current and desired states .......................... 132
Table 5.9  Contextual information requested by commenters by t-unit totals by code .......................................................................................... 135
Table 6.1  Contextual-information types by percentage in t-unit location........... 162
ACKNOWLEDGEMENTS

I express my sincere appreciation to my major professor, Dr. Jo Mackiwicz, for her encouraging mentoring, timely feedback, and dedicated service. In addition, I express gratitude for all of my committee members: Dr. David Russell, Dr. Stacy Tye-Williams, Dr. Bethany Gray, and Dr. Zhu Zhang. They provided thoughtful and excellent feedback, and they were wonderful to work with.

I also thank Keegan Kwirant and Arline Votruba for their help as intercoders. In addition to coding data, they provided valuable insight and feedback throughout the code development process.

Finally, I acknowledge the unwavering support, love, and prayers of my family. The optimism and faith of my wife, Marianne, was a continual source of strength. Our enlightening discussions helped me find many answers to my own questions as I completed this study. My children, Eliza, Linden, Daniel, and Hyrum, were extremely understanding and provided needed diversion that helped me see this study in proper perspective.
ABSTRACT

For technical-knowledge workers seeking information about how to complete software tasks, online social question and answer (SQA) sites represent a valuable resource as an emerging form of software documentation. However, because answerers on these sites respond to questions on a volunteer basis, not all questions receive answers. Current research shows that askers provide contextual information in varying amounts, yet researchers have not yet reliably described contextual information types, disagree on whether more or less information associates with answerability, and have not yet compared the coherence of answered and unanswered questions. To assist technical-knowledge workers posting questions on SQA sites, this study explores the relationship between contextual information and answerability and between logical coherence and answerability.

This study analyzes 3,529 contextual-information t-units and 690 comment t-units from social how-to questions about Microsoft Word that askers posted on the popular SQA site Super User. Content analysis enabled a close examination of not only the amounts of contextual information that askers provided, but also the types of information, relationships among types, and relationships between types and answerability. Establishing and using three reliable codebooks related to social how-to questions, to contextual information, and to answerers’ follow-up comments, the study presents descriptive statistics and examples of contextual-information types and comment types. Further analyzing contextual-information types, the study presents and explores statistical differences in the distinctness, magnitude, variation, efficiency, word count, and logical coherence of contextual information in answered and unanswered questions.
A user of Microsoft Word posts the following question online:¹

I am attempting to create a Macro to insert a textbox populated with pre
determined text when a shortcut key is selected. I am able to record a Macro to
generate the text but I am unable to get it to populate the text box with the text

In the preceding question, the asker details the goal: setting a shortcut key to populate a
textbox with specific text. She² provides information about her current situation: what
can be done and what can’t be done. In addition, she provides important information
about the version of the Word application she is using. The asker carefully describes her
situation to help get relevant information. However, the information the asker provides
does not, it seems, meet the needs of those qualified to help. As a result, an answerer
posts the following comment:

Welcome to Super User. You have done a good job of explaining what you are
ty ing to do. Please edit your question to include the code you are currently
working with and details of any research you have done. This will improve the
quality of your question and the ability for others to give you a detailed answer.
(CharlieRB, 2015)

The answerer respectfully acknowledges the information the asker has already provided,
but also requests additional information that will help this answerer and others better

¹ Throughout this study I will report all quotes from users with edits of only some spacing and of removing
terminal punctuation. I will include no sic notations.
² Because the sex of users is unknown, in this study I will alternate between feminine and masculine
pronouns.
understand the asker’s situation. By providing this information, the asker will presumably enable answerers to give responses that are more complete.

This short dialogue exemplifies many interactions online when users turn to publicly available online services (PAOS) to ask for information about how to do something. Users posting how-to questions, whom I will call “askers” (Gazan, 2011, p. 2303), use PAOSs for help with their day-to-day work tasks because they either do not have these services available to them within their companies or, for some reason, choose not to use services made available by their employers (Ferro & Zachry, 2014). While better-known examples of PAOSs include Twitter, LinkedIn, or Google Docs (p. 121), one type that is less widely known is social question and answer (SQA), which Shah, Oh, and Oh (2009) describe as “community-based, and purposefully designed to support people who desire to ask and answer questions, interacting with one another online” (p. 206). Technical-knowledge workers such as the asker in the example above post specific how-to questions on SQA sites such as Yahoo Answers or Stack Overflow, and they receive specific responses tailored to their desired outcomes, operating systems, software versions, or any number of other contextual variables. The how-to questions that knowledge workers post are too specific to be treated in company documentation, but SQA sites facilitate answers to these individual and specific questions. As Spinuzzi (2009) notes, the sites’ users who answer questions, whom I will call “answerers” (Gazan, 2011, p. 2305), “become active writers as they answer each other’s questions about even the most specific and localized cases” (p. 256). This ability of askers and answerers to generate technical content has caused Kimball (2016) to argue that we exist in the “Golden Age” of technical communication, observing that, “at no time in human
history have more people . . . been involved in helping to accommodate each other to technology and to accommodate technology to their own ends” (p. 12). However, precisely because other users answer questions, askers are not guaranteed a reply. For knowledge workers who depend on these answers to complete their professional work, understanding how to write high-quality social how-to questions to encourage answers is critical.

To assist askers, researchers of information retrieval, computer science, and information science have become interested in discovering what characteristics of social questions correlate with answerability, which is the likelihood or actuality of a question receiving an answer. These researchers agree that questions with simpler, less specialized vocabulary result in questions receiving answers (Asaduzzaman et al., 2013; Chua, & Banerjee, 2015). Other researchers have found that questions devoid of socially insensitive taboo words (Choi, Kitzie, & Shah, 2013; Shah, 2012), multiple question marks (Choi, Kitzie, & Shah, 2013; Kitzie, Choi, & Shah, 2013), too little detail (Asaduzzaman, Mashiyat, Roy, & Schneider, 2013; Li, Jin, Lyu, King, & Mak, 2012) and too much detail (Agichtein, Castillo, Donato, Gionis, & Mishne, 2008; Chua & Banerjee, 2015) also result in answers. While such research has provided valuable insight into the effect of individual words, syntax, and length of information expressed in questions, it has not yet addressed the questions’ substantive content, such as contextual information and logical coherence, and its relationship to questions’ answerability. Understanding substantive content could enable researchers to better predict the answerability of social questions, and help askers understand what content to include in questions so answerers can provide help and answers more efficiently.
Because of potential implications for writers of technical documentation, technical and professional communication researchers have long been interested in various forms of online help discourse. Mirel (1994) analyzed user discussions of complex tasks in an early online bulletin board to assist writers of computer software instructions. Building on Mirel’s construct of complex tasks, Swarts (2015) analyzed uncertainty in online help forums, providing practical guidance for technical communicators who create documentation in these forums. Similar to online help forums, SQA sites may have important implications for technical documentation, yet they differ from forums in many important ways.

Structurally, SQA sites distinctly separate questions, answers, and commentary, whereas in forums these content types intersperse throughout the potentially numerous posts and replies of forum participants. Because these content types are discrete on SQA sites, users who choose to rate content for its quality can rate specific and distinct content—questions, answers, and commentary. Along with other Web 2.0 features, these ratings reveal how users perceive the quality of individual questions and answers. This separation also influences user interaction because the discreteness of the content encourages users to pose questions and provide answers but discourages needless commenting. Whereas forums exist as an online space for individuals to comment back and forth, SQA sites exist as an online space dedicated to questioning and to answering. Analysis of questioning in particular can help identify the ways askers communicate questions and how those different ways relate to answerability. Although analysis of SQA holds much promise—and despite Spinuzzi’s (2009) call for the field of technical communication to begin “thinking through” the implications of social software (p. 260)—
SQA has received little attention among professional and technical communication researchers. By analyzing the way askers communicate their how-to questions, specifically in relation to the contextual information they provide and the logical coherence of their questions, I provide insight for SQA researchers seeking to predict how-to questions’ answerability. In addition, I provide specific question-writing strategies for askers seeking answers to how-to questions.

Researchers have categorized social questions into varying types, and this study focuses specifically on how-to questions. This distinction is necessary because different question types influence the content askers include in their questions and the types of replies answerers provide. One influential study broadly distinguished social questions asked “with the intent of stimulating discussion” from questions asked “with the intent of getting information” (Harper, Moy, & Konstan, 2009, p. 1). Comparing the content of these two question types, the researchers found that discussion-seeking questions more frequently included the words “why” and “you,” whereas information-seeking questions more frequently included the words “I,” “where,” and “how” (p. 7). How-to questions represent a subset of information-seeking questions, and researchers have described how-to questions as questions that “ask for instructions” (Truede, Barzilay, & Storey, 2011, p. 806). As askers write how-to questions to request instructions, they prompt answerers to provide procedures to accomplish the task the asker describes.

Not surprisingly, researchers in computer science have begun to view SQA communication as a “substitute for official product documentation,” especially when official documentation is not yet available or not extensive enough (Truede, Barzilay, & Storey, 2011, p. 804). I argue, therefore, that social how-to questions represent an
emerging form of technical documentation. To do so, I build on documentation frameworks established by Selber (2010) and Farkas (1999).

**SQA Communication as Documentation: A Theoretical Framework**

To examine SQA communication using documentation frameworks, I first situate SQA communication among prevailing theories of documentation. But as can be expected, when technical and professional communicators discuss and define documentation, variation exists. For example, even for its name some researchers use the term “documentation” (Mead, 1998, p. 354) or “help documentation” (Swarts, 2015, p. 154), while others use the terms “instruction set” (Selber, 2010, p. 97), “‘how to’ discourse” (Harris, 1983, p. 139), and “procedural discourse” (Farkas, 1999, p. 42). While many of these terms may be considered “types” of the larger genre of procedural discourse (Swarts, 2014, p. 254), the existence of multiple types of procedural discourse leads one to consider their unifying principles.

Despite the varied names researchers have for documentation, their definitions unify around the ends or goals of documentation. Mead (1998) defines documentation as “any textual or graphical material that instructs the user in how to use a product or service” (p. 354). Regardless of form, documentation becomes a means of instruction that leads to an end goal of use of a product or service. Farkas (1999) focuses use on task performance, defining procedural discourse as “written and spoken discourse that guides people in performing a task—in other words, it is ‘how to’ communication” (p. 42). This performance is not solely physical, however, as Harris (1983) defines how-to discourse as “whole pieces of writing that exist to instruct their readers in the performance of some physical or intellectual task” (p. 139). Selber (2010) similarly emphasizes that instruction
sets “provide step-by-step procedures for accomplishing a physical or mental task” (p. 97). While the terminology, mode, or task type of documentation may differ, the similarity among these definitions is that documentation intends to guide a user’s action toward accomplishing some task. Thus, the main purpose of documentation is to guide user action. As askers post how-to questions on SQA sites, they seek direction about how to act to accomplish tasks. As answerers then reply to those questions, the resulting written answers guide the askers’ actions in accomplishing their tasks. Answers written on SQA sites, therefore, share the same action-oriented purpose as documentation, yet the social capabilities of SQA sites also enable reaction and interaction as the following paragraphs illustrate.

**SQA communication as reaction and interaction**

I use Selber’s (2010) framework of emerging forms of online documentation to show how SQA documentation enables reaction and interaction for both askers and answerers. Selber (2010) created a three-part classification scheme that differentiates instruction sets based partially on users’ ability to participate in the documentation process. The first group is action-enabling instruction sets that are “fixed, static, and absolute” (p. 100). Examples include static PDF documents or web pages. Users can follow these types of instructions sets, but they cannot review or update them at the online location. The second group is reaction-enabling instruction sets that include “user-generated metadata” (p. 103). Examples include screencasts that solicit user-feedback and forums where users post content and then reply to each other. Users not only follow these types of reaction-enabling instruction sets but also review them through user votes, comments, and ratings. The final group is interaction-enabling instruction sets that are “open,” for example, wikis
or other forms of knowledge bases (p. 108). Users not only follow and review interaction-enabling instruction sets, but also create and update them through “a steady practice of reinterpretation and revision” (p. 113). Users, therefore, move from passive recipients of action-enabling instruction sets, to raters and feedback providers of reaction-enabling instruction sets, to active collaborators in the documentation process of interaction-enabling instruction sets.

Selber (2010) characterized SQA as instruction sets that enable reaction but not interaction, a characterization that on its face seems accurate in light of the forum-like, reactionary process of back-and-forth questioning and answering. However, viewed holistically, the resulting product of those exchanges truly is documentation generated through the interactions of askers and answerers. Indeed, because many SQA sites, in contrast to forums, discourage reaction-based chitchat, SQA sites encourage askers and answerers to stay focused on the topic, reducing the amount of irrelevant content users must wade through (“Welcome to Super User,” 2017; “Yahoo Answers,” 2016). Steehouder (2002) described such exchanges as “dialogue,” presciently suggesting that “as opposed to ‘documentation,’ dialogue would seem to be a better metaphor for technical communication in the future” (p. 489). The final product of SQA dialogue then is a series of questions and (hopefully) answers, an interactive combination that produces user-generated documentation. And because SQA sites archive and make searchable this dialogue-driven documentation, the text transcripts remain a valuable knowledge resource for users who might be searching for answers to similar questions in the future (Anderson, Huttenlocher, Kelinberg, & Lescovec, 2012).
The result of users’ SQA dialogue is tertiary, user-generated documentation that differs from more traditional primary and secondary forms that organizations or companies generate. Categorized by author, the “creators of the software” traditionally generate primary documentation (Walters & Beck, 1992, p. 156), and while becoming less prevalent, such primary documentation includes printed user manuals created by software providers and shipped with their out-of-the-box software. Those in the “secondary text market” produce secondary documentation (p. 156), for example, the documentation published by Wiley through the well-known *For Dummies* series (Coney & Chatfield, 2006). In contrast, SQA documentation falls into the category of what I call tertiary documentation that refers to software documentation that users produce.  

Examples of tertiary documentation include do-it-yourself tutorials (Van Ittersum, 2014) or videos (Swarts, 2012) and online software forums where users both post requests for help and respond to requests for help (Steehouder, 2002; Swarts, 2015). While most research focusing on user-generated documentation focuses on users posting content in online formats (Mirel, 1994; Selber, 2010; Swarts, 2014, 2015), the online medium does not adequately distinguish this third type since creators of primary and secondary documentation can also post their documentation online, and creators of tertiary documentation can also print their online documentation. As such, the key difference among these three types of documentation is the originating author, whether the documentation originates from the software’s creator, from the secondary text market, or from users. Notably, technical communication researchers are noticing the importance of  

---

3 When I was in a discussion with my dissertation advisor, Jo Mackiewicz, she came up with the term “tertiary documentation” to describe user-generated documentation. However, a Google Search shows that in 2013, a blogger named “Aaron” coined the term “tertiary documentation” in this 2013 blog post: https://sharknet.us/2013/11/29/categorizing-documentation/.
tertiary documentation (Kimball, 2016; Spinuzzi, 2009), but they have not yet sufficiently studied SQA documentation.

**Contextual information and the logical construction of questions**

Social how-to question content in particular plays an important role in the documentation process occurring on SQA sites because the questions themselves are what initiate dialogue with answerers. Steehouder (2002) observed that users have come to expect more personalized information, and thus companies have responded by producing less “generalized information for a large audience” and providing more “specific information to individuals” (p. 490). As a result, he noted that companies’ “new strategy for external information” centers on “answering questions” (p. 490). Because companies push less documentation on consumers, users face a new burden: they must seek out, or pull, the information they need. In an SQA environment, users must communicate their how-to questions effectively to receive an answer. If the information in these questions is incomplete or otherwise lacking, answerers may not respond, leaving users without the solutions they seek.

Research suggests that askers recognize the need to communicate effectively in order to elicit answers. Jeon and Rieh (2015) interviewed users of one SQA site and noted that askers who posted information-seeking questions reported that they strategically crafted questions to elicit high-quality answers, employing strategies such as adding contextual information, defining the scope of the question, and tailoring the question to a specific audience (p. 6). Interviews with answerers suggest that they are pressed for time, so questions that express a real need motivate them more to answer (Nam, Ackerman, & Adamic, 2009). Other studies suggested that questions coded as
taking more time to answer (Asaduzzaman et al., 2013; Shah, Radford, Connaway, Choi, & Kitzie, 2012), as not providing adequate detail or coherence (Asaduzzaman et al., 2013), and being too complex and unclear (Chua & Banerjee, 2015) received no answers. If answerers perceive questions to be too costly in terms of researching or understanding the questions, they may feel less motivation to answer or, due to time constraints, be unable to provide answers. Askers must, therefore, communicate questions in a way that both persuades answerers to reply and provides the information necessary to diminish the time answerers must spend in answering the question. Therefore, a question that includes sufficient contextual information may increase the likelihood of an asker receiving a reply.

Yet SQA researchers are still debating how much contextual information askers should provide to elicit answers. Researchers generally measure the answerability of a question based on whether the question receives an answer, and researchers have found that the amount of detail askers include predicts question answerability (Li et al., 2012; Saha et al., 2013). However, in relation to how much detail askers should include, some research suggests that questions that do not include enough detail result in less answerability (Asaduzzaman et al., 2013), while other research suggests that questions that include too much detail result in less answerability (Agichtein et al., 2008; Chua & Banerjee, 2015). However, exploring a middle-ground approach, Yang et al. (2011) found that social questions with average amounts of detail in one data set received answers proportionally less often than questions with more or less detail than average. These varied results suggest that the amount of detail askers provide does not relate to answerability, yet answerers clearly need some amount of detail in order to answer
questions. Therefore, more research is needed to more closely investigate the information askers should include to increase the answerability.

Yet including the right types of information may still be insufficient for askers to elicit answers. SQA researchers have found that even if askers include contextual information, the information may still be incoherent or difficult for answerers to comprehend (Asaduzzaman et al., 2013; Kitzie et al., 2013). Coherence consists of writers making logical connections in a text (Brostoff, 1981; Van Dijk, 1980) and structuring the text in ways that meet the audience’s expectations (Kintsch and van Dijk, 1978). Because SQA dialogue represents an emerging form of documentation, askers may enhance coherence in their questions by structuring them according to a logical structure that underlies documentation. Harris (1983) theorized the rhetorical logic of documentation by noting that documentation consists of proving an inductive hypothesis that “this is how you do” something (p. 153). Farkas (1999) provided his theory of the logical structure of documentation that consisted of various states and activities. His model suggested that documentation must describe the “prerequisite state” (with which a user aligns his or her “current state”), the “desired state,” and then provide the actions the user must take in order to move from the prerequisite to desired state (pp. 42–43). He notes that “amid all the variations of format and syntax and as much as computer technologies such as wizards change the presentation of procedures, [the] same logical structure is always in place” (p. 42). Even though SQA dialogue represents a new format for documentation, according to Farkas, it should still exhibit a logical structure. Therefore, to achieve such logical structure in SQA dialogue, askers must provide contextual information concerning their current and desired states so answerers can then
provide the actions and procedures necessary to move between states. If askers do not provide information concerning their current and desired states, answerers may see the question as incoherent and be unable to provide answers. While previous studies of coherence in SQA questions exist (Asaduzzaman et al., 2013; Shah et al., 2012), these studies analyze coherence in only unanswered questions. More research is needed to compare coherence in both unanswered and answered questions in order to validate previous findings.

**The Focus of this Study**

This study describes and analyzes the contextual information and logical coherence of askers’ how-to questions and the relationship of those content features with question answerability. Research suggests that askers post both conversation-seeking and information-seeking questions on SQA sites (Harper et al., 2009), and this study will focus on a subset of information-seeking questions that Trude et al. (2011) describe as “how to” questions (p. 806). They define how-to questions as those questions that askers post to “ask for instructions” (p. 806). The researchers found that 13.91% of how-to questions they sampled from the popular SQA site Stack Overflow received no answers, suggesting a need to further research reasons why some how-to questions go unanswered. Previous research provides little agreement as to whether the amount of detail askers provide in social questions influences their answerability (Agichtein et al., 2008; Asaduzzaman et al., 2013; Chua & Banerjee, 2015; Yang et al., 2011), and other research focusing on the relationship between question coherence and answerability has focused only on unanswered questions (Asaduzzaman et al., 2013; Shah et al., 2012). Therefore, more research is needed to investigate whether the types of contextual information askers
provide relate to answerability and to compare the coherence of answered and unanswered questions.

To extend previous research and to specifically examine how-to questions, I used content analysis (Boettger & Palmer, 2005; Budd, Thorp, & Donohew, 1967; Krippendorf, 2013) to analyze 250 answered and 250 unanswered questions from the SQA site Super User, a subsite of the SQA site Stack Exchange. SQA researchers have extensively studied social questions and answers from another subsite of Stack Exchange, Stack Overflow, due to its popularity and usefulness among programmers. Because I am interested in drawing conclusions for technical-knowledge workers more generally, I analyzed the subsite Super User, which is a “question and answer site for computer enthusiasts and power users” (“Welcome to Super User”, 2017). I specifically examined questions that askers or answerers tagged as relating to the word processing application Microsoft Word to examine questions related to a software application used by numerous technical-knowledge workers (Lanier, 2009; Pringle & Williams, 2005).

To conduct the content analysis of how-to questions, I used separate codebooks to identify how-to questions, identify contextual information that askers provided in their questions, and to identify contextual information answerers elicited in their comments. I based my codebook of how-to questions on Truede, Barzilay, and Storey’s (2011) definition of how-to questions. With how-to questions identified, I then analyzed contextual-information types as informed by Steehouder’s (2002) and Suzuki, Nakayama, and Joho’s (2011) categories of contextual information. This analysis enabled me to analyze the difference in the number of contextual-information types included in answered and unanswered questions. I organized my codebook of contextual-information
types around the previously discussed documentation framework proposed by Farkas (1999) that included a current state and desired state. I then investigated whether logically coherent questions that included information about current and desired states tended to receive answers more frequently than questions that excluded that information. Finally, to increase the validity of these findings, I also analyzed answerers’ requests for additional contextual information as found in answerers’ comments posted to unanswered questions.

The Goal and Contribution of this Study

The goal of this study is to provide technical-knowledge workers with insight into how they can write social how-to questions that lead to answers. Increasingly, technical-knowledge workers are turning to PAOSs for help with their day-to-day work tasks (Ferro & Zachry, 2014), so the results of this study are valuable for individuals who seek information by posting how-to questions on SQA sites. Because social technologies also help unlock knowledge stored within companies (Chui et al., 2012), the results of this study may help employees using internal question-and-answer sites to seek information more effectively.

This study contributes to the field of technical and professional communication. User-generated content plays a central role in today’s Web 2.0 world, prompting Selber (2010) to argue that users’ “sharing of expertise” is the “archetypal task of online engagement” (p. 99). Previous researchers have examined the “sharing of expertise” in various online forms such as videos (Swarts, 2012), user forums (Swarts, 2014), and do-it-yourself tutorials (Van Ittersum, 2014), but SQA documentation has received little to no attention by other researchers. This study on SQA discourse fills that gap and extends this important line of research on procedural discourse in online and social contexts.
Research focusing on the contextual information askers provide in social questions is limited, exploratory, and does not include interrater reliability testing (Steehouder, 2002; Suzuki, Nakayama, & Joho, 2011). This study establishes a reliable coding scheme to describe contextual information and determine differences in the number of contextual-information types in answered and unanswered how-to questions. These findings will help knowledge workers using SQA sites make better informed choices about the number of types of contextual information they include when writing social questions.

Research focusing on the relationship between coherence of social questions and answerability examined only unanswered questions (Asaduzzaman et al., 2013; Shah et al., 2012). This study compares coherence between answered and unanswered how-to questions to validate previous findings. Choi et al. (2013) also suggested that providing a structure to social questions is a “fundamental step toward conceptualizing an effective method for seeking and sharing information” (p. 419). To assess the coherence of questions, this study used an existing documentation framework (Farkas, 1999), providing producers of technical documentation with a structure for understanding, creating, or moderating SQA documentation.

**Conclusion and Overview of this Study**

Informed by theories of documentation and additional literature reviewed in chapter 2 of this study, these are my research questions:

**RQ1:** What types of contextual information do askers provide in social how-to questions?
**RQ2:** Do answered and unanswered how-to questions differ significantly in the number of distinct types of contextual information they include?

**RQ3:** Do answered and unanswered how-to questions differ significantly in whether they include contextual information related to desired states and current states?

**RQ4:** What types of contextual information do answerers most frequently elicit through their comments on unanswered how-to questions?

This chapter has introduced the problems this study seeks to resolve, discussed the theoretical framework of the study, and described the scope, goal, contribution, and research questions of the study. Chapter 2 reviews the prior research that is relevant to this study, and chapter 3 describes the methods of the study. Chapter 4 reports results related to RQ1, and chapter 5 reports results related to RQ2, RQ3, and RQ4. Chapter 6 will conclude this study with a discussion of the results.
CHAPTER II: LITERATURE REVIEW

In this chapter, I review literature related to the answerability of social questions. More specifically, I first review prior research that describes characteristics or predictors of unanswered social questions. Second, I specifically analyze the relationship between contextual information and question answerability, thereby situating my first two research questions, which are these:

**RQ1:** What types of contextual information do askers provide in social how-to questions?

**RQ2:** Do answered and unanswered how-to questions differ significantly in the number of distinct types of contextual information they include?

Third, I analyze the relationship between logical coherence and question answerability, thereby situating my third research question:

**RQ3:** Do answered and unanswered how-to questions differ significantly in whether they include contextual information related to desired states and current states?

Fourth, I analyze the relationship between audience expectations and contextual information, thereby situating my fourth research question:

**RQ4:** What types of contextual information do answerers most frequently elicit through their comments on unanswered how-to questions?

**Characteristics and Predictors of Answerability**

While SQA sites themselves provide relatively little direction to askers for how to write effective questions (“How do I ask,” 2016; “Follow,” 2016), information science and information retrieval researchers provide detailed insight as they examine question
characteristics that correlate with answerability. Researchers agree that social questions posted to SQA sites during early evening hours (Chua & Banerjee, 2015; Yang et al., 2011) and on weekends (Chua & Banerjee, 2015) correlate positively with answerability. SQA researchers also agree that the more questions and answers users post, the more likely those users are to receive answers to their questions (Chua, & Banerjee, 2015; Li et al., 2012; Yang, Hauff, Bozzon, & Houben, 2014). Similarly, researchers agree that users’ reputation scores, which generally accrue as they ask and answer quality questions over time, are also correlated with the answerability of questions (Saha, Saha, & Perry, 2013). Activity level differs from time registered on an SQA site, however, since some users may more actively question and answer than others. Registration on a site generally entails a user creating a username and password, and researchers have found mixed results in whether the length of time users have been registered on a site correlates to their questions’ answerability (Agichtein et al., 2008; Chua & Banerjee, 2015). These research studies suggest that askers should strategically time the posting of their questions and actively ask and answer questions to increase their chances of receiving answers.

In addition to timing, SQA researchers have begun to investigate whether the way an asker words a question relates to answerability. Researchers have found that the presence of interrogative words such as “what,” “who,” “where,” “when,” and “how” enable them to predict answerable questions (Choi et al., 2013). Yet posing too many questions within one question post—measured by the number of question marks in the post—results in fewer answers (Choi et al., 2013; Kitzie et al., 2013). Researchers also agree that questions exhibiting clarity—measured by simpler, less specialized
vocabulary—result in questions receiving answers (Asaduzzaman et al., 2013; Chua & Banerjee, 2015), and questions devoid of socially insensitive words correlate with question answerability (Choi et al., 2013; Shah, 2012). However, most researchers agree that overly polite questions, such as including the words “thanks” and “please,” result in lower answerability (Chua & Banerjee, 2015; Yang et al., 2011; Harper, Raban, Rafaeli, & Konstan, 2008). Askers should, therefore, articulate a single, clear question that is socially sensitive, yet not overly polite.

Researchers have also begun investigating whether the type of question askers post influences answerability. In an influential study on social question types, Harper et al. (2009) coded questions broadly as either informational or conversational, with informational questions being those that askers post with the intent of “getting information” and conversational questions being those askers post with the intent of “stimulating discussion” (p. 1). The researchers achieved high agreement in these codes, providing a broad and reliable classification scheme for social questions. Choi et al. (2013) used this coding scheme to specifically analyze predictors of answerability in information-seeking questions. The researchers found that the presence of interrogative words, precise wording, clear wording, the number of question marks, and additional detail contributed significantly to predicting whether information-seeking questions received an answer that the asker accepted. Ignatova, Toprak, Bernhard, and Gurevych (2009) showed that question types can be more granular than just information or conversation seeking, however, as they coded questions according to a nine-code scheme based on Graesser, McMahan, and Johnson’s (1994) earlier scheme. The researchers found that 46.3% of the questions they coded sought answers in relation to who, what,
where, when questions, with another 45.9% of questions seeking definitions, procedures, or comparisons (p. 4). Providing more insight into procedural questions specifically, Truede et al. (2011) sought to discover types of SQA questions that programmers ask and answer. The researchers found that programmers asked how-to questions seeking instructions most frequently out of 11 question types, with 14% of how-to questions receiving no answers (p. 806). In the sample, questions stating that the asker is a novice all received answers, whereas 30% of questions addressing nonfunctional requirements received no answers (p. 806). These studies suggest that a wide variety of question types exist, that question type may influence whether a question receives an answer, and specific types of users post question types with varying levels of frequency.

Research has found little agreement in how much detail askers should include in questions. Researchers have agreed that the amount of detail included—generally measured by the number of words or characters in a question—helps predict question answerability (Li et al., 2012; Saha et al., 2013), with questions that do not include enough detail resulting in less answerability (Asaduzzaman et al., 2013). However, researchers have also found that questions that include too much detail result in less answerability (Agichtein et al., 2008; Chua & Banerjee, 2015). Given that too much detail and too little detail result in less answerability, askers may be led to provide some amount of detail between those two extremes. However, using average number of words as their measure of detail, Yang et al. (2011) found that social questions with average amounts of detail received answers proportionally less than questions with more or less detail than average. Underscoring these contradicting results, Choi et al. (2013) found that question length did not correlate with question answerability in their sample of social
information-seeking questions. However, Shah et al. (2012) showed that presence or absence of additional details in a question does correlate with answerability. These mixed results suggest that question length may not directly influence question answerability, but including no detail at all may deter answerers. More research is needed to closely investigate what content questions should include to increase answerability as opposed to researching only how much content the questions should include.

Providing insight into the content of answerable questions, a number of studies use human coders rather than relying solely on computational methods. Shah et al. (2012) used grounded theory to discover content qualities associated with unanswered questions. Their resulting typology of reasons that questions do not receive answers includes reasons related to clarity, complexity, multiplicity of questions, and lack of information. They noted that unclear questions are difficult for an average person to understand. Overly complex questions required high amounts of specialized information or time to answer. Questions containing multiple questions detailed more than one question or lacked coherence between a question’s subject and the actual question. Questions lacking information did not provide adequate details for answerers to submit an answer.

Asaduzzaman et al.’s (2013) study of characteristics of unanswered questions corroborates those findings. The researchers similarly took a grounded theory approach, finding that unanswered questions are frequently unclear, incoherent, vague, lacking in adequate information, or complex. The researchers also found that many questions received no answers due to askers’ miscategorizing the questions, not recognizing that similar questions had already been posted, posting questions considered unrelated to the SQA site’s focus, or responding inconsiderately to follow-up questions or comments.
from answerers. As opposed to computational searches for information that consider words and Boolean logic, these results from human coders emphasize the social component of social search: human askers communicate with human answerers. Therefore, askers should carefully consider what details to include based on answerers’ needs, how to communicate those details clearly and coherently for answerers, and how to interact considerately with answerers who voluntarily post follow-up comments.

The Relationship Between Contextual Information and Answerability (RQ1 and RQ2)

Askers appear to sense this need to carefully craft their questions. Jeon and Rieh (2015) interviewed users of SQA sites and noted that askers seeking high-quality answers strategically crafted their questions (p. 6). The researchers report that participants were aware that not all questions receive answers; therefore, the askers employed three general strategies. First, askers included contextual information, such as the asker’s background and preferences, familiarity with the topic in question, detailed problem descriptions, and motivations for asking. Second, askers defined the scope of the question, such as including criteria for decisions, specifying subtopics of interest within a broader question topic, or constraining answers by mentioning the sort of information the asker is and is not interested in receiving. Third, askers tailored questions to a specific audience in order to attract a more qualified answerer, using strategies such as including specialized terms. While the study did not relate these communication strategies to a question’s answerability, it clearly suggests that askers include contextual and constraining information strategically in order to elicit quality answers.
A few studies exist that relate contextual information to quality. In their study of revisions of low-quality social questions, Kitzie et al. (2013) found that higher-quality questions tended to include more unique information while concurrently expressing that unique content in fewer words. Focusing specifically on contextual information in one exploratory study, Suzuki et al. (2011) grouped the contextual information users included into five categories: the asker’s goals, the asker’s own thinking, the asker’s circumstances, the asker’s characteristics, and constraints the asker places on potential answers (p. 1261). The researchers recruited 46 participants to compose six questions: five that included one of the five types of contextual information and a sixth that included no contextual information. Of these six questions, participants wrote three with the intent of seeking opinions from answerers and three seeking information from answerers. The researchers found that varying the type of contextual information included did not significantly influence the number of answers the questions received, but they found some support that askers’ perceptions of the resulting answers’ quality did vary significantly. Harper et al. (2008) similarly asked judges to rate the resulting answers of questions that included varying contextual information. They did not correlate the varied contextual information to the likelihood of receiving an answer, but they found that questions indicating users’ prior effort predicted answer quality at a marginally significant level. These studies provide evidence that some types of contextual information influence perceptions of answer quality. Further, while Suzuki et al. (2011) found that including only one type of contextual information in a question did not influence the number of answers a question received, their research provides a valuable
typology of contextual information that future researchers could use to explore instances when askers provide multiple types of contextual information in their questions.

Other studies provide valuable insight into the types of information users provide when seeking information or help from other users online. Researchers have found that askers include examples of their programming code for programming-related questions, with mixed outcomes in terms of question answerability (Chua & Banerjee, 2015; Saha et al., 2013). Other studies suggest that users frequently include information about their level of experience with the item in question (Mirel, 1994; Treude et al., 2011). More comprehensively and specifically focused on contextual information, Steehouder (2002) analyzed introductory messages on software help forums and found that askers included contextual information in the form of descriptions of their problems, information about hardware and software, the asker’s goal, solutions the asker tried that did not work, potential causes, error messages, and other sources the asker consulted (Steehouder, 2002). Table 2.1 provides a synthesis of Steehouder’s (2002) contextual-information types in relation to the five groups of contextual information that Suzuki et al. (2011) articulated.

**Table 2.1. Contextual-information types in forum and social questions**

<table>
<thead>
<tr>
<th>Steehouder’s (2002) contextual information in forums</th>
<th>Suzuki, Nakama, &amp; Joho’s (2011) contextual factors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scenario</strong></td>
<td><strong>Situation</strong></td>
</tr>
<tr>
<td>A scenario of ‘what happened’ describing the situation.</td>
<td>Circumstances, environment, experience, knowledge, and familiarity.</td>
</tr>
<tr>
<td><strong>Hardware and software specifications</strong></td>
<td><strong>Task</strong></td>
</tr>
<tr>
<td>Hardware and software specifications.</td>
<td>Reasons, motivations, aims, goals.</td>
</tr>
<tr>
<td><strong>Goals</strong></td>
<td></td>
</tr>
<tr>
<td>Goals that the user wants to achieve.</td>
<td></td>
</tr>
<tr>
<td><strong>Attempts that failed</strong></td>
<td></td>
</tr>
<tr>
<td>Attempts made by the help-seeker to solve the problem.</td>
<td></td>
</tr>
<tr>
<td><strong>Suggestions about the cause</strong></td>
<td><strong>Thought</strong></td>
</tr>
<tr>
<td>A suggestion of what may have caused the problem.</td>
<td>Own thought, answer prediction.</td>
</tr>
</tbody>
</table>
Table 2.1. Continued

<table>
<thead>
<tr>
<th>Error message</th>
<th>The error message that might help to understand what the problem is.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sources already consulted</td>
<td>Sources that were already consulted by the help-seeker.</td>
</tr>
<tr>
<td>Attribute</td>
<td>Personal attribute, social status.</td>
</tr>
<tr>
<td>Limit</td>
<td>Conditions on answers and answerers.</td>
</tr>
</tbody>
</table>

As table 2.1 shows, Steehouder (2002) and Suzuki et al. (2011) agree most closely in three areas: situation, goals, and the asker’s own thought. First, Steehouder (2002) observed that askers frequently opened messages in help forums by describing “what happened” (p. 494). Suzuki et al. (2011) found that questions in SQA sites not only described what the asker experienced, but also the circumstances and environment of the situation. The environment and circumstances might include the hardware and software specifications as well as the error messages that Steehouder (2002) coded separately from his situation code. Suzuki et al. (2011) also found that askers included information about their own knowledge and familiarity with the subject of the question, reflecting one of the question-formulation strategies that Jeon and Rieh (2015) reported based on interviews with SQA users. Second, both Steehouder (2002) and Suzuki et al. (2011) observed that askers provided detail about their goals. Suzuki et al. (2011) also found that askers included information about their motivation, also corroborated by Jeon and Rieh (2015) through interviews with askers. Writers may feel a need to include their motivation because, as Nam et al. (2009) found through interviews with SQA answerers, answerers are motivated toward answering questions when they perceive the existence of a real need. Third, Steehouder (2002) and Suzuki et al. (2011) report users speculating on either
the cause or a possible solution, with the speculation reflecting the user’s own thought. Beyond these three areas of agreement, Suzuki et al. (2011) included groupings related to askers’ personal attributes and to constraints the answerers placed on answers; through interviews with askers, Jeon and Rieh (2015) reported both of these codes as question-formulation strategies. This comparison of Steehouder’s (2002) and Suzuki et al.’s (2011) findings suggests that askers frequently include contextual information related to their present circumstances, their desired circumstances, and their thoughts about what caused their present circumstances or what may help them reach their desired circumstances. Askers may also include personal information or information that limits or narrows the types of information they are seeking.

While these studies provide valuable insight into the contextual information that askers provide, they do not yet sufficiently relate that contextual information to the answerability of SQA questions. Steehouder (2002) conducted a valuable preliminary study of 50 opening forum posts; however, his study was limited in size and focused on forums instead of SQA sites, and did not relate the resulting codes to answerability. Steehouder called for further research to explore the “reactions on opening messages,” including the quality of answers that answerers provide (p. 497). Suzuki et al. (2011) focused specifically on SQA sites, yet their study examined questions providing only one type of contextual information. Additionally other studies have closely examined the relationship between answerability, word counts, and the uniqueness of words in questions, yet they have not moved beyond word-level content and answerability (Kitzie et al., 2013; Choi et al., 2013). Because askers provide a wide variety of contextual information as sentences, more research is needed to determine whether combinations of
contextual information at the sentence level influence answerability. Further, while Suzuki et al.’s (2011) and Steehouder’s (2002) groupings were informed by their literature reviews and the their analyses of existing data, the researchers did not perform any interrater reliability testing of their coding schemes. As a result, the reliability of their grouping codes is uncertain. Because of the need to reliably establish a coding scheme for contextual information and to relate multiple instances of sentence-level contextual information to answerability, this study’s first two research questions are as follows:

**RQ1**: What types of contextual information do askers provide in social how-to questions?

**RQ2**: Do answered and unanswered how-to questions differ significantly in the number of distinct types of contextual information they include?

**The Relationship Between Logical Coherence and Answerability (RQ 3)**

Suzuki et al.’s (2011) research suggests that including a single type of contextual information such as circumstances and motivations does not increase the answerability of a question, yet including multiple types of contextual information alone does not necessarily increase answerability either: the included contextual information must be coherent for answerers. To develop a typology of reasons why some programming-related questions go unanswered, Asaduzzaman et al. (2013) examined a random sample of 400 unanswered questions. They found that while some unanswered questions failed to provide samples of code that would have helped answerers better understand the questions, other unanswered questions provided sample code that that was “hard to follow” for answerers (p. 98). Shah et al. (2012) examined 200 unanswered questions and explored reasons why they were not answered. Of these 200 questions, 14% lacked
information that kept coders from understanding what the asker needed, but an additional 10.5% were ambiguous due to a lack of coherence or a clear statement of the asker’s need. In addition, 2% included too much information, with the researchers noting that one lengthy example included a “random assortment of facts” that could be “cognitively overwhelming for a reader to process” (p. 6). Similarly, Kitzie et al. (2013) found that an increased number of words in social questions significantly predicted coders’ perceptions that the questions either lacked information or were ambiguous. They observe that while including more information enables an asker to appropriately detail her needs, more information may also decrease answerers’ ability to understand the questions. These studies suggest that a lack of contextual information may decrease a question’s answerability, but simply including contextual information is not enough: contextual information must also be coherent to potential answerers. Asaduzzaman et al.’s (2013) and Shah et al.’s (2012) studies show that some unanswered questions come across as incoherent to coders, but due to these studies’ sampling of only unanswered questions, they do not provide an adequate comparison of coherence between unanswered and answered questions. Such a comparison would validate findings that suggest that a lack of coherence decreases a question’s answerability.

Coherence comprises textual macrostructures, textual microstructures, contextual surroundings, and reader comprehension. Kintsch and van Dijk (1978) refer to macrostructures as a “topic of discourse” to which text must relate meaningfully (p. 366). The topic represents the main idea of all microstructures, serving as a guide and “global constraint” on the text (p. 366). Among other things, microstructures include cohesion devices (Halliday & Hasan, 1976) and given-new strategies (Kopple, 1983). Brostoff
(1981) has argued that incoherence consists of textual ideas that are close in proximity but are otherwise not connected. She argues that the connections that make coherent text are “logical relationships” that reflect textual patterns readers recognize, such as classification, analogy, and comparison (p. 279). Similarly, Kintsch and van Dijk (1978) refer to microstructures as a “text base” that is ordered in a semantically meaningful way (p. 365). Van Dijk (1980) argued that one way that sentences become meaningful is when they represent “imaginable fact” (p. 32, emphasis in original), and he defined facts as “an event, action, state, or process in some possible world” (p. 32). These facts become meaningful only when they represent likely actions and states and their “condition” upon one another (p. 33). In other words, he argued that because sequences of sentences represent actions and states, they become coherent only when the relationship between action and state makes sense: a state or action must facilitate the other (pp. 33–34). To make social questions coherent then, askers could ensure that the relationship is clear between any states and actions they communicate.

While coherence forms through these various relationships within the text, Witte and Faigley (1981) observe that coherence is also achieved through contextual relationships, relationships to elements outside the text. They argue that even though microstructures may be connected, they may still be incoherent due to audience knowledge and expectations (p. 200). Kintsch and van Dijk (1978) acknowledge these concerns by observing that some microstructures are implicit and are thus excluded from the text base due to an audience’s “general or contextual knowledge of the facts” (p. 365). In their experimental study of writing feedback and coherence, Traxler and Gernsbacher (1995) observed that coherence “enables the reader of listener to build a
mental representation of what the writer or speaker intended to convey” (p. 215).

Consequently, they tested whether readers could correctly choose visuals described by writers’ descriptions; they observed that if writers received feedback concerning the success of their readers between two different writing sessions, the writers apparently altered their writing in the second session so that, in comparison to readers whose writers’ received no feedback, readers were better able to correctly choose the described visuals. Therefore, to make included contextual information coherent and to possibly increase the answerability of their questions, askers could increase their awareness of answerers’ knowledge and needs through feedback.

Yet outside of studies that have found that lack of coherence is a characteristic of some unanswered questions (Asaduzzaman et al., 2013; Shah et al., 2012), SQA researchers have not yet explored coherence and its relationship to answerability. Using van Dijk’s (1980) description of coherence that is based on relationships between actions and states, this study adds to existing SQA research by analyzing questions for coherence, correlating the presence of the coherence to answerability, and then triangulating findings with answerers’ feedback as observed through their comments.

Because SQA discourse represents an emerging form of documentation (Treude et al., 2011), studies related to the logical structure of documentation can provide insight into what makes social questions coherent. Harris (1983) noted that documentation consists of proving an inductive hypothesis that “this is how you do” something (p. 153). Farkas (1999) further argued that a logical structure underlies all documentation, observing that “amid all the variations of format and syntax and as much as computer technologies such as wizards change the presentation of procedures, [the] same logical
structure is always in place” (p. 42). SQA represents a new format for documentation, but according to Farkas, questions should still exhibit his proposed logical structure.

Farkas’s proposed structure consists of five types of states with three types of action that cause changes in states.

1. The first state is the “desired state” or desired outcome for the user.
2. While not emphasized, the second state is the “current state” of the user, or the state the user must move from in order to “align” with the following “prerequisite state.”
3. As mentioned, the third state is the “prerequisite state,” or the state where the user must be located prior to moving on to subsequent states.
4. The fourth states are “interim states” that the user moves through in order to reach the eventual outcome.
5. The fifth states are any “unwanted states” that the user should avoid (pp. 42–43).

The three actions include those (a) of users, (b) of the system the documentation describes, and (c) of any actions outside the system that may influence or act upon the system (pp. 42–43). Notably, van Dijk (1980) argued that likely relationships between states and actions represented coherent text. Therefore, logically constructed and logically coherent documentation clearly communicates where the user should begin, where the user will end up, and what steps the user must take to get there.

Because “procedures exist in a social context,” accurately portraying these prerequisite and desired states brings certain challenges, however (Farkas, 1999, p. 43). This context requires answerers to adapt documentation to asker’s needs, backgrounds, and understanding; to establish credibility; to persuade askers that the benefit of
following the instructions is worth the effort; and to provide information about the
conditional nature of procedural steps without causing confusion for the askers (Farkas,
1999, pp. 43, 48). Indeed, a purely logical model certainly seems appealing as a way to
unify all documentation, yet it can fall short in addressing the realities of practice. As
such, Mirel (1998) questions whether the task-based view on which Farkas (1999) bases
his logic reflects the complex situations in which users find themselves. Rather than
assuming that users recognize the logic of a situation, Mirel (1998) suggests that
“pragmatically, people know the *conditions* under which certain meanings of tasks and
ordering of methods occur” (p. 14, emphasis in original). She continues by noting that
these conditions are what truly matter for users, more so than “universal and standardized
procedures (task syntax) or logical concepts and facts (task semantics)” (p. 14).
Consequently, Swarts (2015) adapts Farkas’s (1999) model of prerequisite, interim, and
desired states into a model that accounts for both the complexity and uncertainty of user
help documentation.

Swarts’s (2015) model incorporates the complexity Mirel (1998) describes by
adding both social and technological constraints to the three states that Farkas (1999)
describes. In this model, the desired state is still clear, yet other stakeholders and
technologies in the users’ networks may influence the way that askers define prerequisite
and interim states. In essence, the complexity model accentuates the influence of what
Farkas (1999) might describe as “actions from outside the system” (p. 43), yet the
complexity is still “manageable” as authors of documentation situate the documentation
task among the given social and technological constraints (Swarts, 2015, p. 167). Driskill
(1989) describes how such contextual constraints “converge to define the situations in
which workers participate” (p. 130). This context lends meaning and knowledge to both writers and readers within those situations (p. 129). Notably, as seen in table 2.2, the contextual-information types that Steehouder (2002) and Suzuki et al. (2011) describe clearly relate to the states in Farkas’s and Swarts’s logical models.

**Table 2.2.** The states and actions of logical documentation aligned with observed contextual information in forums and social questions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired state</td>
<td>The goal that is presented to the user.</td>
<td>Goals</td>
</tr>
<tr>
<td></td>
<td>Goals that the user wants to achieve.</td>
<td>Scenario</td>
</tr>
<tr>
<td></td>
<td>A scenario of ‘what happened’ describing the situation.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Suggestions about the cause</td>
<td>A suggestion of what may have caused the problem.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prerequisite states</td>
<td>The state that is a condition for moving toward the desired state. This is often stated at the beginning of a procedure so that the user can align his or her current state with the prerequisite state.</td>
<td></td>
</tr>
<tr>
<td>Interim states</td>
<td>States we enter as we move toward our goal. These are our milestones or subgoals. We create or reach these states through our actions.</td>
<td></td>
</tr>
</tbody>
</table>
The four studies shown in table 2.2 align most closely in five areas.

- First, Farkas’s observation that logically constructed documentation presents a goal to the user is mirrored by Steehouder’s and Suzuki et al.’s observations that SQA askers detail the goals they are trying to achieve.
- Second, while Farkas mentions a current state only as a precursor to the prerequisite state, Steehouder and Suzuki et al. suggest that askers provide multiple types of contextual information related to their current state.
• Third, Swarts suggests that documentation users face constraints, and Steehouder’s and Suzuki et al.’s contextual-information types also suggest that askers provide contextual information that constrains answers and answerers.

• Fourth, Farkas notes that users take action to achieve goals, and askers’ contextual information reflects actions they take to consult sources and to solve their problems.

• And fifth, Farkas notes that systems also act, and Steehouder’s observations suggest that SQA askers provide contextual information related to their systems’ actions.

Yet as shown in table 2.2, Steehouder and Suzuki et al. report no contextual-information types that appear to relate to Farkas’s prerequisite states, interim states, unwanted states, and actions from external events. The absence of contextual-information types in these areas accentuates that social question and answer consists of both questions and answers. When askers post questions, they provide only part of the exchange, with answerers providing the other part. Presumably, askers provide information related to the desired state, the current state, constraints, and some human actions and system actions, and answerers then provide information related to the prerequisite states, interim states, and unwanted states. As both parties in the exchange provide the necessary information, the resulting question-answer exchange becomes logically constructed and logically coherent documentation. If an asker fails to provide a goal or current state, then the question’s answerer is unable to provide the steps (i.e., actions) necessary to move from the current state to the goal. Therefore, building on the first two research questions and the preceding studies, this study’s third research question is as follows:
**RQ3:** Do answered and unanswered how-to questions differ significantly in whether they include contextual information related to desired states and current states?

**The Relationship Between Audience Expectations and Contextual Information (RQ4)**

The logical structure of questions enhances the coherence of text, but as previously discussed, the expectations and knowledge of the audience also play an important role in how coherent the audience perceives a question. This perception may influence an answerer’s decision to answer or not answer a question. But outside of studies that report coders finding lack of coherence as a characteristic of some unanswered questions (Asaduzzaman et al., 2013; Shah et al., 2012), little research exists that provides insight into the relationship between audience expectations and knowledge, question coherence, and answerability. One possible way to gain insight into answerers’ expectations and knowledge is to examine the feedback comments and follow-up questions that answerers post in response to askers’ questions.

Only a few studies focus on the comments associated with social questions, but these studies provide some insight into the expectations of audiences as well as the relationship between those expectations to answerability. Yang et al. (2014) examined comments that appeared to trigger the editing of a question. For example, if an answerer asked about a programmer’s source code, and then the asker updated the question with code, then the researchers considered this comment an “important edit” to the question (p. 181). To initially understand edit types, the researchers examined other important
question edits that occurred immediately before an answer was posted to a question.

Based on 600 randomly selected questions, edits fell into these categories:

- Changes to source code
- Contextual information to clarify the asker’s goal or other information to broaden answerers’ understanding of the question
- Details about hardware and software
- Examples of inputs or of what result the asker hopes to receive as output
- Problem-statement clarification that provides an error message or other error logs
- Attempt information about what the user has already tried
- Solution information if the asker has found the right answer on his own
- Formatting changes to spelling or code syntax.

Based on random sample of comment-initiated or answer-initiating edits, coders then annotated a sample of edits. Due to difficulties they experienced in distinguishing between edit types and the small number of instances of some categories, the researchers reduced their coding categories to these: source code refinement; one category comprised of problem statement, example, and context; another merged category of solution and attempt; and a final category of detail. They later omitted the detail category because these occurred infrequently. The researchers then used these coded edits to train a classifier to predict the quality of questions based on whether a question received answers and edits. The coded edits of answerers’ comments provide important insight into the expectations and knowledge needs of answerers: answerers expect to see information related to the asker’s current situation, goal, and previous attempts to resolve the problem or answer the question.
One other study provides insight into the relationship between audience expectations, comments, and answerability. Ahn, Butler, Weng, and Webster (2013) investigated whether social feedback on SQA sites influenced an asker’s future question quality, as measured by the SQA community’s upvoting and downvoting of the asker’s question. While Ahn et al.’s study focused on votes rather than answerability, Yang et al. (2014) also found a high correlation between votes and the number of answers a question receives. Ahn et al. (2013) hypothesized that the community’s commenting on questions would provide important feedback to question askers that could improve the asker’s future questions. Notably the researchers found that the number of follow-up comments posted on questions did not significantly predict future question quality. However, because the study investigated the number of comments and not the content of these comments, the researchers suggested that “future work might consider content analysis” in order to more definitively establish whether these comments predict future question quality (p. 8). This study suggests that the raw number of comments may not influence the answerability of questions, but the researchers suggest that the content of these questions may include valuable information related to audience expectations of what information answerable questions should include. Therefore, to better understand audience expectations regarding the information askers should include in questions, this study’s fourth research question is as follows:

**RQ4:** What types of contextual information do answerers most frequently elicit through their comments on unanswered how-to questions?

This chapter discussed the literature that informs the research questions of this study. The next chapter will present the methods that I followed to answer these questions.
CHAPTER III: METHODS

In the previous chapter, I reviewed literature that underscored the need for more research to understand the relationship between contextual information and answerability of social how-to questions. Because previous research on SQA contextual information and answerability did not include interrater reliability testing of coding schemes (Suzuki et al., 2011), this study contributes significantly to SQA research by establishing a coding scheme that reliably describes and categorizes the contextual information askers provided in their questions.

Previous research also disagreed on the amount of contextual information askers should provide to increase question answerability (Agichtein et al., 2008; Asaduzzaman et al., 2013; Choi et al., 2013; Chua & Banerjee, 2015; Li et al., 2012; Saha et al., 2013; Shah, 2012; Yang, Bao, & Lin, 2011). This disagreement suggests the need for more research into differences between the types of content askers include in both unanswered and answered questions. This study filled that need by using a reliable coding scheme to compare the difference in the number of types of contextual information askers provided in answered and unanswered questions.

Results from previous studies suggest that some content of unanswered questions lacks coherence (Asaduzzaman et al., 2013; Shah et al., 2012), but these studies analyzed only unanswered questions and did not compare coherence in unanswered questions to coherence in answered questions. Because a comparison of both unanswered and answered questions would enable more valid conclusions about differences in coherence between unanswered and answered questions, this study analyzed coherence in a sample of both unanswered and answered questions. However, coherence comprises of both
textual relationships and audience expectations, so this study, extending previous research, analyzed the follow-up comments of answerers to triangulate audience expectations with textual relationships.

This chapter presents the methods of the study and proceeds as follows. I first describe the research site and study participants, provide an overview of the data and data collection procedures, review the research methodology, and report on codebook development.

**Research Site and Participants**

The website where the data originated is called Stack Exchange and is accessible on the internet at www.stackexchange.com. According to its “about us” page, Stack Exchange consists of over 150 SQA subsites with topics ranging from computer programming to gardening to the English language. For 2015, Stack Exchange reported 101 million unique visitors per month to its collection of subsites, with 5 million total registered users. In 2015, these users represented askers who posted 3.7 million questions and answerers who responded with 4.6 million answers. Askers and answerers also posted 17.9 million comments during that same period (“About,” 2016).

The specific subsite of Stack Exchange analyzed in this study is Super User, a “question and answer site for computer enthusiasts and power users” (Super User, 2016). A power user is defined as a “computer user who uses advanced features of computer hardware, operating systems, programs, or websites” (“Power user,” 2017). Thus most users of the site likely have considerable experience with the software and hardware they post about. The majority of SQA studies examine data from Yahoo Answers or Stack Overflow, which is another subsite of Stack Exchange intended for “professional and
enthusiast programmers” (“Welcome to Stack Overflow,” 2017); however, I examined Super User because I am interested in drawing conclusions for technical-knowledge workers generally rather than for only computer programmers. Whereas question tags on Stack Overflow focus on specific programming languages, functions, and syntax, those on Super User focus mainly on common hardware and software that numerous technical-knowledge workers might use on the job.

While I did not collect data directly from human subjects, askers and answerers of the Super User subsite generated the archived question and comment data that I analyzed in this study. In an effort to respect “all possible research participants as autonomous individuals who have the capability of making decisions about their participation in a research project” (Wrench, Thomas-Maddox, Richmond, & McCroskey, 2013, p. 30), I met with Dr. Kerry Agnitsch, Co-Chair of the Iowa State University Institutional Review Board (IRB), on September 24, 2015, to investigate whether I would need IRB approval to analyze the archived data from Stack Exchange. Because the data I analyzed is not private, she stated that I most likely would not need IRB approval. However, she recommended that I examine the user agreement on the site I pull data from, and if any expectation is set that the data will be private, then I would need to consult with the IRB office further.

I reviewed the privacy policy posted on the Stack Exchange site, and I found that the policy explicitly states that identifying information that site users provide is not private. Here is an excerpt from the privacy policy that states that if users voluntarily provide personal information in their communication on the site, they lose privacy rights:
Some users may elect to publicly post personally identifying or sensitive information about themselves in their normal use of the network. This could occur through use of the optional profile fields, in question or answer posts, or when an individual posts a job history on the Careers site. Information like that, which is voluntarily posted in publicly visible parts of the network, is considered to be public, even if it would otherwise be considered to be personally identifying or sensitive. As such, it is not subject to the protocols listed below, because we don’t control it; you do. Additionally, voluntarily publicizing such information means that you lose any privacy rights you might normally have with regards to that information. It also increases your chances of receiving unwanted communications, like spam. (“Stack Exchange, Inc.,” 2016)

On September 19, 2016, I emailed the preceding excerpt along with a link to the full privacy policy to Dr. Agnitsch and asked her whether she thought I should seek IRB approval for this study. I also mentioned that if any data happened to include identifying data then I would not publish that data. She replied that if my study involves publicly available data from stackexchange.com, if users are informed by stackexchange.com that identifying data they provide is public, and if I do not publish any identifying data that users happened to provide, then she agreed that I did not need IRB approval. Specifically, she stated that my project did not involve “human subjects (as federally defined)” because I would not be “interacting or intervening with participants to gather data” and because the data I used “is not private” (K. Agnitsch, personal communication, September 23, 2016). Acting in accordance with the guidelines set forth in her reply email, I proceeded to collect the data for this study.
Data and Data Collection Procedures

Stack Exchange updates a downloadable archived copy of public data from its database on the Stack Exchange data explorer each Sunday evening around 3:00 UTC (https://data.stackexchange.com/superuser/queries). The data explorer consists of a text box where anyone can submit standard query language (SQL) commands to view and download the data in the archive. I downloaded the data for this study using the SQL code found in appendix A.

Because I designed this study to draw conclusions specifically for technical-knowledge workers (Ferro & Zachry, 2014; Johnson-Eilola, 1996; Wick, 2000), I downloaded questions that askers or answerers tagged as relating to Microsoft Word. Microsoft Word is a word-processing software application that most technical-knowledge workers use (Lanier, 2009; Pringle & Williams, 2005). The Super User site allows any user to create tags, but it recommends that askers rely on previously created tags (“How to Tag,” 2017). I analyzed all available tags related to Microsoft Word by searching all tags for the word “word,” and I determined that most askers used one of nine different tags: microsoft-word, microsoft-word-2000, microsoft-word-2003, microsoft-word-2007, microsoft-word-2008, microsoft-word-2010, microsoft-word-2011, microsoft-word-2013, microsoft-word-2016 (“Tags,” 2017). To capture questions tagged with any of these tags, I included questions that included the search terms “microsoft-word,” as appendix A shows. Table 3.1 shows the nine tags, the number of questions on the Super User site to which users appended the tags, and the number of questions in the final data set to which users had appended the tags. Overall, the distribution of tags in the present study reflects the distribution of tags on the Super User site.
Table 3.1. Tag frequency on Super User and in the present study’s data set

<table>
<thead>
<tr>
<th>Tag</th>
<th>Questions tag appended to on site</th>
<th>% of total</th>
<th>Questions tag appended to in data set</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>microsoft-word</td>
<td>4,725</td>
<td>65.44</td>
<td>406</td>
<td>64.75</td>
</tr>
<tr>
<td>microsoft-word-2000</td>
<td>5</td>
<td>0.07</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>microsoft-word-2003</td>
<td>112</td>
<td>1.55</td>
<td>6</td>
<td>0.96</td>
</tr>
<tr>
<td>microsoft-word-2007</td>
<td>623</td>
<td>8.63</td>
<td>31</td>
<td>4.94</td>
</tr>
<tr>
<td>microsoft-word-2008</td>
<td>21</td>
<td>0.29</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>microsoft-word-2010</td>
<td>976</td>
<td>13.52</td>
<td>79</td>
<td>12.60</td>
</tr>
<tr>
<td>microsoft-word-2011</td>
<td>110</td>
<td>1.52</td>
<td>8</td>
<td>1.28</td>
</tr>
<tr>
<td>microsoft-word-2013</td>
<td>528</td>
<td>7.31</td>
<td>91</td>
<td>14.51</td>
</tr>
<tr>
<td>microsoft-word-2016</td>
<td>120</td>
<td>1.66</td>
<td>6</td>
<td>0.96</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>4,725</strong></td>
<td><strong>65.44</strong></td>
<td><strong>627</strong></td>
<td><strong>64.75</strong></td>
</tr>
</tbody>
</table>

*Note:* Because users can append multiple tags to questions, the total number of questions does not reflect the actual number of questions I analyzed in this study. The total number of questions I analyzed in the study is 500.

In September 2016, I downloaded data of questions posted between January 1, 2009, and June 30, 2016, for two pilot studies. The total number of questions tagged as relating to Microsoft Word during this time totaled 5,044. In January 2017, I downloaded data of questions posted between April 1, 2013, and September 30, 2016, for the final study. The total number of questions tagged as relating to Microsoft Word during this time totaled 3,128 questions. Here are the key fields I pulled from the database to capture question and comment information:

- Question ID
- Question Title
- Question Text
- Question Tags
- Question Date
- Question Edits
- Question Answers
- Question Comments
I then separated my pilot study data into two time periods, enabling me to use one subset of the data for the first pilot study and the remaining data for the second pilot study. Before further sampling the question data, the data for the first pilot study consisted of 2,108 (41.79%) of the 5,044 originally downloaded questions that askers posted from January 1, 2009, through March 31, 2013, and the data for the second pilot study data consisted of 2,936 (58.21%) of the 5,044 originally downloaded questions askers posted from April 1, 2013, through June 30, 2016.

Because I learned more about the Super User site and the nature of the questions throughout the pilot study, I removed some questions from the final study data that I had included during the pilot study. Specifically, I had learned that moderators can close questions. When questions are closed, answerers can no longer post answers. Moderators close questions for various reasons: the question does not relate to hardware or software; the question seeks opinions; the question has numerous answers; the question would require a long answer; the question relates to hardware that does not plug in to computers; a question relates to a website or a web service; a question asks about purchase recommendations; or a question seeks support for business-related information technology (“Welcome to Super User,” 2017). Because this study focused on analyzing the relationship between only contextual information and answerability, I decided to control for moderators’ influence by eliminating 97 closed questions from the final study’s data. After I eliminated these questions, 3,031 questions remained in the final data set.

From both the pilot- and final-study data, I created two additional groups of data within each data set:
• **Answered questions:** Questions that received answers but no comments or edits. I did not include questions with comments because I wanted to analyze questions that provided enough contextual information that answerers could provide an answer without seeking additional information from the asker. Super User also exhibits wiki features that enable any askers or answerers with a certain level of experience to edit anyone’s questions or answers (“Edit Questions,” 2016). Because I wanted to examine the communication of only the original asker with no one else’s edits, I did not include questions that had been edited. Of the original 2,108 questions in the first pilot study, 543 (25.76%) questions had been answered with no edits or comments. Of the original 2,936 questions in the second pilot study, 661 (22.51%) fell into this category. Of the original 3,031 questions for the final study, 648 (21.38%) fell into this category.

• **Unanswered questions:** Questions that did not receive answers or edits but could have received comments. I similarly restricted edits for this group to ensure I was analyzing only the communication of the original asker; however, to answer RQ4 related to the types of contextual information answerers request, I included unanswered questions in this group that had received comments from answerers. Of the original 2,108 questions in the first pilot study, 54 (2.56%) questions fell into this category; of the 2,936 original questions in the second pilot study, 504 (17.17%) fell into this category. Of the original 3,031 questions for the final study, 493 (16.27%) fell into this category.

Again because I had learned more about the question data throughout the pilot study, I further refined the questions in the final study. I had discovered that some
answerers posted comments on unanswered questions that could deter other answerers from answering. For example, answerers sometimes posted comments stating that another asker had already posted a similar question, thus suggesting the redundancy of the question and referring the asker to the answer of the previously posted question. In addition, answerers sometimes posted comments suggesting that what the asker wanted was not possible; with such definitive statements, other answerers could have passed by these questions without attempting to provide answers. Therefore, I eliminated five questions with the statement “possible duplicate” and nine questions with the statement “isn’t possible” in the comments. I also ensured that no other askers had contested the comments before eliminating the unanswered questions from the data set. The new total of unanswered questions in the final study’s data set was 480 (15.84% of the original 3,031 questions).

I also explored answerability and the length of time a question existed on the site. As I have already noted, the data in the final study between the years of 2013 and 2016, whereas the data in the first pilot study originated between 2009 and 2013. Notably, the percentage of answered questions fell from 25.76% in the data of the first pilot study data to 22.51% in the data of the second pilot study and to 21.38% in the final study. To explore the likelihood that the question’s time on the site played a role in answerability and to ensure that more recently posted questions in the final data set would not necessarily have a lower probability of receiving an answer than questions posted less recently, I examined the answer date of all questions in the original data download of the final study. Of the 3,128 questions in the download, 2,214 (70.78%) had an answer date. For all questions with an answer date, the average time that elapsed between the
question’s initial posting and receiving an answer was 20 days, with the median time being 121 minutes. Further, 2,096 (94.67%) had received answers within 90 days of their initial posting date. Askers had posted questions in the final data set as late as September 30, 2016, and I downloaded the final on January 19, 2017. Because over three and a half months had passed between the two dates, I determined that the majority of answerable questions posted in the year 2016 would have already received answers by the time I downloaded the final data set.

To ensure that I had a sample size large enough to draw meaningful conclusions and to follow sample size precedents set in other SQA studies (Agichtein et al., 2008; Treude et al., 2011), the sample of the final study comprised 250 randomly selected answered and 250 randomly selected unanswered how-to questions. Aiming for my pilot-study sample sizes to be 10% of the final-study sample (Boettger & Palmer, 2010), I included 25 randomly selected answered and 25 randomly selected unanswered questions in the first and second pilot-study samples. Later in this chapter when I discuss code development, I provide additional details about my sampling process.

In this study, as well as its pilots, I divided question text into my unit of analysis: t-units. A t-unit is an independent clause with any related dependent clauses, described by Hunt (1965) as the “shortest grammatically allowable sentences” (p. 21). For example, returning to the example question presented in chapter 1, the numbered brackets designate t-units:

\(<t1>\)I am attempting to create a Macro to insert a textbox populated with pre
determined text when a shortcut key is selected.\(<t1>\) \(<t2>\)I am able to record a
Macro to generate the text\(<t2>\) \(<t3>\)but I am unable to get it to populate the text
box with the text inside. Using Microsoft word 2010. Any suggestions please?

Although I indicated t-units in this example and in other examples throughout this dissertation, when unitizing data I used different fonts colors to distinguish t-units. I then used a Microsoft Excel macro to extract individual t-units during my analysis.

The questions represented users’ natural language, and users sometimes abbreviated sentences or misapplied punctuation. To guide my unitizing during instances in unclear cases, I developed these guidelines (used in the pilot studies as well):

- When askers wrote incomplete sentences, I relied upon terminal punctuation to mark the end of the t-unit. When askers included no terminal punctuation, I relied upon paragraph breaks, capitalization, or my own sense of the English language.
- When askers included a comma between independent clauses instead of a semicolon, I marked the two clauses as separate t-units.
- I marked as t-units any independent clauses located in parenthetical references.
- When askers embedded t-units in other t-units—for example, between em dashes or in parenthetical references—I marked the embedded t-unit as a separate t-unit from its surrounding text.
- I marked as t-units any independent clauses following introductory colons.
- I marked as t-units any independent clauses in numbered or bulleted lists.
- When askers included two questions in one question with the question mark occurring only after the second question, I marked each question as a separate t-unit.
When askers included links, examples, or specifications, but did not reference or introduce them in a t-unit, I appended them to the previous t-unit.

For the final study, final reliability check, and pilot studies, table 3.2 provides summary statics shows the number of questions, number of t-units, mean number of t-units per question, and median number of t-units per question. These numbers varied slightly throughout the coding process as I discovered instances where I mistakenly unitized data incorrectly; however, table 3.2 shows the final number of units in each phase of the study.

**Table 3.2.** Number of t-units and t-units per question within pilot studies and final study

<table>
<thead>
<tr>
<th>Study</th>
<th>N questions</th>
<th>N t-units</th>
<th>M</th>
<th>Mdn</th>
<th>Max</th>
<th>Min</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final study</td>
<td>500</td>
<td>3,529</td>
<td>7.06</td>
<td>6</td>
<td>37</td>
<td>1</td>
<td>4.26</td>
</tr>
<tr>
<td>Pilot study 1</td>
<td>50</td>
<td>336</td>
<td>6.72</td>
<td>6</td>
<td>16</td>
<td>2</td>
<td>2.97</td>
</tr>
<tr>
<td>Pilot study 2</td>
<td>50</td>
<td>292</td>
<td>5.84</td>
<td>5</td>
<td>19</td>
<td>2</td>
<td>3.32</td>
</tr>
</tbody>
</table>

To answer RQ4 regarding the types of contextual information answerers request in their comments, I also unitized answerers’ comments on unanswered questions into t-units. When unitizing comments, I followed the same unitizing guidelines as I did when unitizing questions, with these additions:

- When answerers quoted phrases from the asker as separate sentences, I appended those to the t-unit following the quotation.
- When short clarification questions followed another question, I appended the clarifying question to the previous t-unit. For example, I would have unitized “What did you install first? Word or Excel?” as only one t-unit.

Of the 250 unanswered questions in the final study, 148 (59.20%) included comments; of the 25 unanswered questions in the first pilot study, 16 (64.00%) included comments; and of the 25 unanswered questions in the second pilot study, 12 (48.00%) included
comments. For these studies, table 3.3 provides summary statistics showing the number of comments, number of t-units, mean number of t-units per comment, and median number of t-units per comment.

<table>
<thead>
<tr>
<th>Study</th>
<th>N comments</th>
<th>N t-units</th>
<th>M</th>
<th>Mdn</th>
<th>Max</th>
<th>Min</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final study</td>
<td>261</td>
<td>690</td>
<td>2.64</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td>1.75</td>
</tr>
<tr>
<td>Pilot study 1</td>
<td>63</td>
<td>35</td>
<td>1.80</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>1.11</td>
</tr>
<tr>
<td>Pilot study 2</td>
<td>20</td>
<td>50</td>
<td>2.50</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Note: The data in this table represent only questions that included comments.

Research Methodology and Data Analysis

To analyze the t-units, I used content analysis (Boettger & Palmer, 2005; Budd et al., 1967; Krippendorf, 2013) because my research questions related to communication and because I needed to analyze communication to answer those questions (Carney, 1971, p. 52). Historically, researchers in communication, social science, and business disciplines have used content analysis in their research; however, the method is growing in popularity among researchers in medical and library science fields as well (Neuendorf, 2017). Content analysis enables researchers to analyze communication data with numerous syntactic or lexical forms that appear in varied contexts (Neuendorf, 2017, p. 17). SQA researchers have successfully used content analysis to analyze user-generated social questions generated within the context of SQA sites (Chua & Banjeree, 2015; Ignatova et al., 2009; Shah, 2012). I therefore built on previous SQA research by using content analysis to answer my research questions.

Researchers define content analysis by its objectivity (Kassarjian, 1977, p. 9; Neuendorf, 2017, p. 17; Smith, 2000, p. 314) with an emphasis on replicability and validity (Krippendorf, 2013, p. 24; Neuendorf, 2017, p. 17). This study established
reliability by using independent coders to assist in the code-development process, and throughout that process I calculated Cohen’s kappa to provide measurable reliability results. (I report these statistics later in the chapter.) Further, I created codes based on an extensive review of previous literature. To establish validity, I recursively refined codes throughout numerous coding cycles. My analysis of answerers’ comments also enabled me to contrast my findings about the contextual information that askers provided, with the contextual information answerers requested, thereby validating supply with demand.

Researchers employing content analysis can analyze information represented in communication and then relate findings to information outside the communication (Budd et al., 1967). For example, SQA researchers have used content analysis to analyze question types and other question characteristics and then related findings to answerability (Chua & Banjeree, 2015; Shah, 2012). In the present study, I first used content analysis to analyze the contextual information that askers provided in both answered and unanswered social how-to questions and then related contextual-information types to answerability. To complete this analysis, I followed these steps to answer my research questions: analyzing contextual-information types, analyzing differences between answered and unanswered questions, analyzing differences in the presence and absence of current- and desired-state information in answered and unanswered questions, and analyzing comments.

**Analyzing contextual-information types**

In both the pilot studies and the final study, I first developed and tested a reliable codebook for coding contextual-information types, which I describe in more detail in the next section and which enabled me to answer my first research question related to the
types of contextual information askers provide. Once I completed coding of the contextual-information types in the 500 questions in the final data set, I calculated the total frequency of each type and described and discussed examples of each type. I report these results in chapter 4 of this study.

**Analyzing differences in contextual-information types of answered and unanswered questions**

Using the data I coded in the previous step, I then analyzed differences between the answered and unanswered questions in the data set. This analysis enabled me to answer my second research question related to the difference in the number of types of contextual information that askers provide in answered and unanswered questions. To examine this difference, I compared the means of the number of types of contextual information provided in answered and unanswered questions. The number of types of contextual information represented ratio-level data because askers could have feasibly included no contextual information in their questions; therefore, I intended to use an independent *t*-test assuming a 95% confidence interval to test differences between means of the number of types of contextual information present in answered and unanswered questions (Wrench et al., 2013). Before running the *t*-test, however, I visually analyzed the distributions using a histogram and discovered that the data was left-skewed. Because *t*-tests assume that both the answered and unanswered data would be normally distributed and not exhibit skew (Wrench et al., 2013, p. 373), I explored non-parametric methods.

I determined to use non-parametric, randomization tests that would enable me to calculate *p*-values without the normal-distribution requirement (Edgington & Onghena, 2007, p. 13). In a randomization test, basic assumptions include that the data is
independent and exchangeable (Good, 2005). Independence means that the probability of one question in the data set being answered or unanswered would not affect the probability of another question in the data set being answered or unanswered (Wilcox, 2009). I randomly sampled data to maximize likelihood of independence.

Exchangeability means that the distribution of a test statistic would be the same regardless of the label we placed on the data (Nichols & Holmes, 2001). Labels in the present study would be answered and unanswered questions. Because the null hypothesis for all difference tests in this study assumed no differences between answered and unanswered questions, I assumed that the data were exchangeable and proceeded with the randomization tests (Nichols & Holmes, 2001). 4

To run the randomization tests, I used the statistical program R, adapting code from Ford (2014) to run the test. Edgington and Onghena (2007) state that all computer-based randomization tests must follow four steps:

1. First, the test must compute a test statistic from the original populations. In the case of the present study, the test statistic was the difference in mean number of types of distinct contextual information in the answered and unanswered questions.

2. Second, the program must randomly select data from both sample populations. In the present study, the two populations were data from the number of types of distinct contextual information in answered and unanswered questions.

Randomization tests assume that the population data is exchangeable; therefore, the data randomly selected values from both populations. The

4 See Appendix B for my additional considerations of unequal variance.
sample size of values was 250, equaling the sample size of the original answered and unanswered questions.

3. Third, the program must compute the test statistic for the randomly selected data. The program then runs steps 2 and 3 iteratively to create a distribution of the test statistic. In the case of the present study, I sampled (without replacement) 10,000 times.

4. Finally, the program must compute a $p$-value. This program calculates the $p$-value by taking the sum of all test statistics from step 3 that are greater than (or less than when appropriate) the original test statistic from step 1. (p. 45)

The result is an estimated $p$-value that provides the probability of receiving a test statistic greater than (or less than) the observed statistic calculated in step 1. I report the specific results of the randomization test in chapter 5 of this study.

In addition to examining the difference in the average number of distinct types of contextual information in answered and unanswered questions, I completed additional data exploration to explore answerability in other ways. First, using a chi-square test, I explored whether the nominal category of presence or absence of specific contextual-information types associated with the nominal category of answerability. In addition to running the chi-square tests, I calculated and reported Cramér’s $\phi$ to determine the percentage of variability in answerability that were explained by the differences in the presence or absence of the contextual-information types. Second, I examined whether the amount of contextual-information types varied between answered and unanswered questions. Because of the skew present in this data, I completed a randomization test, as I have already described earlier in this section, to assess differences in the mean number of
contextual-information types present in answered and unanswered questions. I repeated a similar test to examine whether the average number of individual contextual-information types differed between answered and unanswered questions. Third, to more closely examine the presence or absence of contextual-information types and their amount, I explored whether the proportion of distinct t-units (the number of unique contextual-information types present in each question) to total t-units (the total contextual-information types present in each question) related to answerability between answered and unanswered questions in total and in individual contextual-information types. To test whether the two proportions were equivalent, I used a binomial proportions test as described by Crawley (2005). Finally, I examined whether the word count of contextual-information types varied between answered and unanswered questions. Because of the skew present in this data, I completed a randomization test, as I have already described earlier in this section, to assess differences in the mean number of words present in answered and unanswered questions. I repeated a similar test to examine whether the average word count of individual contextual-information types differed between answered and unanswered questions. I describe the results of all tests in chapter 5.

Because of the exploratory nature of these tests and the additional exploratory tests I conducted throughout this study, I did not set a $p$-value a priori to determine the significance of the tests. If I had set a $p$-value for the tests, I would have set it to .05 and corrected it for possible Type I errors due to compounding because of the multiple tests. However, because of the exploratory nature of my analysis, I interpreted $p$-values less than .10 to explore only potential associations that existed in the data set. Additional studies could test the validity of the findings.
Analyzing presence and absence of current- and desired-state information

In the codebook I used as the basis of coding contextual information types, two broad categories of types were “Current State” and “Desired State.” These categories enabled me to answer RQ3, which related to whether answerability depends on the presence or absence of codes within these categories in answered and unanswered questions. Using the categories and the coded contextual information from the previous steps, I calculated frequencies for the following subsets of answered and unanswered questions: (a) questions with no current-state codes, (b) questions with no desired-state codes, (c) questions with neither code, and (d) questions with both types of codes. Because I examined proportions of questions that separated into two nominal categories, I conducted a chi-square test of independence to determine whether answerability associated with the presence or absence of the current-state and desired-state categories.

Because the previous test assessed differences between only the presence or absence of the current-state and desired-state categories, I explored whether the proportion of information within those categories differed between answered and unanswered questions. To explore these proportions, I examined proportions based on both t-unit and word count using the binomial proportions test described by Crawley (2005). I present all results in chapter 5.

Analyzing comments

I adapted the codebook that I used for analyzing contextual information for use in coding answerer comments. The new codebook reliably described contextual information requested by answerers rather than information provided by askers, as the original codebook described. In the next section, I describe this comment codebook. The
comment codebook enabled me to answer RQ4, which related to the types of contextual information answerers request in unanswered questions. After analyzing the comments, I calculated the total frequency of each comment type and described and discussed examples of each type. I report these results in chapter 5.

In the next section, I discuss how I developed the codebooks described in the previous four steps.

**Codebook Development**

I developed three codebooks for this study. The first codebook enabled me to reliably isolate how-to questions from other types of questions. The second codebook enabled me to reliably code contextual information that askers provided. The third codebook enabled me to reliably code contextual information that answers requested. To develop these three codebooks, coding proceeded in three separate phases.

**Phase 1: How-to questions**

This study specifically focused on how-to social questions, so I first needed to develop a codebook that enabled me to select how-to questions from other types of questions. This codebook also enabled me to further refine my sample of questions. In another study of social questions on Stack Overflow, a subsite similar to Super User, Truede et al. (2011) found that 43% of sampled questions were how-to questions. Consequently, I recognized the need to sample an adequate number of questions to account for some of my data not being how-to questions. Therefore, in my final study, I applied this initial how-to codebook to all 648 answered questions and 493 unanswered questions to ensure that I would have the 250 answered and 250 unanswered questions I desired for my final sample size. Similarly, in the two pilot studies, I initially randomly sampled and coded an
adequate number of questions to ensure that I would have the 25 answered and 25 unanswered questions I needed for analyzing contextual information in the second phase of the study.

Previous studies have shown that human coders can reliably code social question types (Harper, Moy, and Constan, 2009), but no reliable coding scheme existed for coding how-to questions. To prepare the initial codebook, I used Truede et al.’s (2011) definition of how-to questions that stated that how-to questions ask for instructions. I informed subsequent codebooks based on other researchers’ insights as well. For example, Farkas (1999) defined procedural discourse as that which “guides people in performing a task” (p. 42). Therefore, in addition to analyzing questions for askers’ queries for instructions, coders also ensured that askers described a task.

As coders attempted to code questions based on the presence of askers’ queries for instructions and their descriptions of tasks, they faced difficulties determining whether askers were actually asking for instructions. For example, instead of phrasing questions beginning with words such as, “How do I,” many askers began questions with “Is there a way” or “Is it possible” or without explicitly using interrogative words at all. In his analysis of question-and-answer systems, Pilkington (1992) called “How do I” questions “enablement” questions where the intention is to “elicit a plan of action to accomplish a task” (p. 469). In addition, he called “Is it possible to” questions “exploration” questions where the writer’s intention was to “check or seek approval for intended actions or plans” (p. 469). Because “Is there a way?” questions seemed to focus on both enabling a way and the possibility of an answer, I decided to code those questions as how-to questions. Because Pilkington described “Is it possible to” questions
as simply assessing the possibility of questions, I initially determined not to view them as how-to questions.

My closer analysis of Pilkington’s (1992) research and other research, however, suggested that I should include “is it possible” questions as how-to questions if askers intended to receive procedures to accomplish a task. In his analysis of opening messages in software forums, Steehouder (2002) focused more on the intent of askers rather than the explicit ways they asked the questions. For example, he described “is it possible to” questions as “referring to the possibility of something when one really wants advice about how to do it” (p. 497). In addition, he described “Does anyone know how I can” questions as “asking if somebody knows a solution, when one naturally wants to know the answer” (p. 497). Finally, he stated that “I’m wondering how” questions as “emphasizing [sic] one’s ignorance without explicitly asking a question” (p. 497). In each case, Steehouder suggested that the intent of the askers was to seek for information about how to accomplish tasks. I also observed that Pilkington (1992) noted that the intention of most users in his data set was to “ask for instructions” (p. 472) in both enablement questions (for example, “How do I?”) and exploration questions (for example, “Is it possible?”). Based on my own observations and on this prior research, I determined that coders needed to look less at the specific way askers phrased questions and instead look more at the askers’ intentions: how-to questions manifested themselves by the presence of askers’ intention to receive instructions and the presence of askers’ description of a task.

Yet, my closer analysis of the questions suggested that askers directly and indirectly stated their intentions to receive instructions; in addition, some askers merely
hinted their intention. In his analysis of software forums, Steehouder (2002) observed that, “asking help from strangers is a risky social action” because the asker signals weakness through his apparent ignorance and because he places strain on volunteer answerers by asking them to invest resources into composing an answer. Therefore, building on politeness theory developed by Brown and Levinson (1987), Steehouder observed that asking for help could be regarded as a face-threatening act (p. 496). In their theory, Brown and Levinson (1987) assumed that individuals have a “face,” which can be “lost, maintained, or enhanced” (p. 61). Certain acts in conversation can “threaten” face for both speakers and hearers, colloquially known as causing people to lose face (p. 65). For example, a face threat includes a speaker’s comment that binds the hearer to some action, much like when an asker obligates (potential) answerers to post an answer. Politeness theory assumes that individuals generally work together to maintain each other’s faces, and Brown and Levinson (1987) describe numerous politeness strategies to help people mitigate threats to face. For example, strategies for negative politeness include, among others, indirect requests, hedging, and questioning. In his analysis of software forum posts, Steehouder (2002) observed that while some askers directly asked for help, other askers used many politeness strategies, including either indirectly asking for help or not including an explicit request for help at all. He observed that askers indirectly asked for help by including statements such as, “Opinions / suggestions welcome,” or by phrasing the beginning of questions using words such as, “Is it possible to” or “Does anybody know” (p. 497). As I developed the how-to codebook for the present study, I observed that askers of social how-to questions used similar indirect strategies. Mackiewicz and Riley (2003) observed that some indirect requests change the
meaning of the initial request so extensively that they end up “obscuring the underlying” request (p. 90). Building on Blum-Kulka’s (1989) work, they called these hints. For example, instead of asking “Is it possible to insert page numbers,” askers might instead simply state, “I want to insert page numbers.” These hinted requests seemed to reflect what Steehouder (2002) observed in software forums when askers did not include any explicit request at all. Therefore, I incorporated direct, indirect, and hinted requests into this study’s codebook, as shown in table 3.4.

**Table 3.4. How-to question codebook**

<table>
<thead>
<tr>
<th>Code title</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
</table>
| Direct how-to question | A user’s question that directly indicates intent to elicit a plan of action, instructions, or procedures to guide the user in accomplishing a mental or physical task. | • How can one do that?  
• Question: is there a shortcut and what is the shortcut? |
| Indirect how-to question | A user’s question that indirectly indicates intent to elicit a plan of action, instructions, or procedures to guide the user in accomplishing a mental or physical task. | • Is there any way of hiding all of them so I can print like they are empty?  
• Does anyone know how to fix this??  
• Is it possible?  
• Help?  
• Put more simply, can I connect the headers of section 5 and section 7 to section 3 without also including section 4 and section 6? |
| Hinted how-to question | A user’s question that hints at intent to elicit a plan of action, instructions, or procedures to guide the user in accomplishing a mental or physical task. | • I hit something that now makes accented vowels come up whenever I hit the apostrophe key. I don’t know how to fix it.  
• I have seen a lot of questions on page numbering in word 2010, and how to add them, so I know that. I have a particular question regarding the number of pages. For example: I have a document of 5 pages, I want to insert page numbering on page 3 and I want to see the format of page numbering page x of y. If I don this according the the many questions answered here, I start at page 3 with page 1 of 5, but what I really want to see is page 1 of 3, starting at page 3. Page 2 of 3 at page 4, and 3 of 3 at page 5, etc. I couldn’t find the answer on this specific question for far.  
• What am I missing. Is this a bug or a mentioned design feature? If second, then what reason is behind this?  
• In the MS Word world, if someone took a 3rd party template, modified it, and then exported it to a pdf, does the third party name as author persist, or does the licence holder for the application which modified and exported become the author in the metadata? |
| Not a how-to question | A user’s question that does not indicate intent to elicit a plan of action, instructions, or procedures to guide the user in accomplishing a mental or physical task. | • |
In addition to developing the codebook through a review of literature, I developed the codebook and established its reliability through multiple coding sessions with independent coders. Following procedures common to content analysis (Neuendorf, 2016, pp. 40-41), I established reliability in two pilot studies and in the final study. In the pilot studies, the independent coder was an undergraduate student in his senior year who was majoring in technical communication. Because he graduated after the second pilot study, a second independent coder assisted with reliability checks during the final study. She was a first-year master’s student majoring in rhetoric and professional communication. I was the second coder in both cases.

During each coding round, I trained the independent coder on the codebook by reviewing the codes. This coder and I then normed on a subsample of the data, discussed results, adjusted the codebook as necessary, and then coded the full sample. We coded in a quiet room using Excel spreadsheets to capture our codes. On the spreadsheet, the coder would read the question in one cell and indicate the code in an adjacent cell. As we coded, we made notes to help inform the development of the codebook. After each round, we assessed reliability by calculating Cohen’s kappa and then discussed our notes and thoughts to inform the codebook for the next round (if necessary). In the first pilot study, we conducted three rounds of coding and achieved a Cohen’s kappa of .88. In the second pilot study, we conducted four rounds of coding and achieved a Cohen’s kappa of .96. When I assessed reliability in the final study with a second independent coder, we conducted two rounds of coding and achieved a Cohen’s kappa of .89. Landis and Koch (1977) describe all of these kappa statistics as “almost perfect” (p. 165).
As I noted previously, because not all questions would be how-to questions, I
needed to include enough questions to ensure that I would have at least 25 answered and
25 unanswered how-to questions for the pilot studies and the final reliability check. In the
final reliability check, we coded 120 randomly selected questions, 60 answered and 60
unanswered. After achieving our agreement level, I then further sampled 25 answered
and 25 unanswered questions from the 111 questions upon which we had agreed. When I
coded the full data set, I coded all available questions, 648 answered questions and 493
unanswered questions, and then sampled 250 answered and 250 unanswered questions
from the coded results. Table 3.5 shows the distribution of questions according to their
codes for the final study.

**Table 3.5.** How-to question codebook with coding results

<table>
<thead>
<tr>
<th>Code title</th>
<th>Answered total</th>
<th>% of total</th>
<th>Unanswered total</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct how-to question</td>
<td>222</td>
<td>34.25</td>
<td>135</td>
<td>28.13</td>
</tr>
<tr>
<td>Indirect how-to question</td>
<td>256</td>
<td>39.51</td>
<td>210</td>
<td>43.75</td>
</tr>
<tr>
<td>Hinted how-to question</td>
<td>106</td>
<td>16.36</td>
<td>90</td>
<td>18.75</td>
</tr>
<tr>
<td>Not how-to question</td>
<td>64</td>
<td>9.88</td>
<td>45</td>
<td>9.38</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>648</strong></td>
<td></td>
<td><strong>480</strong></td>
<td></td>
</tr>
</tbody>
</table>

Before sampling the data further, I wanted to ensure that answerability did not
associate with the various codes. Therefore, I conducted a chi-square test that showed
that the answerability was independent of the codes: $\chi^2(3, N = 500) = 5.46, p = .141$.
Because answerability appeared not to differ based on the codes, I calculated the
proportion of each how-to code type and stratified my sampling based on those
proportions. Thus in both the 250 answered and unanswered questions, I included 88
(35.20%) direct how-to questions, 114 (45.60%) indirect how-to questions, and 48 (19.20%) hinted how-to questions. Therefore, the numbers of direct, indirect, and hinted how-to questions were constant between the answered and unanswered questions in the final data set.

Once I had determined the final data sets in both pilot studies and the final study, I then proceeded to unitize the sampled questions for contextual-information types. In the next section, I will discuss the codebook for these types.

**Phase 2: Contextual information**

To answer RQ1 related to the types of contextual information askers provide, RQ2 related to contextual-information type and answerability, and RQ3 related to logical coherence and answerability, I then coded the contextual information t-units that I had unitized from the how-to questions selected in phase 1. As shown previously in table 3.2, my final-study unitizing resulted in 3,529 contextual information t-units.

No reliable coding scheme existed for coding contextual information, but I developed the initial coding scheme based on Suzuki et al.’s (2011) five groups of SQA contextual information and Steehouder’s (2002) contextual-information types in forum opening messages. In addition, I grouped the codes based on the current and desired states Farkas (1999) described. In addition, I included an “other” category for the two codes that did not relate specifically to either state. Table 3.6 shows the final codebook.
<table>
<thead>
<tr>
<th>Title</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current state</td>
<td>Information related to the asker’s present circumstances or environment, what led to the asker’s present circumstances or environment, the effects of the current situation, the way the asker currently attains the goal, or successful workarounds.</td>
<td>• I am writing a memo that will go out to all our customers in which we refer to a period of dates (from 06-10-2016 to 12-06-2016 inclusively) • Inserting table of Figures is not pulling in the hyperlink in order to build the separate List of Figures or the List of Tables</td>
</tr>
<tr>
<td>Situation</td>
<td>This code includes no Specifications, Sources, or Examples.</td>
<td></td>
</tr>
<tr>
<td>Specifications</td>
<td>All information related to the Situation Code that also communicates the operating system, software, file types, programming languages, or other system details the asker is using.</td>
<td>• I’ve got a Win7 computer with Office 2013 installed • System language is German</td>
</tr>
<tr>
<td>Sources</td>
<td>All information related to the Situation Code that also communicates sources the asker consulted or references.</td>
<td>• All the information I can find on the web tells me how to do this based on the built-in paragraph style names Heading1, Heading2, etc.</td>
</tr>
<tr>
<td>Examples</td>
<td>All information related to the Situation Code that also communicates screen shots, sample documents, textual examples, code examples, listed steps, and example comparison.</td>
<td>• and when i click the bullet button it shows this:</td>
</tr>
<tr>
<td></td>
<td>This code may include Specifications and Sources.</td>
<td>&lt;a href=&quot;https://i.stack.imgur.com/9DWjV.png&quot; rel=&quot;nofollow noreferrer&quot;&gt;&lt;img src=&quot;https://i.stack.imgur.com/9DWjV.png&quot; alt=&quot;left aligned&quot;&gt;&lt;/a&gt;</td>
</tr>
<tr>
<td>Frustration</td>
<td>All information related to the Situation Code that also communicates the asker’s frustration.</td>
<td>• This is extremely frustrating as there are thousands of these words and fragments</td>
</tr>
<tr>
<td>Knowledge</td>
<td>All information related to the Situation Code that also communicates the asker’s knowledge.</td>
<td>• I’m aware of both &lt;em&gt;fields&lt;/em&gt; and &lt;em&gt;quick parts&lt;/em&gt; features in Word • And I know this isn’t just wishful thinking because Windows Updates does exactly that • I know how to use the Transform section to make circular text shapes</td>
</tr>
<tr>
<td>Thought</td>
<td>All information related to the Situation Code that communicates the asker’s own speculative causes of the current situation.</td>
<td>• i can’t seem to set the default style properly</td>
</tr>
</tbody>
</table>
**Table 3.6. Continued**

<table>
<thead>
<tr>
<th>Previous attempts</th>
<th>All information related to the Situation Code that also communicates the asker’s unsuccessful attempts to satisfactorily accomplish the Task.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• I tried to create a Microsoft WORD &quot;inventory&quot; document that contains a 5-columns/20-rows &quot;Table&quot;; with the end of each row containing an emdedded picture to show the item</td>
</tr>
<tr>
<td></td>
<td>• I’d assumed that to generate a TOC, all I’d have to do would be to tell Word which paragraph styles to include</td>
</tr>
<tr>
<td>Error</td>
<td>All information related to the Situation Code that communicates an error message that the asker receives.</td>
</tr>
<tr>
<td></td>
<td>• Error when I am trying to Combine 2 Word documents - The Security level is set to High. Please run the application which created this document, in the “Security Warning” dialog select the check box &quot;Always trust macros from this source&quot; and enable macros created by Adobe Systems Inc</td>
</tr>
<tr>
<td>Desired state</td>
<td>Information related to the task or goal the asker seeks to attain.</td>
</tr>
<tr>
<td>Task</td>
<td>This code includes no Specifications, Sources, or Examples.</td>
</tr>
<tr>
<td></td>
<td>• but we need to prevent changes to the images (size, position, etc.) in the header and footer areas of the documents</td>
</tr>
<tr>
<td></td>
<td>• We’re trying to create Word templates based on our corporate identity</td>
</tr>
<tr>
<td></td>
<td>• Is there a way I can keep this date period from splitting on two lines?</td>
</tr>
<tr>
<td></td>
<td>• In other words, they should appear if we are exactly on a page break</td>
</tr>
<tr>
<td>Specifications</td>
<td>All information related to the Task Code that also communicates the operating system, software, file types, programming languages, or other system details the asker is using. This does not include document specifications.</td>
</tr>
<tr>
<td></td>
<td>• I am trying to understand styles in Microsoft Word</td>
</tr>
<tr>
<td></td>
<td>• so is there a way to redirect the application to Word 2010?</td>
</tr>
<tr>
<td>Examples</td>
<td>All information related to the Task Code that also communicates screen shots, sample documents, significant textual examples, code examples, or example comparisons.</td>
</tr>
<tr>
<td></td>
<td>• This is how I would like the extracted text to end up after it gets pasted to the word doc - &lt;a href=&quot;https://www.dropbox.com/s/4npic52yt96bxkz/Coiling%20Dragon%20Book%201%20Ch%201.docx?dl=0&quot; rel=&quot;nofollow noreferrer&quot;&gt;Example&lt;/a&gt;</td>
</tr>
<tr>
<td>Motivation</td>
<td>All information related to the Task Code that also communicates an asker’s motivation for accomplishing the task or goal.</td>
</tr>
<tr>
<td></td>
<td>• I would like this feature because I type notes on my computer for my classes</td>
</tr>
<tr>
<td>Thought</td>
<td>All information related to the Task Code that communicates the asker’s own speculative answers.</td>
</tr>
<tr>
<td></td>
<td>• For this reason, I believe the image will need to be embedded in the document</td>
</tr>
<tr>
<td></td>
<td>• I think the the answer probably revolves around a formula field type based on (DOCTYPE &quot;Approval Date&quot;)</td>
</tr>
</tbody>
</table>

• I tried to create a Microsoft WORD "inventory" document that contains a 5-columns/20-rows "Table"; with the end of each row containing an emdedded picture to show the item
• I’d assumed that to generate a TOC, all I’d have to do would be to tell Word which paragraph styles to include
• Error when I am trying to Combine 2 Word documents - The Security level is set to High. Please run the application which created this document, in the “Security Warning” dialog select the check box "Always trust macros from this source" and enable macros created by Adobe Systems Inc
Table 3.6. Continued.

<table>
<thead>
<tr>
<th>Limit</th>
<th>All information related to the Task Code that communicates a constraint on answers or answerers. This code may include Specifications, Sources, and Examples.</th>
<th>but I wouldn’t really want to “recover” my files every time I open it on a different OS. Is there a way I can select the new cell without closing the &quot;Insert Merge Field&quot; box?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>Information that communicates appreciation for answerers’ help or welcomes answers politely. This code may include Specifications, Sources, and Examples.</td>
<td>Thank you!</td>
</tr>
<tr>
<td>Other questions/ comments</td>
<td>A user’s question that does not communicate intent to elicit a plan of action, instructions, or procedures to guide the user in accomplishing a mental or physical task.</td>
<td>What am I missing. Is this a bug or a mentioned design feature? If second, then what reason is behind this? Why?? Why does that happen? Why the red stays, but the blue goes away?</td>
</tr>
<tr>
<td></td>
<td>A comment that does not communicate any contextual information described in the other codes or that is unrelated to the task and situation described in the question.</td>
<td>I can’t get this to work, where do I set it up in Word 2007?</td>
</tr>
</tbody>
</table>

One of the significant challenges I faced when developing the codebook was ensuring that codes were mutually exclusive. In their overview of content analysis, Boettger and Palmer (2010) argue that mutual exclusivity among codes is a necessary prerequisite before exploring statistical relationships in content analysis studies. Because I needed to explore statistical relationships in the data in order to answer the research questions, I worked to ensure that codes were as mutually exclusive as possible.

Achieving mutual exclusivity posed difficulties primarily because the unit of analysis enabled content related to software and hardware specifications, sources, and examples to appear within the same t-unit as types of contextual information. For example, askers would frequently include specifications about their software version when they articulated their tasks or situations.
To achieve mutual exclusivity in the data, I carefully defined the codes, nesting them hierarchically in an order based on what I perceived to be the codes’ frequency. The nesting occurred in the order of the codes appearing in table 3.6. For example, the situation code includes “no Specifications, Sources, or Examples.” By defining the code this way, I captured information relating only to the situation. I defined the next code, specifications, and noted that it included information related to the situation code, but it also included “the operating system, software, file types, programming languages, or other system details the asker is using.” By defining the code that way, I captured specifications information that appeared within the same t-unit as situation information. Continuing, I defined the next code, sources, and noted that it included information related to both the situation code and could include information related to the specifications code, but it also included “sources the asker consulted or references.” This same pattern occurred for the examples code. After the examples code, all other codes could include specifications, sources, or examples, but they primarily focused on the information listed in the code’s definition.

During each coding round, I followed similar procedures as I did in the how-to question coding. Specifically, I trained the independent coder on the codebook by reviewing the codes, norming on a small sample of data, discussing results, adjusting the codebook as necessary, and then coding the full sample. We coded in a quiet room using Excel spreadsheets to capture our codes. On the spreadsheet, the coder would review units within the context of the question, but with each t-unit separated as a distinct color. The coder would make a decision and indicate the appropriate code in an adjacent cell. As we coded, we made notes to help inform the development of the codebook. After each
round, we assessed reliability by calculating Cohen’s kappa and then discussed our notes and thoughts to inform the codebook for the next round (if necessary). In the first pilot study, we participated in nine rounds of coding and achieved a Cohen’s kappa of .75. In the second pilot study, we conducted four rounds of coding and achieved a Cohen’s kappa of .76. When I assessed final reliability with a second independent coder, we conducted five rounds of coding and achieved a Cohen’s kappa of .75. Landis and Koch (1977) describe all of these kappa scores as “substantial” (p. 165).

My third research question depended upon the reliability of the broader categories of the codes. These categories were current state, desired state, and other. To ensure reliability, I assessed Cohen’s kappa based on these categories. To calculate this reliability coefficient, I collapsed all codes into the three larger categories and assessed whether the coders agreed upon the categories. This did not assess whether coders had agreed upon the specific codes, but rather upon the categories that represented the states described by Farkas (1999). In the first pilot study, achieved a Cohen’s kappa of .88. In the second pilot study, we conducted achieved a Cohen’s kappa of .80. When I assessed final reliability with a second independent coder, we achieved a Cohen’s kappa of .85. Landis and Koch (1977) describe kappa scores above .80 as “almost perfect” (p. 165).

Phase 3: Comments

To answer RQ4, which related to the types of contextual information answerers elicit, I coded comments of unanswered questions for the contextual information that answerers elicited. As shown previously in table 3.3, unitizing in my final study resulted in 690 contextual information t-units.
No coding scheme existed for coding the contextual information answerers elicited, but I developed the initial coding scheme based on the codebook I developed in phase 2 and as shown in table 3.6. I redefined the codes within the coding scheme to reflect the same contextual information, but the definitions reflected contextual information the answerers requested as opposed to the contextual information askers provided. Because answerers did not request all types of contextual information, I eliminated codes throughout the codebook-development process based on the codes that the other coders and I encountered in the two pilot studies and in the reliability check for the final study. Table 3.7 reflects the final codes and their definitions.

**Table 3.7. Contextual-information types requested by askers**

<table>
<thead>
<tr>
<th>Title</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current state</td>
<td>A request for information related to the asker’s present circumstances or environment, what led to the asker’s present circumstances or environment, the effects of the current situation, the way the asker currently attains the goal, or successful workarounds.</td>
<td>• Are there any sync conflicts reported in OneDrive?</td>
</tr>
<tr>
<td>Situation</td>
<td>A request for information related to the Situation Code that also requests the operating system, software, file types, programming languages, or other system details the asker is using. This does not include requests for document specifications.</td>
<td>• Exactly which version of Word/Office 2013 are you using?</td>
</tr>
<tr>
<td>Specifications</td>
<td>A request for information related to the Situation Code that also requests screen shots, sample documents, textual examples, code examples, listed steps for reproducing the situation, and example comparison.</td>
<td>• What code have you got so far?</td>
</tr>
<tr>
<td>Example</td>
<td>A request for information related to the Situation Code that also requests the asker’s knowledge.</td>
<td>• Do you know about Styles, namely - Paragraph styles?</td>
</tr>
<tr>
<td>Knowledge</td>
<td>A request for information related to the Situation Code that also requests information about the asker’s unsuccessful attempts to satisfactorily accomplish or research the Task.</td>
<td>• What did you try exactly? • have you given writer administrator privileges? • I was just looking at filtering for Excel again, and wondering whether you had actually tried the “Contains” option</td>
</tr>
</tbody>
</table>
Table 3.7. Continued

<table>
<thead>
<tr>
<th>Error</th>
<th>A request for information related to the Situation Code that also includes a request for an error message the asker receives.</th>
<th>• What is the exact error message about <code>normal.dotm</code>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desired state</td>
<td>• Are you looking simply for changes to your standard template to affect future documents at creation time using that template, or to automatically revise documents after the fact if you later change the master template?</td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>A request for information related to the task or goal the asker seeks to attain.</td>
<td>• Does this question help? <a href="http://superuser.com/questions/487000/ms-word-how-to-disable-spell-grammar-checking-on-a-custom-style?rq=1">http://superuser.com/questions/487000/ms-word-how-to-disable-spell-grammar-checking-on-a-custom-style?rq=1</a></td>
</tr>
<tr>
<td>Thought</td>
<td>A request asking the asker to evaluate a possible answer or proposed solution.</td>
<td>• would putting in a Page Break at the bottom of the previous page work?</td>
</tr>
<tr>
<td>Other</td>
<td>• Just a guess, but if you still have your scanned image selected, you probably won’t be able to create form fields</td>
<td></td>
</tr>
<tr>
<td>General comment</td>
<td>A request or comment that does not request any specific contextual information described in the other codes or that is unrelated to the task and situation described in the question.</td>
<td>• If you click in the document body, the controls may be enabled</td>
</tr>
</tbody>
</table>

During each coding round, the other coder and I followed similar procedures as we did in the previous phases. In the first pilot study, we participated in two rounds of coding and achieved a Cohen’s kappa of .78. In the second pilot study, we conducted one round of coding and achieved a Cohen’s kappa of .91. When I assessed reliability with a second independent coder for the final study, we conducted four rounds of coding and achieved a Cohen’s kappa of .88, which Landis and Koch describe as “almost perfect” (p. 165).

**Conclusion**

In this chapter, I have discussed the methods of this study, including the procedures for gathering the SQA data and for developing and applying the codes that I used to classify and analyze the SQA data. Chapter 4 reports results from the first
research question related to contextual information; the following chapter will report results from the second, third, and fourth research questions related to answerability, logical coherence, and contextual information answerers elicit through their comments; the final chapter will present a concluding discussion of the results.
CHAPTER IV: CONTEXTUAL INFORMATION TYPES

In this chapter, I report on the contextual-information types I discovered in this study’s data, thereby answering the first research question of this study:

**RQ1:** What types of contextual information do askers provide in social how-to questions?

Social how-to questions represent an emerging form of documentation. Similar to the “current states” and “desired states” mentioned and described by Farkas (1999, pp. 42–43), the contextual-information types I discovered coalesced around askers’ current and desired states. In the first section of this chapter, I provide an overview of the information types and their frequency. In the second and third sections, I exemplify and analyze each state’s contextual-information types. In the fourth section, I similarly exemplify and analyze other information types not related to the current or desired state of askers. Lastly, I conclude with a discussion of the findings.

**RQ1: Overview of Contextual Information Coding Results**

As table 4.1 shows, askers provided information related to the current state more than twice as often as information related to the desired state. Given the proportion of desired-state information to current-state information, askers appear to have favored current-state information to desired-state information when communicating their questions. Within information related to the current state, askers provided situation and specifications information most frequently. Within information related to the desired state, askers provided task and limit information most frequently. Askers thus provided substantially more situation information that likely provided context and background information for their tasks.
Table 4.1. Contextual-information t-unit totals by code

<table>
<thead>
<tr>
<th>Code description</th>
<th>Total t-units</th>
<th>% of total</th>
<th>% of subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current state</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>1,084</td>
<td>30.72</td>
<td>48.50</td>
</tr>
<tr>
<td>Specifications</td>
<td>534</td>
<td>15.13</td>
<td>23.89</td>
</tr>
<tr>
<td>Sources</td>
<td>42</td>
<td>1.20</td>
<td>1.88</td>
</tr>
<tr>
<td>Examples</td>
<td>234</td>
<td>6.63</td>
<td>10.47</td>
</tr>
<tr>
<td>Frustration</td>
<td>7</td>
<td>0.20</td>
<td>0.31</td>
</tr>
<tr>
<td>Knowledge</td>
<td>43</td>
<td>1.22</td>
<td>1.92</td>
</tr>
<tr>
<td>Thought</td>
<td>68</td>
<td>1.93</td>
<td>3.04</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>205</td>
<td>5.81</td>
<td>9.17</td>
</tr>
<tr>
<td>Error</td>
<td>18</td>
<td>0.51</td>
<td>0.81</td>
</tr>
<tr>
<td>Subtotal</td>
<td>2,235</td>
<td>63.33</td>
<td>100.00</td>
</tr>
<tr>
<td>Desired state</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>504</td>
<td>14.28</td>
<td>46.67</td>
</tr>
<tr>
<td>Specifications</td>
<td>170</td>
<td>4.82</td>
<td>15.74</td>
</tr>
<tr>
<td>Examples</td>
<td>116</td>
<td>3.29</td>
<td>10.74</td>
</tr>
<tr>
<td>Motivation</td>
<td>28</td>
<td>0.79</td>
<td>2.59</td>
</tr>
<tr>
<td>Thought</td>
<td>70</td>
<td>1.98</td>
<td>6.48</td>
</tr>
<tr>
<td>Limit</td>
<td>192</td>
<td>5.44</td>
<td>17.78</td>
</tr>
<tr>
<td>Subtotal</td>
<td>1,080</td>
<td>30.60</td>
<td>100.00</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gratitude/welcome</td>
<td>152</td>
<td>4.31</td>
<td>71.03</td>
</tr>
<tr>
<td>Other questions/comments</td>
<td>62</td>
<td>1.76</td>
<td>28.97</td>
</tr>
<tr>
<td>Subtotal</td>
<td>214</td>
<td>6.06</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>3,529</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

While the data in table 4.2 provide a different view, they provide further support for how important askers likely found both situation and task information. Whereas table 4.1 provides an estimate of the amount of content dedicated to each contextual-information type in the questions, table 4.2 shows whether each contextual-information type was present or absent in each of the 500 social how-to questions I analyzed in this study. The percentages in table 4.2 show that askers provided situation and task information with highest frequency in questions, with askers providing the situation contextual information in 42 more questions than questions in which they provided task information.
Table 4.2. Contextual-information t-unit presence or absence by total questions

<table>
<thead>
<tr>
<th>Code description</th>
<th>Total questions containing type</th>
<th>% of total questions (500)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current state</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>370</td>
<td>74.00</td>
</tr>
<tr>
<td>Specifications</td>
<td>291</td>
<td>58.20</td>
</tr>
<tr>
<td>Sources</td>
<td>29</td>
<td>5.80</td>
</tr>
<tr>
<td>Examples</td>
<td>151</td>
<td>30.20</td>
</tr>
<tr>
<td>Frustration</td>
<td>7</td>
<td>1.40</td>
</tr>
<tr>
<td>Knowledge</td>
<td>35</td>
<td>7.00</td>
</tr>
<tr>
<td>Thought</td>
<td>57</td>
<td>11.40</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>134</td>
<td>26.80</td>
</tr>
<tr>
<td>Error</td>
<td>15</td>
<td>3.00</td>
</tr>
<tr>
<td><strong>Desired state</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>328</td>
<td>65.60</td>
</tr>
<tr>
<td>Specifications</td>
<td>142</td>
<td>28.40</td>
</tr>
<tr>
<td>Examples</td>
<td>98</td>
<td>19.60</td>
</tr>
<tr>
<td>Motivation</td>
<td>25</td>
<td>5.00</td>
</tr>
<tr>
<td>Thought</td>
<td>59</td>
<td>11.80</td>
</tr>
<tr>
<td>Limit</td>
<td>145</td>
<td>29.00</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gratitude/welcome</td>
<td>132</td>
<td>26.40</td>
</tr>
<tr>
<td>Other questions/comments</td>
<td>50</td>
<td>10.00</td>
</tr>
</tbody>
</table>

In addition to appearing to find important both situation and task information, askers also appeared to find important both specifications and examples information, whether related to their current or desired states. Of the most frequently included information related to the current state, askers provided specifications, examples, and previous attempts. Of the most frequently included information related to the desired state, askers provided specifications, limit, and examples. Regardless of whether askers provided information related to current or desired states, they included both specifications and example information frequently in their questions.

**Current-state information**

Closer analysis of the contextual information that askers communicated in relation to the current state showed that it fell into nine categories: situation, specifications, sources, examples, frustration, knowledge, thought, previous attempts, and error. All of this
information described either the askers’ current state or what askers have done in the past to lead up to their current state. Table 4.3 provides examples of each type of code from the study’s data.

Table 4.3. Examples of current-state contextual information

<table>
<thead>
<tr>
<th>Code description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current state</strong></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>• I have a paragraph of text with 5 different colors</td>
</tr>
<tr>
<td></td>
<td>• Since about two weeks, when I try to close file (Using File/Close), it hangs up</td>
</tr>
<tr>
<td>Specifications</td>
<td>• I am running Mac for Word Version 15.12.3</td>
</tr>
<tr>
<td></td>
<td>• When copy pasting text from another editor to Microsoft Word, paragraphs come with tabs in the beginning</td>
</tr>
<tr>
<td>Sources</td>
<td>• see &lt;a href=&quot;http://superuser.com/questions/155752/change-the-language-of-fields-in-microsoft-word&quot;&gt;Change the language of fields in Microsoft Word&lt;/a&gt;</td>
</tr>
<tr>
<td></td>
<td>• The &lt;a href=&quot;http://office.microsoft.com/en-us/word-help/field-codes-formula-field-HP005186218.aspx&quot; rel=&quot;nofollow noreferrer&quot;&gt;examples&lt;/a&gt; use curly braces</td>
</tr>
<tr>
<td>Examples</td>
<td>• Example of stuff I cannot enter: { IF( =MOD(PAGE, 2) ) = 0 &quot;even&quot; &quot;odd&quot;}</td>
</tr>
<tr>
<td></td>
<td>• However, when I click the Create button it opens PowerPoint to the following page. <a href="https://i.stack.imgur.com/48IbC.png">https://i.stack.imgur.com/48IbC.png</a></td>
</tr>
<tr>
<td>Frustration</td>
<td>• This is extremely annoying</td>
</tr>
<tr>
<td></td>
<td>• I have a really strange and irritating issue using Microsoft Office and OneDrive on my Windows 8.1 machine</td>
</tr>
<tr>
<td>Knowledge</td>
<td>• I know how to get everything, except the &quot;4&quot; i.e. the page number inside the document not considering sections</td>
</tr>
<tr>
<td></td>
<td>• I am aware that this may not be an index in the formal sense</td>
</tr>
<tr>
<td>Thought</td>
<td>• I think it is probably finding all those multiple instances of the text</td>
</tr>
<tr>
<td></td>
<td>• I assume this is because Acrobat sees no visible text in the paragraphs, and hence omits them</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>• So then I tried making an outrageous change to it (16 point red text), thinking I could scroll through text and find it that way</td>
</tr>
<tr>
<td></td>
<td>• I’ve updated the video driver on the system (Intel HD 2500) but to no benefit</td>
</tr>
<tr>
<td>Error</td>
<td>• but I got an error message that said “^” is not a valid special character for the Find What box”</td>
</tr>
<tr>
<td></td>
<td>• and the following error appears before the Figure and Table numbers: Error! No text of specified style in document</td>
</tr>
</tbody>
</table>

In the following section, I show how askers used these types to convey their questions—the request for information at the heart of their communication.
**Situation**

Askers provided 1,084 t-units of situation information, representing 30.72% of the t-units coded in this study’s sample. Askers communicated situation information to describe their current state or what led to their current state. Consequently, askers communicated situation information in various forms of both past and present tenses. When describing the current state, askers used a wide variety of verbs “is,” “are,” “has,” and “have” that indicated that they were focused on their current state. For example, Asker 81’s (A81) question included two t-units that described her current state in present tense:

A81:  

\[
\text{I have a list of 3000 words} <t1>\text{and they are custom words between legal, scientific, and common errors} <t2>\text{I want to add it to my custom dictionary and want to add in one step.} <t3>\text{I want that when the user writes the title it automatically changes the document title so I can have a constant updated title in the header.} <t4>\text{how to do that please?} <t5>\]

In t1, A81 described the list of words that she has, conveying a sense of possession in her current state, and in t2 she described what those words are, conveying a sense of what existed in her current state.

Askers also highlighted the action occurring in the current state. Askers often described action using present-progressive tense to indicate the ongoing nature of their current state. A question from A113 is an example:

A113:  

\[
\text{I am making a template} <t1>\text{and it has a field named Title.} <t2>\text{I want that when the user writes the title it automatically changes the document title so I can have a constant updated title in the header.} <t3>\text{I cannot use VBA.} <t4>\]
\]
In t1, A113 described his action of making a template in Microsoft Word and provided the sense of doing that existed in his current state. A113’s use of present-progressive tense also underscored the continuing nature of the present circumstances because it suggested that the situation persisted even as he wrote the question.

In addition to describing the action occurring in the current state, askers also described the action that led to their current states. Because this information preceded the current state, I call this information “pre-situational” information. Askers generally wrote pre-situational information in simple-past or past-perfect tense to detail actions that led to the present circumstances. A283’s question is an example:

A283: <t1> I copied some rows from a different table, </t1> <t2>now those rows have a different length and cell sizes than the other rows of the table </t2>. <t3>Is there a way to automatically adjust the layout of the row to match the rest of the table? </t3> <t4>(My version is Word 2003)</t4>

In t1, A283 described what she did that led up to the current state. In t2, she then described those current circumstances. Finally, after providing that background information, she asked in t3 about what future actions she could take to bring about the outcome she desired. In addition to illustrating how askers used pre-situational information to detail their actions that led to the current state, A283’s example also illustrates how askers used situational information to provide background and context for their desired states.

Although pre-situational information described past actions that led to the current state, askers also provided pre-situational information in present tense by using the
adverbs “when” and “if” to connect the actions they took in the past to their current state. In his analysis of opening forum messages in software forums, Steehouder (2002) described such “when I” statements as users providing a “scenario of ‘what happened’ describing the situation” (p. 494). Askers included similar “when I” statements in this study’s data set, as exemplified by A45’s question:

A45:  
	<t1>I made a .dot template with a few macros.</t1>  
	<t2>**However, when** I create a new document based on the template I run into trouble.</t2>  
	<t3>**If I reference activedocument, I get the message that no document is open.**</t3>  
	<t4>So how do I select part of the new document?</t4>

A45 began her question by describing in past tense her relevant past actions. Then to bridge from past circumstances to the present circumstances, she included t2 and t3 that described her actions that led to her problematic, present circumstances. She concluded by asking for information about how to accomplish what she wanted.

As askers combined this pre-situational information with information about their current situations, the result was often a description of a problematic situation. The situation was problematic because the asker apparently did not know how to proceed to circumvent the situation. In addition, askers communicated the situation as problematic because it provided rationale to answerers for why the asker posted a question at all. For answerers responding on a volunteer basis, this rationale may signal the reality of the situation and, thereby, motivate answerers to respond (Nam et al., 2009, p. 784). A265 provided an example of a problematic situation:
A265: <t1>I have a table of contents that i have used to access documents in a manual using hyperlinks.</t1> <t2>However, when i click on the document in the TOC, the TOC shows up in the document itself.</t2> <t3>If i scroll up or down and click on the document, the TOC that was in that document disappears.</t3> <t4>This only happens to documents that are 2 or more pages.</t4> <t5>Anyway to get rid of this happening?</t5>

In t1, A265 described the current state of her table of contents and described pre-situational information about how she has used the table of contents. In t2 and t3, she continued describing pre-situational information using “when I” and “if I” statements to describe a scenario of what happened. In t4, she provided additional information about her current state. Taken together, these first four t-units suggested problematic document behavior. However, to confirm that what she described is a problem, in t5 A265 included a request for help to make the document’s behavior stop. A265’s use of situation information to communicate her problem provided important contextual information that may have helped answerers not only to have the information they needed to provide her an answer, but also to understand why she needed instructions to resolve the problem. A265’s task of eliminating her document’s problematic behavior was, therefore, closely connected to her problematic circumstances.

In summary, askers provided 1,084 t-units of situation information, representing 30.72% of the t-units coded in this study’s sample. Such a heavy emphasis on information that led up to or that described the asker’s current circumstances or problems underscores the importance of situation contextual information. As table 4.2 shows
earlier in this chapter, 370 of the 500 (74.00\%) questions included situation information. In his analysis of software user forums, Steehouder (2002) similarly found that a majority of user’s opening messages included situational information (p. 495). Suzuki et al. (2011) specifically analyzed social questions, and they also found that users included situation information related to the askers’ “circumstances” (p. 1261). The findings in the present study reliably confirm the findings in prior studies that askers provide an abundance of situation information in social how-to questions.

**Specifications**

Askers provided 534 t-units of specifications information, representing 15.13\% of the t-units in this study’s sample. Similar to their situation information discussed in the previous section, askers provided specifications information to describe either their current state or what led to their current state as well as information about their operating system, software, file types, programming languages, or other system details. Because all questions in this study related to Microsoft Word, document-focused specifications such as page size, font selections, margin size, list types, document themes, page numbering, and template details manifested ubiquitously in the data set. As noted in Chapter 3, however, this information type included only specifications related to the environment outside the document.

Askers provided specifications information either as the sole focus of the t-unit or as a secondary focus. When providing this information as the sole focus, A64 wrote, “I am using Office 2011 (Norwegian version) for Mac, . . .” In the t-unit, the majority of the information in the sentence related to A64’s software, software versions, and operating system. When providing specifications as secondary focus, A169 wrote, “In MS Word
2010, I have a numbered list.” In the t-unit, A169 communicated the software and version while also including additional information about the list present in her document. Other askers provided specifications in parentheses, again emphasizing the secondary focus of the t-unit.

In his study of software forum messages, Steehouder (2002) found that 23 of 50 (46.00%) forum messages included “hardware and software specifications” (p. 495). As table 2 shows, 291 of the 500 (58.20%) questions I analyzed included specifications information. With the majority of askers providing specifications information, askers clearly find this information relevant to their questions.

Examples
Askers provided examples in 234 of 3,529 (6.63%) t-units in the study’s sample to convey a depiction or even an image of the askers’ current states. When providing examples, askers included abstract textual descriptions, specific text examples, extensive text examples, duplication steps, and sample images and files.

- **Descriptions**: A244 provided a simple description as an example: “For example: Letter 1 has a number of bullet points in the body of the letter populated by different columns in an Excel spreadsheet.” A244 relied upon descriptive words to convey an image of a letter and the mechanism that populated data into that letter.

- **Specific Text Examples**: A25 provided a simple text example that he parenthetically included in one part of his question: “I have some docx documents that contain a lot of footnotes with usually a short footnote text (e.g. "Richard 2010." or "see section xy").” While A25 provided some descriptive text, she also
made that descriptive information more concrete by parenthetically providing specific examples of what she meant by “short footnote text.”

- *Extensive Text Examples:* A146 provided a more extensive text example in her question that duplicated what she saw in her Microsoft Word application on her computer:

  A146: The main body on page 2 looks like this:

  1.1 Text in numbered list
  1.2 Text in numbered list
  2.1 Text in numbered list
  2.2 Text in numbered list

  A146 relied not on descriptions but rather a recreated, concrete example to ensure that answerers understood her current state. Such textual examples as A146’s would likely only work for users conveying images of text and not graphics. Consequently, for questions focusing on Microsoft Word, a text processor, extensive textual examples likely appear more frequently than questions related to more visual software programs, like Adobe Photoshop.

- *Duplication Steps:* Askers also provided the steps answerers could take to duplicate what the askers saw on their computers. A444 provided this type of example:

  A44: Steps to reproduce:

  1. In word 2013, create a new document
  2. Using the "Send to mail recipient" button (which is not on the ribbon by default), show e-mail related fields (To, Cc, Subject, ...)


3. Write your e-mail, including a recipient and a subject, then attach a file.

4. Use "Save as..." (.doc or .docx, it doesn't matter) then close it.

5. Open it again: recipient address and attachment are gone.

In this example A44 provided step-by-step directives that enabled answerers to not only read a description of but also duplicate her current situation on their own computers.

- Sample Images and Files: The most concrete type of examples was actual images or files. A21 provided such an example here:

  A21: But instead of seeing a radio button I see something like this:

  ![image](https://i.stack.imgur.com/RHRbP.png)

  In this example, A21 provided a link to a visual image of an improperly working radio button. Rather than relying on a description of this visual example or using text to recreate the text portion of the image and describe the visual portion, A21 provided an image that concretely and accurately conveyed what he saw on his own computer.

Askers provided examples with relatively high frequency. As shown previously in table 4.1, askers provided 234 of 3,529 (6.63%) t-units in the study’s sample. Table 4.2 also shows that askers included examples related to the current state in 151 out of 500 (30.20%) questions; I discuss examples related to the desired state in the next section; however, it is important to point out here that askers also included examples related to the desired state in 98 out of 500 (19.60%) questions. Neither Steehouder (2002) in his study of forum messages nor Suzuki et al. (2011) in their study of social questions noted the
use of examples in their studies’ data. In contrast, the data in the present study show that askers provide this type of contextual information frequently in their questions.

**Previous attempts**

Askers communicated previous attempts information in 205 of 3,529 (5.81%) t-units to describe unsuccessful attempts they had made to achieve the task or solve the problem. Previous attempts information generally exhibited these characteristics: (1) the asker had communicated the task or problem in a previous t-unit, (2) the subject was a first-person pronoun or gerund, (3) the verb was often in simple-past or past-perfect tense, (4) the t-unit included details about some action the asker took, and (5) the attempted action was unsuccessful. A428’s question exemplified previous-attempts information:

A428: <t1>I use a case management system at work that generates Word documents by inserting data from its database using fields.</t1> <t2>I have a couple of fields that I need to style (bold, capitalize and underline), but only part of the field.</t2> <t3>Is it possible for me to pre-format the string in the database to apply this style to the part of the field before it’s inserted into Word? </t3> <t4>I have already tried applying HTML and STYLEREF but these just display as plain text inside Word.</t4> <t5>Any help would be much appreciated.</t5> <t6>Thanks, [NAME]</t6>

Exhibiting the characteristics described, A428 communicated her task in t2 and t3. She then included previous-attempts information in t4 in past perfect tense and using a first-person pronoun. A428 then communicated her lack of success in t5.
As noted, askers communicated previous-attempts information in 205 of 3,529 (5.81%) t-units in the sample data, making this type of information the sixth most frequent. In addition, previous-attempts information appeared in 134 of 500 (26.80%) questions, making it the seventh most likely information type for askers to include in a question. Steehouder (2002) found that 16 of 50 (32.00%) opening messages in software forums included previous-attempts information. While askers do not devote as much content to describing previous attempts as to other types of contextual information, askers still provide previous-attempts information with relatively high frequency.

**Thought**

Askers provided thought information in 68 of 3,529 (1.93%) t-units to speculate about the causes of the problems they communicated in their question. Coders identified this information by the presence of speculative words such as “think,” “believe,” “appears,” and “seems,” and by the presence of a cause. A244 included thought information when he faced with a problem with mail merge functionality in Word: “I’m wondering if this might something related to a weird hidden formatting option somewhere within Word that I am not aware of that is different between the Merge Master and the Letter 1 templates causing things that were hidden to reappear.” The use of “might” suggested the speculative nature of A244’s thinking; in addition, he clearly stated that he was thinking about the cause of the problem in the latter part of the t-unit.

While askers provided thought information with a relatively low frequency when compared to the total t-units in the data set, askers included the information in 57 of 500 (11.40%) questions. Steehouder (2002) similarly observed that users in forums provided contextual information in the form of “suggestions of what may have caused the
problem” (p. 494). While askers wrote few t-units dedicated to their thoughts about causes of problems, the presence of thought information in 1 in 10 questions suggests that askers view this contextual information as fairly relevant to many questions.

Knowledge

In 43 of 3,529 (1.22%) t-units asked communicated their knowledge to accomplish a wide variety of purposes. Askers expressed knowledge using phrases such as “I know” and “I am aware.” For example, A33 simply stated, “I know how to export PDF from Word.” Coders could easily and reliably identify these t-units based on the content manifest in the questions. However, a closer examination of these t-units in context suggested that intent varied from asker to asker. Rather than simply expressing their knowledge, askers used these t-units to (a) express their current state through knowledge, (b) justify a question, and (c) manage answers.

• Express Their Current State: Askers communicated their current state by describing what they knew and what they did not know. For example, A254 wrote this question:

A254: <t1>I know how to make comments appear on the right side of the document</t1>. <t2>But is there a way to list down all comments made in the document so I see all of them at once instead of scrolling down the entire doc looking for them</t2>?

In t1, A254 expressed what he knew. This knowledge served to communicate the current state of his knowledge. He followed that statement with t2 that communicated the knowledge that he desired to know. The gap between what he
knew and what he desired to know was the instructions that he needed from answerers.

- **Justify a Question:** Askers used knowledge to justify why they were asking their questions. For example, A473 communicated a task of needing to “specify a page range to print in MS Word 2010.” Her problem, however, was that the field to specify a page range was grayed out. She justified the question by stating, “I know that this is achievable on a .doc in Word 2010 because I’ve created a blank .doc in which the Pages option is not grayed out.” In this t-unit, she called upon her previous experience to justify that her task was possible, but she just did not know how to accomplish it.

- **Manage Answers:** Askers also stated what they knew in order to manage the types of answers that answerers would provide. For example, here is an excerpt from A78’s question:

A78:  
<br/>&lt;t1&gt;I know about Quick Parts&lt;/t1&gt;, &lt;t2&gt;but they require the salesperson to go to the right place in the document and select the right quick part&lt;/t2&gt;. &lt;t3&gt;I’d prefer something that’s more standardized&lt;/t3&gt;.

In t1, A78 stated that he knew about the Quick Parts functionality of Microsoft Word. However, his knowledge operated in tandem with t2 where he detailed why this functionality would not work in his situation. In t3, he explicitly stated his preference to not use Quick Parts, thereby reducing the likelihood that an answerer would suggest Quick Parts as a solution to his problem.

In their study of social questions, Suzuki et al. (2011) included asker knowledge as a facet of situation information, yet the authors did not provide information as to how
frequently askers provided knowledge. In the present study, explicit statements of what
the askers knew appeared in 43 of 3,529 (1.22%) t-units. These 43 t-units appeared in 35
of the 500 (7.00%) questions analyzed, representing a relatively small set of all questions.
However, as evidenced by the preceding examples, askers appeared to strategically use
their knowledge for a wide variety of purposes.

Sources

Askers provided 42 out of 3,529 t-units (1.20%) of sources information to describe a
source they consulted or otherwise referenced in their current state. Generally, these
sources took the form of hyperlinked text, although sometimes askers would refer to a
source by name. For example, A300 mentioned “a Microsoft forum.”

Askers provided sources information with relatively low frequency, and these t-
units appeared in only 29 of 500 (5.80%) questions. However, as discussed in chapter 3,
other information types may have subsumed sources information. Steehouder (2002)
theorized that askers in user forums might include sources to “justify” the asker’s use of
the forum (p. 496). By showing sources they consulted, askers may effectively signal that
other resources provided no help and that the asker has already put some effort into
answering her own question. Askers may, therefore, include sources strategically in
order to convey other contextual information, such as previous attempts, which may have
subsumed some sources information in the data set.

Error

Askers provided error information in 18 of 3,529 (.51%) t-units in the study to convey the
specific error messages the askers encountered. Errors manifested as the result of an
askers’ unsuccessful attempts to take some action. A370 provided an example in this excerpt of her question:

A370: <t1>I am using Word 2007.</t1> <t2>I modified a version of the file which was saved to my harddrive and saved my changes.</t2> <t3>When I wanted to open the file again, I went ‘Recent Documents and clicked the file.</t3> <t4>However, I received an error message: "This file could not be found" which referenced the path to the file in the harddrive.</t4>

In this question, A370’s action was to modify and save a file, and then to try to load the file again. The result was the error message communicated in t4.

These error messages occurred in 15 of 500 (3.00%) questions in the study. Steehouder (2002) observed error messages in 6 of 50 (12.00%) opening forum messages in his study. The difference may arise from the better documentation companies now make available. Askers can easily search for errors on the internet, resulting in askers posting fewer questions online.

**Frustration**

Askers included frustration information in 7 of 3,529 (.20%) t-units to convey their unpleasant feelings about their current state. Common terms askers used included “annoying” and “irritating.” For example, A48 stated, “This is also really annoying as when I print labels, the printed area over laps the size of the stickers.” A172 expressed ongoing frustration when she stated, “One thing I’ve always hated about Windows is that to type a non-English character, like __, on a US keyboard you have to hold the Alt key and type in a four digit code on your num-pad.” In these examples, the askers expressed
clear disdain for the software and undesirable circumstances that the software caused for them.

Most instances of frustration focused on Word and its interface with other hardware and software. In addition to the previous example from A172 that expressed frustration related to Word and typing characters on a keyboard, other askers expressed frustration related to Word and printing, Word and online templates, Word and email, Word and cloud software, and Word and file management. Microsoft Word interacts with many other software programs and hardware, and askers appear to express frustration most frequently when this interaction fails to meet their expectations.

Yet despite many askers describing problems with Word in their questions, the data set included relatively few instances of frustration. In their study of computer users and frustration, Ceaparu, Lazar, Bessiere, Robinson, and Shneiderman (2004) found that besides web browsing, word processing caused more frustration for users than all other sources in their study. As noted, askers included frustration information in only 7 of 3,529 (.20%) t-units, and these t-units appeared in 7 of 500 (1.40%) questions, the fewest present in the data set. This amount shows that askers do not express frustration frequently in social how-to questions. Further, as shown by the equality of t-units to questions in which frustration information occurred, askers tended to express frustration once rather than multiple times, suggesting that askers did not use questions to express lengthy, frustrated rants.

 Desired-state Information

In addition to providing contextual information related to the current state, askers also provided contextual information related to their desired state. Closer analysis of this
contextual information showed that it fell into six categories: task, specifications, examples, motivation, thought, and limit. All of this information described the askers’ goals, described reasons for the goals, or asked answerers how to achieve the goals. Table 4.4 provides examples of each type of code from the study’s data.

**Table 4.4. Examples of desired-state contextual information**

<table>
<thead>
<tr>
<th>Code description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Desired state</strong></td>
<td></td>
</tr>
</tbody>
</table>
| **Task**         | • How can I go back to seeing 1 page at a time?  
                   • I need to update the Rich text content control with some new lines |
| **Specifications** | • How do I remove all field codes from a Word document?  
                      • I need to specify a page range to print in MS Word 2010 |
| **Examples**     | • How can I create the following effects (red circles) for a table in Microsoft Word?  
                      https://i.stack.imgur.com/5rVvp.png  
                      • I am trying to make a document in which I can add a specific date at the top (for example, 03 October 2014) |
| **Motivation**   | • The reason it is a protected document is that we have about 200 companies with different directors  
                      • I’d like to have the caption associated with it to be at the top of each page so that the reader doesn’t forget which table they are reading |
| **Thought**      | • I think the answer probably revolves around a formula field type based on {DOCTYPE "Approval Date"})  
                      • Maybe, there is a way to use VBA or some scripting to fix this |
| **Limit**        | • How to fix it (without learning and switching over to latex)? I am aware that this may not be an index in the formal sense  
                      • I cannot use VBA |

In this section, I discuss these information types in order of their frequency of occurrence.

**Task**

Askers provided task information in 504 of 3,529 (14.28%) t-units to declare their needs and wants and to ask for instructions. When declaring needs and wants, askers used statements such as “I need,” “I want,” “I have to,” and “I’d like to,” to communicate their desires. An excerpt from A7’s question provided an example: “I need to be able to view, measure and resize the objects on the screen with a real ruler, and then print them to that
exact size.” By stating her need, A7 described and communicated her task and desired state. However, in this t-unit she left unstated any invitation for answerers to provide instructions.

Other answerers were more direct in their asking for instructions. Rather than only declaring their needs and wants in task information, askers also invited responses by structuring the information as questions. They phrased questions using words such as, “How do I,” and “Is there a way to,” to invite answerers to provide instructions for accomplishing the task. A365’s question provided an example: “Is there a way to auto fit the shape to a cell size permanently?” In this example, A365 stated the task (“to auto fit the shape to a cell size permanently”), and the question’s interrogative syntax invited responses from answerers. By using question syntax, askers included in one t-unit both the task and an invitation for answerers to provide instructions.

Other askers communicated task information less explicitly in their questions because of their use of vague pronouns. An excerpt from A352’s question exemplified this practice:

A352: <t1>I have created a table in MS Excel</t1> <t2>and after that I pasted that as image in MS Word.</t2> <t3>The affair is that when I exported the word file as pdf, the border of some lines is viewed thicker than it is</t3> <t4><a href="https://i.stack.imgur.com/YSJnw.png" rel="nofollow noreferrer">check it out</a></t4>. <t5>However, if I zoom in, it is looked fine without any problem</t5>. <t6><b>How can I fix this?</b></t6>

In t6, A352 communicated a question that invited answerers to respond. However, the task she communicated is vague due to her use of the pronoun “this.” A352 relied upon a
lengthy discussion of her situation to provide any necessary detail about her task. Once she had detailed her situation in the question’s first five t-units, she then referenced the t-units generally by asking how to fix the problem they described. Task statements, such as the one A352 used, showed how tightly tasks related to the situation of askers and how askers relied upon situation information to communicate their tasks. Such statements also illustrated how askers used pronouns to cut down on question length. For answerers, these statements were likely less clear because the pronoun required answerers to synthesize multiple situation statements to understand the task.

As noted, askers provided task information in 504 of 3,529 (14.28%) t-units. In their study of social questions, Suzuki et al. (2011) found that askers provided task information related to their “reasons, motivations, aims, and goals” (p. 1261). In his analysis of opening messages in software forums, Steehouder (2002) also observed task information that he described as “goals that the user wanted to achieve” (p. 493). The findings in the present study reliably confirm the findings in these previous studies that askers provide task-related information.

The findings also suggest the relative importance of task information for askers. Askers provided task information as the third most frequent type of contextual information, falling behind situation information (1,084 of 3,529 t-units, 30.72%) and specifications information (534 of 3,529 t-units, 15.13%) that related to askers’ current states. Even though askers provided situation t-units in nearly double the amount of task t-units, much of this difference disappears in a comparison of the number of questions where the two types of contextual information appeared. Askers provided task information in 328 of 500 (65.60%) questions and situation information in 370 of 500
(74.00%) questions. Askers thus devoted slightly more content toward situation information than task information, yet they included both types of information in a majority of questions.

**Limit**

Askers provided limit information in 192 of 3,529 (5.40%) t-units to communicate answers that would not work. Askers limited answers in three ways: stating directly that they did not want or could not accept a certain type of answer, including limiting prepositions in their questions, and comparing their current procedures to their desired procedures. Examples of each method follow:

- **Direct Statement:** A160 communicated a task in a t-unit and then followed that task statement with this t-unit: “To be clear, I’m not interested in "ignoring" it once.” To circumvent any answerers related to “‘ignoring’ it once,” A160 directly stated what she did not want in an answer. Worded slightly differently, A19 also directly stated what he did not want: “and i really don’t want to make a table for it manually.” By including such direct statements, askers tailored their questions to their needs and minimized the likelihood of answerers providing undesired answers.

- **Limiting Prepositions:** Some askers phrased their questions to include limiting prepositions. A101’s question provided an example: “Is there a way to make form letter variants in MSWord without VBA?” In this question, A101 communicated both the task and the limiting information into one t-unit through her use of the preposition “without.”
•  **Comparing Current Procedures to Desired:** Askers also provided limit information after first stating procedures that they can follow to successfully complete a task. The limiting information appeared in a following question that included a comparative adjective that showed the asker did not want to follow the previously stated procedures. A184 provided an example. He first provided situation information that stated how he could manually create a table. He then provided this limiting t-unit: “But is there a quicker way?” The question implied that A184 did not want to manually create a table, effectively communicating to askers that they should not provide manual procedures as an answer.

Askers provided limit information in 145 of 500 (29.00%) questions. In their study of social questions, Suzuki et al. (2011) also observed that askers provided “conditions on answers and answerers” as contextual information. In interviews with askers, Jeon & Rieh (2015) also found that askers seeking high-quality answers used strategies to “narrow down options,” such as including information about “what is not an option for the asker” (p. 6). The present study reliably confirms findings in these studies and provides examples of how askers communicated this limiting information.

**Specifications**

Askers provided specifications information in 170 of 3,529 (4.82%) t-units to convey specifications information related to their tasks. Similar to specifications that askers provided in relation to the current state, specifications related to the current state manifested in the form of information about operating systems, software programs, file types, programming languages, or other system details.
Similar to task information, specifications information manifested in both declarative task statements and task-related questions. A76 provided an example of a declarative statement in his question: “I want to reorganise my Word 2010 document and have the heading styles update accordingly.” A341 provided an example of a task-related question: “How can I do this in Word 2016?” Because software can vary in look, feel, and operation in different versions and on different operating systems, specifications information likely provided answerers important contextual information as they responded with instructions about how to accomplish these askers’ tasks.

As noted, in relation to the desired state, askers provided specifications information in only 170 of 3,529 (4.82%) t-units. This percentage contrasts greatly from the specifications information that askers provided in relation to the current state. As discussed previously, askers provided specifications information (current-state) in 534 of 3,529 (15.13%) t-units. Askers thus provided specifications information much more frequently when communicating about the current state than about the desired state.

Askers also tended to communicate their specifications in one state or the other and then not communicate their specifications in relation to the other state. Askers communicated specifications information related to either the current state or the desired state in 347 of 500 (69.40%) of all questions in the study’s data. However, only 86 of those 347 (24.78%) included specifications information related to both current state and the desired state. Of the remaining 261 questions, askers provided specifications related to only the desired state in 56 (21.46%) questions, and specifications related to only the current state in 205 (78.54%) questions. Askers, therefore, tended to provide specifications information in relation to either the desired state or the current state, but
not both. In addition, when askers provided specifications information in relation to only one state, askers favored providing that information in relation to the current state.

**Examples**

Askers provided examples related to the desired state in 116 of 3,529 (3.30%) t-units to provide descriptions or images of what they wanted to accomplish. Similar to the examples they provided in relation to the current state, askers provided examples in text, image, and file formats. While the formats were the same, askers provided examples related to the desired state differently through two types of comparison:

- **Simile**: Because the askers’ desired states did not yet exist, askers had to find a way to communicate a vision of what they wanted. One way they communicated this vision was to compare their desired state to something that already existed. A8’s question provided an example:

  A8:  
  <t1>I want to replace fonts of digits and some other characters automatically as I type in Microsoft Word 2013 (something like word Auto-correct option)</t1>. <t1>How can I do this by Macros? ’</t1>

  In his question, A8 described his want and then parenthetically compared it to existing functionality in Word. By describing his want in this way, A8 provided an indirect example of his desired state through a comparison.

- **Before and After**: Askers also provided an example of both the current state and the desired state to communicate what they wanted. A274’s question provided an example:

  A274:  
  <t1>I have a large word document with multiple cross-references to figures, tables etc.</t1>  
  <t2>How can I add a page number to these
So, for example, "See Table 2" would become "See Table 2 (p. 123)."

Is this possible? Perhaps with a Macro or VB script?

In t2, A274 provided a question that included her task. In t3, she provided an example related to her current state and showed what that example would look like in her desired state. This comparative example clearly enabled answerers to see the gap between what is and what should be: the addition of page numbers.

While askers provided examples related to the desired state in only 116 of 3,529 (3.30%) t-units in the study’s data, these 116 t-units were present in 98 of 500 (19.60%) questions. As noted previously, neither Steehouder (2002) nor Suzuki et al. (2011) mentioned the existence of examples in their studies of contextual information in forums and social questions; the present study’s data showed that askers relied frequently on examples to communicate their desired states.

However, a comparison of examples related to the current state and the desired state shows that askers provided examples related to the current state much more frequently and in different formats. Askers provided examples in 234 of 3,529 (6.60%) t-units, or nearly double the amount of examples that askers provided related to the desired state. When askers post images in their questions, the Super User system stores the images on a site called Imgur. I completed a simple keyword search for Imgur images in the t-units where askers provided examples related to both current and desired states. I found that askers shared images in 33 of 116 (28.44%) t-units related to the desired state and 107 of 234 (45.73%) related to the current state. These statistics suggested that askers provided more image examples related to the current state than to the desired state. This
finding likely stems from the ease with which askers could take screenshots of what they saw on their screens (i.e., the current state). When they wanted to take a screenshot of the desired state, however, askers had to create an example since it did not yet exist. Providing descriptive textual examples then is likely more time efficient and, therefore, the askers’ likely preferred method for communicating examples related to the desired state.

**Thought**

Askers provided thought information related to the desired state when they tentatively or speculatively offered potential solutions to their problems or answers to their questions. Thought information manifested through the words “think” and “believe” and through a question mark at the end of askers’ suggested solutions or answers. In addition, thought information manifested as askers opened up possibilities for answers by suggesting types of answers that might work for the asker’s situation. Examples of these strategies follow:

- **Think or Believe:** A223’s question provided an example of tentative wording: “For this reason, I believe the image will need to be embedded in the document, . . .” In this t-unit, A223 provided her own thought about what needed to happen for her own situation. However, she tempered her thought using the word “believe” to emphasize her own thinking.

- **Questions:** Answerers also emphasized their own thinking by structuring their thinking in the form of questions. An excerpt from A198’s question exemplified this method: “Maybe apply a registry fix?” The use of a question mark emphasized the speculative nature of A198’s suggestion, showing that A198 did
not did not know how to proceed with the fix, but had at least thought about the fix as a possibility.

- **Open Possibilities:** In addition to emphasizing their thinking, askers also opened up possibilities for answerers more directly. A168 provided an example: “Even a VBA fix is welcome.” In this t-unit, A168 communicated a possible way for answerers to accomplish the task. Rather than limiting answerers to a certain answer type or style, however, A168 opened up possibilities by suggesting that she had the ability and knowledge to use VBA if that is what the answerer required.

Askers provided thought information in 59 of 500 (11.80%) questions in this study’s sample. In their study of social questions, Suzuki et al. (2011) similarly observed that askers provided their “own thought, answer prediction” as contextual information. The present study thus reliably confirms findings from that study.

Askers used words such as “think” and “believe” and question marks to convey their thoughts. Researchers have also interpreted these strategies as ways to downgrade the directness of communication (Mackiewicz & Riley, 2003, p. 86). As askers wrote questions, they placed an obligation on answerers to respond; by using strategies to downgrade the types of answers they seek, askers may have attempted to communicate politely rather than share their own thoughts. Despite nuances in intention, this study shows that askers provided thought related to their desired state in a moderate number of questions.
Motivation

Askers communicated motivation information in 28 of 3,529 (.79%) t-units to describe what motivated their tasks. Askers communicated their motivation for accomplishing their tasks using phrases and conjunctions such as “so that,” “since,” “in order to,” and “because.” Askers viewed their tasks as means, whereas the motivation was the ends. For example, A312’s was to select a font, but she revealed her ends when she communicated her motivation: “I need to select this font in order to have the template rendered correctly.” A312 saw the font selection as a means to the end of having a properly rendered template. A75 provided a similar example: “The reason I want to copy the code is because I want to adjust some sizes for certain parts and print it out.” A75 similarly asked for instructions about how to perform some action, but the motivation was the true outcome he sought: adjusting sizes and printing. Askers, therefore, viewed their tasks as the means to effect greater ends.

Askers communicated these ends as effects on documents, as shown in the previous examples, but they also communicated these ends as effects on other people. A65’s motivation was an effect on readers: “I’d like to have the caption associated with it to be at the top of each page so that the reader doesn’t forget which table they are reading.” A65 saw the task as associating a caption that led to an effect of readers not forgetting what they were reading. A438’s task was to convert dumb quotes to smart quotes, and her motivation was consistency with others’ settings: “The purpose of this is that this document/template should always work like this, regardless of what computer/user opens it.” These people-focused motivations highlighted that askers
worked and interacted with others, and accomplishing their tasks enabled them to achieve goals beyond making changes in documents.

Askers provided motivation information third-to-least frequently: 28 of 3,529 (0.79%) t-units and in 25 of 500 (5.00%) questions. Despite this relatively small amount, Suzuki et al. (2011) also found that askers provided “motivation” contextual information in their study of social questions (p. 1261). The present study, therefore, reliably confirms prior research findings by Suzuki et al. and shows that effects on both software and people motivated askers. However, motivation information is complicated because, although askers explicitly communicated motivations as the intended outcomes they sought to achieve, they also communicated motivations explicitly as situations they wanted to circumvent. In that way, askers may have indirectly communicated their motivations as they described problematic situation information. Despite this nuance, this study revealed that askers infrequently communicated their motivation in the direct ways discussed in this section.

**Other information**

In addition to providing information related to the current and desired states, askers also provided other information. This information coalesced into two categories: (1) gratitude and welcome, and (2) other questions and comments.

**Gratitude/welcome**

Askers provided gratitude and welcomed answers in 132 of 500 (26.40%) questions to politely express appreciation for answerers’ help or to otherwise politely welcome answerers. Askers generally communicated gratitude using words such as “thank you” and “appreciate.” They communicated gratitude succinctly and toward the end of their
questions. A3 exemplified brevity: “Thanks!” Notably, of the askers’ 152 instances of gratitude or welcoming information, 116 (76.31%) occurred in the last t-unit in the question. Similar to other social communication, ending questions with a statement of gratitude thus appeared to be the convention in these questions.

Information that politely welcomed answers manifested most frequently in two ways: (a) through the user of the word “please” and (b) through other short inquiries such as “any ideas?” A139 provided an example of politely welcoming answers: “Please help me with this.” A427 similarly welcomed answers politely, but less directly: “Any help would be much appreciated.” In addition to using polite words, askers also included short, polite invitations to welcome answers. A19’s invitation to answerers provided an example: “Any ideas?” Steenhouder (2002) similarly found that users seeking help in forums frequently included indirect requests for help, such as “Any solutions?” or “suggestions welcome” (p. 497). He theorized that users communicated requests indirectly as a politeness strategy. The findings in the present study confirm that askers use similar politeness strategies in social questions.

Askers provided gratitude and welcomed answers in 132 of 500 (26.40%) questions in the study, placing this information as the eighth most frequently type of information in the questions. Askers posed questions to volunteer answerers, yet their statements of gratitude and politeness appeared in only one of four questions. While this number might seem low, askers might view their relationship with answerers as mutually beneficial. Askers may be participating on the site both as askers and answerers, so they may consider their gratitude fulfilled when they themselves post answers to others’ questions.
Other questions/comments

Askers communicated other questions and comments in 62 of 3,529 (1.76%) t-units to ask questions that did not elicit instructions from answerers or to provide information not related to the current or desired state. When not eliciting instructions, askers structured questions using interrogatives such as “why” and “what.” Such questions signaled the intent of askers to find out the cause or effect of their problems. In the infrequent commentary that did not relate to the current or desired state, askers commented on the functionality of the Super User site itself and on the process of writing their questions. Some askers also included their names.

However, as evidenced by the fact that such commentary occurred in only 62 of 3,529 (1.76%) t-units, askers provided relatively little other questions and comment information. Of these 62 t-units, 49 (79.03%) were other questions and 13 (21.0%) were other comments. Askers included other questions in 40 of 500 (8.00%) questions in the study’s sample, thus showing that the majority of the how-to questions in the sample did not ask additional questions unrelated to eliciting instructions. Other comments occurred in only 12 of 500 (2.40%) questions. Social question and answer sites distinguish themselves from forums by their focus on asking and answering questions and not on unnecessary chitchat; the results show that askers tended to stay on topic by minimizing information that was unrelated to the current and desired state.

Conclusion

The results in this chapter provide examples and frequency statistics of the contextual-information types that askers provided in social how-to questions. Specifically, results show that askers provided situation, knowledge, thought, task, and limit information.
These results provide reliable support for similar contextual information factors observed by Suzuki et al. (2011) in social questions. Results also show that askers provided presituational, specifications, error, thought, sources, and previous-attempts information.

These results provide reliable support for similar contextual-information types described previously by Steehouder (2002) in his study of user software forums. Results also show that askers provided examples with moderately high frequency. Other researchers have not yet fully described examples as a contextual-information type, so this finding contributes to the information types already described by other researchers.

The results in this chapter also show that askers provided contextual information related to their situation most frequently. As noted in chapter 2, researchers investigating documentation have theorized about the importance of situation information. Specifically speaking of users who engage in complex computer work, Mirel (1998) argued that situation information is what lends significance and meaning to tasks and, therefore, called for documentation writers and researchers to recognize the importance of situation information (p. 14). As shown in table 1, askers in the present study provided twice as much situation information as task information. The relative abundance of situation to task information lends support to Mirel’s argument.

In summary, chapter 4 described the contextual-information types that askers provided. In chapter 5, I compare the frequency of this information between answered and unanswered questions. After comparing the types of contextual information askers provided, I then describe the contextual information answerers requested in their follow-up comments to unanswered questions.
CHAPTER V: CONTEXTUAL INFORMATION, ANSWERABILITY, AND COMMENTS

In the previous chapter, I examined the types of contextual information askers provided in social how-to questions. In this chapter, I investigate how contextual information relates to answerability in questions, thereby answering the second, third, and fourth research questions in this study:

**RQ2**: Do answered and unanswered how-to questions differ significantly in the number of distinct types of contextual information they include?

**RQ3**: Do answered and unanswered how-to questions differ significantly in whether they include contextual information related to desired states and current states?

**RQ4**: What types of contextual information do answerers most frequently elicit through their comments on unanswered how-to questions?

The first section of this chapter answers RQ2, investigating the relationship between answerability and the questions’ distinct types of contextual information. The second section answers RQ3, investigating the relationship between the logical coherence of questions and answerability. The last section answers RQ4, focusing on the contextual information answerers requested in unanswered questions.

**RQ2: Contextual Information and Answerability**

Previous scholarship related to contextual information examined answerability based on only one type of contextual information (Suzuki et al., 2011) or did not report specifically whether a question received an answer (Harper et al., 2008; Suzuki et al., 2011). The first section of this chapter, therefore, expands on previous research by examining multiple contextual-information types and their relationship to both answered and unanswered
questions. As discussed in chapter 3, I compared the contextual information in 250 answered and 250 unanswered questions. Throughout this chapter I refer to distinct (i.e., unique) contextual-information t-units as types, and the magnitude (i.e., amount) of these contextual-information t-units as tokens.

In this section, I first examine whether answerability relates to distinct contextual-information types, or the presence or absence of information. I found that while answered and unanswered questions did not differ in the number of types of contextual information they included, the presence or absence of thought information (current-state) and gratitude/welcome information did associate with answerability. In each case, answered questions provided fewer of each type.

Second, I move beyond presence and absence to explore whether the magnitude (i.e., the amount) of contextual-information tokens relates to answerability. I found that answered and unanswered questions differed in the number of contextual-information tokens they included, with answered questions including fewer tokens than unanswered questions. I found that a difference in the number of situation-information tokens largely explained this difference.

Third, I examine contextual-information variation by looking at both contextual-information types and tokens by exploring whether the proportion of distinct information to magnitude (i.e., type-token ratio) relates to answerability. I found that answered and unanswered questions differed in their type-token ratios, with answered questions including more t-unit variation. To measure efficiency, I also examined the type-token ratio of individual contextual-information types. The efficiency of examples (desired-
state), situation information, and task information related to answerability. Answered questions were more efficient for each type.

Finally, I compare the word count between answered and unanswered questions. I found that answered questions included fewer words than unanswered questions. When I compared individual contextual-information types, I found that in comparison to unanswered questions, answered questions included fewer words of thought information (current state) and examples (current state). Comparing the average word length of t-units, I found that error information, situation information, other information, and situation examples differed based on whether they were in answered or unanswered questions, with answered questions including shorter t-units in all cases except situation information.

**Distinctness**

I first compared the number of distinct contextual-information types in the answered and unanswered questions. By distinct, I mean the number of unique contextual-information types present in each question. For example, if a question included two t-units of situation information, two t-units of task information, and one t-unit of gratitude/welcome information, the total number of contextual-information types would be three—one for each of the three types of contextual information present in the question.

I analyzed the quantile-quantile (Q–Q) plots of both answered and unanswered questions as compared to a normal distribution and visually discovered that their distributions were skewed and did not align with the normal distribution. Therefore, I conducted a randomization test to determine if answered ($M = 4.07$, $SD = 1.51$) and unanswered ($M = 4.20$, $SD = 1.72$) questions differed in the total number of distinct types
of contextual information they contained: \( p = .328 \). The results of the test show that, on average, answered and unanswered questions did not differ in the total number of contextual-information types they contained. Askers and answerers on average both provided around four contextual-information types in their questions. Therefore, on average, questions that included more contextual-information types had no greater chance of receiving an answer than those that included fewer types.

Because the number of contextual-information types did not vary significantly between answered and unanswered questions, I further explored whether the presence or absence of each contextual-information type in questions related to answerability. Whereas the initial randomization test compared whether the average number of total contextual-information types within each question differed between answered and unanswered questions, here I compared the number of questions that included each contextual-information type to assess whether the presence or absence of contextual information associated with answerability. I conducted chi-square tests to determine whether answered and unanswered questions related to the presence and absence of each contextual-information type. I also calculated Cramér’s phi for each type, which illustrated the amount of variability that the contextual-information types accounted for in answerability. As shown in table 5.1, the results of these tests varied in significance and effect size.
Table 5.1. Chi-square test results for answerability and contextual-information type

<table>
<thead>
<tr>
<th>Contextual-information type</th>
<th>Answered questions including type</th>
<th>Unanswered questions including type</th>
<th>Answered questions without type</th>
<th>Unanswered questions without type</th>
<th>N</th>
<th>df</th>
<th>Χ²</th>
<th>p</th>
<th>Φ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>181</td>
<td>189</td>
<td>69</td>
<td>61</td>
<td>500</td>
<td>1</td>
<td>.51</td>
<td>.475</td>
<td>.03</td>
</tr>
<tr>
<td>Specifications</td>
<td>141</td>
<td>150</td>
<td>109</td>
<td>100</td>
<td>500</td>
<td>1</td>
<td>.53</td>
<td>.468</td>
<td>.03</td>
</tr>
<tr>
<td>Sources</td>
<td>16</td>
<td>13</td>
<td>234</td>
<td>237</td>
<td>500</td>
<td>1</td>
<td>.15</td>
<td>.702</td>
<td>.02</td>
</tr>
<tr>
<td>Examples</td>
<td>74</td>
<td>77</td>
<td>176</td>
<td>173</td>
<td>500</td>
<td>1</td>
<td>.04</td>
<td>.845</td>
<td>.01</td>
</tr>
<tr>
<td>Frustration</td>
<td>3</td>
<td>4</td>
<td>247</td>
<td>246</td>
<td>500</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>Knowledge</td>
<td>16</td>
<td>19</td>
<td>234</td>
<td>231</td>
<td>500</td>
<td>1</td>
<td>.12</td>
<td>.726</td>
<td>.02</td>
</tr>
<tr>
<td>Thought</td>
<td>19</td>
<td>38</td>
<td>231</td>
<td>212</td>
<td>500</td>
<td>1</td>
<td>6.42</td>
<td>.011</td>
<td>.11</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>66</td>
<td>68</td>
<td>184</td>
<td>182</td>
<td>500</td>
<td>1</td>
<td>.01</td>
<td>.920</td>
<td>.00</td>
</tr>
<tr>
<td>Error</td>
<td>11</td>
<td>4</td>
<td>239</td>
<td>246</td>
<td>500</td>
<td>1</td>
<td>2.47</td>
<td>.116</td>
<td>.07</td>
</tr>
<tr>
<td><strong>Desired state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>170</td>
<td>158</td>
<td>80</td>
<td>92</td>
<td>500</td>
<td>1</td>
<td>1.07</td>
<td>.300</td>
<td>.05</td>
</tr>
<tr>
<td>Specifications</td>
<td>71</td>
<td>71</td>
<td>179</td>
<td>179</td>
<td>500</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>Examples</td>
<td>54</td>
<td>44</td>
<td>196</td>
<td>206</td>
<td>500</td>
<td>1</td>
<td>1.03</td>
<td>.311</td>
<td>.05</td>
</tr>
<tr>
<td>Motivation</td>
<td>13</td>
<td>12</td>
<td>237</td>
<td>238</td>
<td>500</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>.00</td>
</tr>
<tr>
<td>Thought</td>
<td>31</td>
<td>28</td>
<td>219</td>
<td>222</td>
<td>500</td>
<td>1</td>
<td>.08</td>
<td>.782</td>
<td>.01</td>
</tr>
<tr>
<td>Limit</td>
<td>70</td>
<td>75</td>
<td>180</td>
<td>175</td>
<td>500</td>
<td>1</td>
<td>16</td>
<td>.693</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gratitude/welcome</td>
<td>56</td>
<td>76</td>
<td>194</td>
<td>174</td>
<td>500</td>
<td>1</td>
<td>3.72</td>
<td>.053</td>
<td>.08</td>
</tr>
<tr>
<td>Other questions/comments</td>
<td>25</td>
<td>25</td>
<td>225</td>
<td>225</td>
<td>500</td>
<td>1</td>
<td>0</td>
<td>1.00</td>
<td>.00</td>
</tr>
</tbody>
</table>

The p-values in table 5.1 suggest that two possible associations exist. The first association relates to questions that included thought information related to the current state. The chi-square tests suggest that the presence or absence of thought information related to answerability, with unanswered questions more frequently including thought information. As described in chapter 4, askers provided thought information when they speculated about the causes of the problems they communicated in their questions.

Speculating about the cause of a problem could signal the complexity of the problem in question. In his study of software forums, Swarts (2015) observed that software users go online to ask about different types of tasks. First, askers post about “tamed tasks” whose solutions are clearly defined. Second, askers post about “more complex” tasks whose possible solutions are less clear (p. 168). Swarts observed that answerers posted solutions in response to complex tasks in “tentative, speculative, and conditional ways” (p. 170). In
the present study, askers wrote their thoughts about problem causes in similarly tentative and speculative ways; for example, A27 wrote, “the problem seems to relate to Word trying to access File Manager.” Askers’ use of speculative language suggests that they asked about more complex tasks. The complexity of the tasks could explain why questions that included speculative thought information received answers less frequently than those questions that did not. In their analysis of 3,000 social questions and answers, Chua and Banerjee (2015) found that questions coded as more complex were less likely to receive answers. They defined complexity as the “understandability” of questions (p. 3). In another study of social questions, Shah et al. (2012) found that 36% of the unanswered questions they analyzed exhibited characteristics of complexity, described as questions that “contain and/or demand an excessive amount of information to be exchanged, which can deter a response due to the implied effort required to provide a quality answer” (p. 5). Conceivably, askers who are less sure about causes of their problems could signal to answerers that more work will be involved in answering the askers’ questions. Yet neither Chua and Banerjee (2015) nor Shah et al. (2012) noted the presence of speculative language in the complex questions they analyzed, but inasmuch as speculative language signals task complexity as Swarts (2015) found, the findings in the present study add support for findings in Chua and Banerjee’s (2015) and Shah et al.’s (2012) studies of social questions.

The second possible association relates to questions that included gratitude/welcome information. The chi-square tests suggest that the presence of gratitude/welcome information related to answerability, with unanswered questions more frequently including gratitude/welcome information. Chua and Banerjee (2015) observed
that questions posted on the site Stack Overflow that included more politeness and subjective wording were less likely to receive answers. Yang et al. (2011) analyzed questions on Yahoo Answers! for the polite words “thanks,” “thanks,” “please,” “could,” “would,” and “help,” and found that questions containing such politeness markers were less likely to receive answers. They observed that questions that included politeness words were usually those in which the askers described “long” and “complicated or troublesome experiences” (p. 1277). In their study of social questions, Harper et al. (2008) analyzed politeness in questions based on the quality of the resulting answers. They found that politeness in questions did not alone predict answer quality, but they found some interaction effect between politeness and the site where the question originated. They concluded that each site they studied seemed to react differently to politeness. The data in the present study originate from the site Super User, and findings appear to confirm the association between lack of politeness and answerability that Chua and Banerjee (2015) found on Stack Overflow and that Yang et al. (2011) found on Yahoo Answers!

With the exception of thought information (current state) and gratitude/wELCOME information, however, the findings shown in table 5.1 suggest that only a weak relationship existed between the number of answered and unanswered questions that included the various contextual-information types. In other words, answerability does not appear to hinge upon the presence or absence of the other 14 contextual-information types shown in table 5.1.

These findings expand findings in previous research studies. In their study of contextual information in social questions, Suzuki et al. (2011) studied five groups of
contextual information that related to six contextual-information types shown in table 5.1: task, thought related to the current state, situation, knowledge, and limit. The researchers found that askers who communicated one of the five types received no more answers than askers who provided another type of contextual information. Inasmuch as the number of answers correlates with the probability of receiving any answer, the findings in the present study lend support to Suzuki et al.’s findings. Yet the findings also expand Suzuki et al.’s findings by providing seven more information types that appear to not associate with answerability. Further, in their experimental study of social questions, Harper et al. (2008) examined another type of contextual information. The researchers investigated whether askers’ communicating their own prior effort resulted in differences in answer length, perceived answerer effort, and perceived answer quality. They found that askers’ including their prior effort did not have any effect on length or perceived effort, but it did have some marginal effect on answer quality. The findings in the table 5.1 augment Harper et al.’s findings by suggesting that including prior effort appears to not associate with increased answerability.

**Magnitude**

Because of the possible association between the presence or absence of some types of contextual information and answerability, I further explored whether the number of contextual-information tokens present in questions associated with answerability. I examined Q–Q plots of both answered and unanswered questions and concluded that I could not assume normality; therefore, I completed a randomization test with 10,000 random samples to determine if answered ($M = 6.64, SD = 3.91$) and unanswered ($M = 7.47, SD = 4.54$) questions differed in the total number of contextual-information tokens
they contained: \( p = .026 \). This exploratory finding suggests that, on average, answered and unanswered questions differed in the contextual-information tokens they contained. These tokens represented individual t-units, and t-units represent the “shortest grammatically allowable sentences” (Hunt, 1965, p. 21). Therefore, unanswered questions appear to have included more sentences on average than answered questions.

To more closely examine the differences in the number of t-units answered and unanswered questions included, I analyzed contextual-information tokens by each contextual-information code and compared totals of answered and unanswered questions (see table 5.2). In the majority of cases, askers provided more tokens of contextual information in unanswered questions, with few exceptions.

**Table 5.2.** Contextual-information token totals by code and by answered and unanswered questions

<table>
<thead>
<tr>
<th>Code description</th>
<th>T-units in 250 answered questions</th>
<th>T-units in 250 unanswered questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>Mdn</td>
</tr>
<tr>
<td><strong>Current state</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>485</td>
<td>1.94</td>
</tr>
<tr>
<td>Specifications</td>
<td>260</td>
<td>1.04</td>
</tr>
<tr>
<td>Sources</td>
<td>25</td>
<td>0.10</td>
</tr>
<tr>
<td>Examples</td>
<td>108</td>
<td>0.43</td>
</tr>
<tr>
<td>Frustration</td>
<td>3</td>
<td>0.01</td>
</tr>
<tr>
<td>Knowledge</td>
<td>20</td>
<td>0.08</td>
</tr>
<tr>
<td>Thought</td>
<td>22</td>
<td>0.09</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>102</td>
<td>0.41</td>
</tr>
<tr>
<td>Error</td>
<td>14</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>Desired State</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>247</td>
<td>0.99</td>
</tr>
<tr>
<td>Specifications</td>
<td>85</td>
<td>0.34</td>
</tr>
<tr>
<td>Examples</td>
<td>59</td>
<td>0.24</td>
</tr>
<tr>
<td>Motivation</td>
<td>13</td>
<td>0.05</td>
</tr>
<tr>
<td>Thought</td>
<td>34</td>
<td>0.14</td>
</tr>
<tr>
<td>Limit</td>
<td>90</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gratitude/welcome</td>
<td>66</td>
<td>0.26</td>
</tr>
<tr>
<td>Other questions/comments</td>
<td>28</td>
<td>0.11</td>
</tr>
</tbody>
</table>
To assess the significance of these differences between answered and unanswered questions, I examined density plots of the tokens associated with each contextual-information code. I observed that each plot exhibited characteristics of right skew. To confirm my visual observations of the skewed data, I completed a Shapiro-Wilkes test of normality for each contextual-information distribution, and the results of each test suggested that I could not assume normality for the distributions of any of the information types ($p < .05$). Therefore, I completed randomization tests to determine whether contextual-information tokens within each contextual-information code differed significantly between answered and unanswered questions. Table 5.3 shows the results and $p$-value estimates from the tests.

**Table 5.3.** Results of randomization tests comparing the number of contextual-information types in answered and unanswered questions

<table>
<thead>
<tr>
<th>Code description</th>
<th>N</th>
<th>T-units in answered questions</th>
<th>M</th>
<th>SD</th>
<th>T-units in unanswered questions</th>
<th>M</th>
<th>SD</th>
<th>Estimated p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>250</td>
<td>485</td>
<td>1.94</td>
<td>2.05</td>
<td>599</td>
<td>2.40</td>
<td>2.46</td>
<td>.024</td>
</tr>
<tr>
<td>Specifications</td>
<td>250</td>
<td>260</td>
<td>1.04</td>
<td>1.33</td>
<td>274</td>
<td>1.10</td>
<td>1.34</td>
<td>.614</td>
</tr>
<tr>
<td>Sources</td>
<td>250</td>
<td>25</td>
<td>0.10</td>
<td>0.44</td>
<td>17</td>
<td>0.07</td>
<td>0.33</td>
<td>.315</td>
</tr>
<tr>
<td>Examples</td>
<td>250</td>
<td>108</td>
<td>0.43</td>
<td>0.90</td>
<td>126</td>
<td>0.50</td>
<td>1.45</td>
<td>.519</td>
</tr>
<tr>
<td>Frustration</td>
<td>250</td>
<td>3</td>
<td>0.01</td>
<td>0.11</td>
<td>4</td>
<td>0.02</td>
<td>0.13</td>
<td>.462</td>
</tr>
<tr>
<td>Knowledge</td>
<td>250</td>
<td>20</td>
<td>0.08</td>
<td>0.34</td>
<td>23</td>
<td>0.09</td>
<td>0.35</td>
<td>.611</td>
</tr>
<tr>
<td>Thought</td>
<td>250</td>
<td>22</td>
<td>0.09</td>
<td>0.32</td>
<td>46</td>
<td>0.18</td>
<td>0.46</td>
<td>.005</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>250</td>
<td>102</td>
<td>0.41</td>
<td>1.39</td>
<td>103</td>
<td>0.41</td>
<td>0.87</td>
<td>.952</td>
</tr>
<tr>
<td>Error</td>
<td>250</td>
<td>14</td>
<td>0.06</td>
<td>0.32</td>
<td>4</td>
<td>0.02</td>
<td>0.13</td>
<td>.028</td>
</tr>
<tr>
<td><strong>Desired state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>250</td>
<td>247</td>
<td>0.99</td>
<td>0.97</td>
<td>257</td>
<td>1.03</td>
<td>1.05</td>
<td>.629</td>
</tr>
<tr>
<td>Specifications</td>
<td>250</td>
<td>85</td>
<td>0.34</td>
<td>0.59</td>
<td>85</td>
<td>0.34</td>
<td>0.59</td>
<td>.944</td>
</tr>
<tr>
<td>Examples</td>
<td>250</td>
<td>59</td>
<td>0.24</td>
<td>0.48</td>
<td>57</td>
<td>0.23</td>
<td>0.57</td>
<td>.801</td>
</tr>
<tr>
<td>Motivation</td>
<td>250</td>
<td>13</td>
<td>0.05</td>
<td>0.22</td>
<td>15</td>
<td>0.06</td>
<td>0.30</td>
<td>.626</td>
</tr>
<tr>
<td>Thought</td>
<td>250</td>
<td>34</td>
<td>0.14</td>
<td>0.38</td>
<td>36</td>
<td>0.14</td>
<td>0.45</td>
<td>.745</td>
</tr>
<tr>
<td>Limit</td>
<td>250</td>
<td>90</td>
<td>0.36</td>
<td>0.66</td>
<td>102</td>
<td>0.41</td>
<td>0.72</td>
<td>.409</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gratitude/welcome</td>
<td>250</td>
<td>66</td>
<td>0.26</td>
<td>0.53</td>
<td>86</td>
<td>0.34</td>
<td>0.56</td>
<td>.093</td>
</tr>
<tr>
<td>Other questions/comments</td>
<td>250</td>
<td>28</td>
<td>0.11</td>
<td>0.35</td>
<td>34</td>
<td>0.14</td>
<td>0.45</td>
<td>.443</td>
</tr>
</tbody>
</table>
The estimated $p$-values suggest that little difference existed between the average number of contextual-information tokens askers provided in answered and unanswered questions, except in four instances: (1) thought information (current state), (2) error information, (3) situation information, and (4) gratitude/welcome information. The proportion of presence and absence of two of these contextual-information types also associated with answerability, as shown in table 5.1: thought (current-state) and gratitude/welcome information. Table 5.3 augments the findings in table 5.1: its data suggest that not only the presence and absence of these two types associated with answerability but also the magnitude of their presence or absence.

Specifically, the data in table 5.3 suggest that askers provided fewer situation tokens in answered than unanswered questions. In total, askers provided 1,661 tokens of information in answered questions and 1,868 tokens in unanswered questions. The difference between the two is 207 tokens. Notably, the difference in situation tokens accounted for 114 (55.07%) of the difference. While previous technical communication scholars theorized that situation information provided context and meaning for tasks in documentation (Mirel, 1998), these findings suggest that too much situation information can be detrimental to questions receiving answers. Answerers may perceive some situation information to be irrelevant, thus only adding to the length or complexity of the questions. Answerers may then become confused by so many irrelevant details and be unable to clearly see the asker’s task in the midst of so much information about the askers’ situations.

The data in table 5.3 also show that answered questions included more tokens of error information than unanswered questions did. In their study of 385 social questions,
Truede et al. (2011) similarly found that questions that included error messages were more likely to receive answers. Likely reasons for this difference are these: (1) error messages communicate valuable information and (2) answerers may be able to easily search for error messages on the internet. Documentation researchers have long advocated for the value of including information about error messages in documentation (Laonder & van der Meij, 1995). Practitioners have also advocated that documentation be made available on the internet to reduce the stress on technical-support agents (O’Keefe & Pringle, 2012). With potentially more organizations and more users posting technical content about error messages online, answerers may be able to easily answer questions including error messages by simply conducting a web search. Including error messages, therefore, appears to increase chances of askers receiving answers.

**Variation and efficiency**

To explore the findings from the previous two sections further, I explored whether the contextual-information type-token ratio related to answerability. The proportion of total types to tokens provides a measure of variation within the answered and unanswered data sets. Answered questions included a total of 1,661 tokens (6.64 tokens per question) of which 1,017 were types (61.23%). In contrast, unanswered questions included a total of 1,868 tokens (7.47 tokens per question) of which 1,051 were types (56.26%). I conducted a test to assess the equality of proportions. The type-token ratio in answered questions (61.23%) was higher than the type-token ratio in unanswered questions (56.26%): $z = 3.01, p = .003$.

Answered questions thus included more contextual-information variation than unanswered questions. In other words, on average answered questions included more
unique information than unanswered questions. Notably, Kitzie et al. (2013) found that higher-quality questions tended to include more unique information while concurrently expressing that unique content in fewer words. While quality was not the focus of the present study, the findings suggest that Kitzie et al.’s findings may hold true for answered and unanswered questions as well.

I further explored the type-token ratio based on contextual-information codes within individual questions. Because the type-token ratio of contextual-information codes within individual questions assesses the ratio of one specific contextual-information type to its tokens, the ratio merges distinctness and magnitude to provide a measure of how efficiently askers communicated types. To illustrate this concept of efficiency, I examined A334’s answered question below:

A334: <t1>I have 500+ doc files that I need to merge in one file.</t1> <t2>How do I do that?</t2> <t3>I have tried insert -- from a text file,</t3> <t4>but it merges only like the first 50 of them.</t4> <t5>Any ideas?</t5>

In her question, A334 provided five t-units of contextual information: situation, task, previous attempts, situation, and gratitude/welcome. Of these five t-units, four were distinct types. (Situation information repeated once.) The question then had a type-token ratio of 4 to 5 (.8), which shows the contextual-information variation of the question. However, the type-token ratio of the four distinct types was 1 (i.e., the proportion of one distinct type to one token), and the type-token ratio of the situation information was .5 (i.e., the proportion of one distinct type to two tokens), which show how efficiently the asker provided each type of contextual information. In comparison, A406’s unanswered question was even less efficient in relation to situation information and overall:
A406: <t1>I have over a hundred documents that I’ve inserted into a new master document,</t1> <t2>each subdocument has its own header which contains the document number</t2> <t3>(the numbering convention for subdocuments is A-1, A-2, B-1, C-1, C-2, etc.).</t3> <t4>When I create a TOC in the master document, it shows page numbers.</t4> <t5>How do I change the page numbers to display the document numbers instead?</t5>

In his question, A406 similarly provided five t-units of contextual information: situation, situation, situation, situation, and task. Of these five t-units, two were distinct: situation information repeated three times. Therefore, the question’s type-token ratio was 2 to 5 (.4). The task information’s type-token ratio was 1, while the situation information’s type-token ratio was .25. Overall, the unanswered question exhibited less contextual-information variation and less situation efficiency than A334’s previously discussed answered question.

Table 5.4 shows the tokens of each contextual-information type by answered and unanswered question. In addition, it shows the percentage of total contextual-information types within individual questions. The percentages in table 5.5 show that answered questions appeared to be more efficient in communicating four areas: situation information, examples related to the current state, thought information (current state), and almost all of the information related to the desired state. Answered questions tied in efficiency with unanswered questions when describing specifications (desired state). These findings suggest that askers of answered questions communicated much information about the current state and almost all of the information about the desired state more efficiently than askers of unanswered questions.
In contrast to askers of answered questions, askers of unanswered questions were more efficient in communicating sources, knowledge, previous attempts, and error messages. Notably, askers of unanswered questions communicated information related to their tasks and situations less efficiently, suggesting that they tended to more frequently include multiple tokens in the same question about what led to their current state, what their current state was, and what they wanted as a desired state.

Table 5.4. Type-token ratio for contextual information within questions

<table>
<thead>
<tr>
<th>Contextual-information type</th>
<th>Not distinct contextual-information types in answered questions</th>
<th>Distinct contextual-information types in answered questions</th>
<th>Total</th>
<th>% distinct of total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>181</td>
<td>304</td>
<td>485</td>
<td>37.32</td>
</tr>
<tr>
<td>Specifications</td>
<td>141</td>
<td>119</td>
<td>260</td>
<td>54.23</td>
</tr>
<tr>
<td>Sources</td>
<td>16</td>
<td>9</td>
<td>25</td>
<td>64.00</td>
</tr>
<tr>
<td>Examples</td>
<td>74</td>
<td>34</td>
<td>108</td>
<td>68.52</td>
</tr>
<tr>
<td>Frustration</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>100.00</td>
</tr>
<tr>
<td>Knowledge</td>
<td>16</td>
<td>4</td>
<td>20</td>
<td>80.00</td>
</tr>
<tr>
<td>Thought</td>
<td>19</td>
<td>3</td>
<td>22</td>
<td>86.36</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>66</td>
<td>36</td>
<td>102</td>
<td>64.71</td>
</tr>
<tr>
<td>Error</td>
<td>11</td>
<td>3</td>
<td>14</td>
<td>78.57</td>
</tr>
<tr>
<td><strong>Desired state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>170</td>
<td>77</td>
<td>247</td>
<td>68.83</td>
</tr>
<tr>
<td>Specifications</td>
<td>71</td>
<td>14</td>
<td>85</td>
<td>83.53</td>
</tr>
<tr>
<td>Examples</td>
<td>54</td>
<td>5</td>
<td>59</td>
<td>91.53</td>
</tr>
<tr>
<td>Motivation</td>
<td>13</td>
<td>0</td>
<td>13</td>
<td>100.00</td>
</tr>
<tr>
<td>Thought</td>
<td>31</td>
<td>3</td>
<td>34</td>
<td>91.18</td>
</tr>
<tr>
<td>Limit</td>
<td>70</td>
<td>20</td>
<td>90</td>
<td>77.78</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gratitude/welcome</td>
<td>56</td>
<td>10</td>
<td>66</td>
<td>84.85</td>
</tr>
<tr>
<td>Other questions/comments</td>
<td>25</td>
<td>3</td>
<td>28</td>
<td>89.29</td>
</tr>
</tbody>
</table>

To assess the differences in the ratios shown in Table 5.4, I conducted proportion tests to assess their equality. Table 5.5 shows the resulting z statistics and p values from these tests. The results show that the percent of examples (current-state) information in answered questions (68.52%) was higher than the percentage in unanswered questions (61.11%); the percent of situation information in answered questions (37.32%) was higher than the percentage in unanswered questions (31.55%); and the percent of task information in answered questions (68.83%) was higher than the percentage in
unanswered questions (61.48%). In each case, askers of answered questions provided the information more efficiently. Therefore, when askers of answered questions provided examples related to the current state, situation information, and task information, on average they provided the information in fewer t-units than askers of unanswered questions provided the information. In essence, askers of answered questions provided the information more efficiently.

Table 5.5. Proportion test results for distinctness and contextual-information type

<table>
<thead>
<tr>
<th>Contextual-information type</th>
<th>Distinct contextual-information types in answered questions</th>
<th>Distinct contextual-information types in unanswered questions</th>
<th>Not distinct contextual-information types in answered questions</th>
<th>Not distinct contextual-information types in unanswered questions</th>
<th>N</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>181</td>
<td>189</td>
<td>304</td>
<td>410</td>
<td>1,084</td>
<td>1.99</td>
<td>.046</td>
</tr>
<tr>
<td>Specifications</td>
<td>141</td>
<td>150</td>
<td>119</td>
<td>124</td>
<td>534</td>
<td>.12</td>
<td>.905</td>
</tr>
<tr>
<td>Sources</td>
<td>16</td>
<td>13</td>
<td>9</td>
<td>4</td>
<td>42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Examples</td>
<td>74</td>
<td>77</td>
<td>34</td>
<td>49</td>
<td>234</td>
<td>1.18</td>
<td>.238</td>
</tr>
<tr>
<td>Frustration</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td>16</td>
<td>19</td>
<td>4</td>
<td>4</td>
<td>43</td>
<td>.22</td>
<td>.826</td>
</tr>
<tr>
<td>Thought</td>
<td>19</td>
<td>38</td>
<td>3</td>
<td>8</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous attempts</td>
<td>66</td>
<td>68</td>
<td>36</td>
<td>35</td>
<td>205</td>
<td>.20</td>
<td>.843</td>
</tr>
<tr>
<td>Error</td>
<td>11</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Desired state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>170</td>
<td>158</td>
<td>77</td>
<td>99</td>
<td>504</td>
<td>1.73</td>
<td>.0837</td>
</tr>
<tr>
<td>Specifications</td>
<td>71</td>
<td>71</td>
<td>14</td>
<td>14</td>
<td>170</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Examples</td>
<td>54</td>
<td>44</td>
<td>5</td>
<td>13</td>
<td>116</td>
<td>2.13</td>
<td>.033</td>
</tr>
<tr>
<td>Motivation</td>
<td>13</td>
<td>12</td>
<td>0</td>
<td>3</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thought</td>
<td>31</td>
<td>28</td>
<td>3</td>
<td>8</td>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limit</td>
<td>70</td>
<td>75</td>
<td>20</td>
<td>27</td>
<td>192</td>
<td>.68</td>
<td>.495</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gratitude/welcome</td>
<td>56</td>
<td>76</td>
<td>10</td>
<td>10</td>
<td>152</td>
<td>.64</td>
<td>.524</td>
</tr>
<tr>
<td>Other questions/comments</td>
<td>25</td>
<td>25</td>
<td>3</td>
<td>9</td>
<td>62</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Chi-square tests assume that 2 × 2 tables will have no fewer than five items in each quadrant, so the table does not include statistics (left blank) for contextual-information types that did not meet that minimum threshold.*
**Word count**

The previous sections explored answerability based on t-unit types and tokens, and this section explores answerability based on word count. I examined Q–Q plots of both answered data ($M = 91.03, SD = 64.00$) and unanswered data ($M = 104.70, SD = 54.24$) and determined that they did not distribute normally. Therefore, I completed randomization tests on 10,000 random samples to estimate the $p$-value: 0.01. Therefore, on average answered and unanswered questions differed in word count.

While previous research reports mixed results in relation to length and answerability, the findings in this study confirm those of Chua and Banerjee (2015), who examined the word count of 3,000 answered and unanswered questions. In their hierarchical logistic regression analysis, they found that question length significantly predicted answerability, with longer questions negatively relating to answerability. However, these findings do not support the findings of Choi et al. (2013), who used machine learning to classify unanswered and answered questions and determined that word count did not contribute significantly to the model. The findings in the present study suggest that answered and unanswered questions differed in word count, with answered questions including fewer words on average.

To explore this difference further, I examined whether the average word count in each contextual-information token differed. Distribution plots showed that all data skewed right, and I examined Q–Q plots to confirm that the data did not distribute normally. Therefore, I completed randomization tests using 10,000 random samples to determine whether answered and unanswered contextual-information tokens differed in their average word counts. Table 5.6 shows that answered ($M = 1.50, SD = 5.60$) and
unanswered questions \((M = 2.66, SD = 7.98)\) differed in average number of words in thought information related to the current state \((p = .053)\). In addition, answered \((M = 7.28, SD = 15.64)\) and unanswered questions \((M = 10.40, SD = 23.98)\) differed in average number of words in examples related to the current state \((p = .082)\). In both cases, unanswered questions provided more words, suggesting that on average askers of unanswered questions provided lengthier tokens when describing examples (current state) and their thoughts (current state).

I more closely analyzed the examples (current-state) tokens, and I found that answered questions included images as examples more frequently than unanswered questions included images. Answered questions included 108 tokens of examples, and within those examples the questions included 42 images (.39 images per token). In contrast, unanswered questions included 126 tokens of examples, and within those examples the questions included 29 images (.23 images per token). Answered questions may, therefore, include fewer words because they appear to communicate their examples more frequently through images than words. As the saying goes, a picture is worth a thousand words.

**Table 5.6.** Results of randomization tests comparing the average number of words in answered and unanswered questions by contextual-information type

<table>
<thead>
<tr>
<th>Code description</th>
<th>Words in answered questions</th>
<th>(n)</th>
<th>(M)</th>
<th>(SD)</th>
<th>Words in unanswered questions</th>
<th>(n)</th>
<th>(M)</th>
<th>(SD)</th>
<th>(p)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current state</strong></td>
<td>------------------------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>-----------------------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>Situation</td>
<td>6095</td>
<td>250</td>
<td>24.38</td>
<td>27.08</td>
<td>7044</td>
<td>250</td>
<td>28.18</td>
<td>30.29</td>
<td>.140</td>
</tr>
<tr>
<td>Sources</td>
<td>411</td>
<td>250</td>
<td>1.64</td>
<td>7.66</td>
<td>255</td>
<td>250</td>
<td>1.02</td>
<td>5.00</td>
<td>.288</td>
</tr>
<tr>
<td>Examples</td>
<td>1820</td>
<td>250</td>
<td>7.28</td>
<td>15.65</td>
<td>2600</td>
<td>250</td>
<td>10.40</td>
<td>23.98</td>
<td>.082</td>
</tr>
<tr>
<td>Frustration</td>
<td>61</td>
<td>250</td>
<td>0.24</td>
<td>2.67</td>
<td>49</td>
<td>250</td>
<td>0.20</td>
<td>1.74</td>
<td>.843</td>
</tr>
<tr>
<td>Knowledge</td>
<td>254</td>
<td>250</td>
<td>1.02</td>
<td>4.29</td>
<td>342</td>
<td>250</td>
<td>1.37</td>
<td>6.09</td>
<td>.463</td>
</tr>
<tr>
<td>Thought</td>
<td>374</td>
<td>250</td>
<td>1.50</td>
<td>5.60</td>
<td>666</td>
<td>250</td>
<td>2.66</td>
<td>7.98</td>
<td>.053</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>1431</td>
<td>250</td>
<td>5.72</td>
<td>15.61</td>
<td>1415</td>
<td>250</td>
<td>5.66</td>
<td>12.71</td>
<td>.958</td>
</tr>
<tr>
<td>Error</td>
<td>245</td>
<td>250</td>
<td>0.98</td>
<td>5.06</td>
<td>108</td>
<td>250</td>
<td>0.43</td>
<td>30.29</td>
<td>.162</td>
</tr>
</tbody>
</table>
As shown in table 5.3, the number of tokens of thought (current-state) information, situation, error information, and gratitude/welcome information differed between answered and unanswered questions. Unanswered questions included more tokens of each information type with the exception of error information. Because unanswered questions included more tokens than answered questions for three of the four types and fewer tokens for one of the types, not surprisingly, table 5.6 reflects word counts congruent with the differences in types shown in table 5.3. Notably, however, the $p$-values for the differences between unanswered and answered questions differ between the two tables, with only thought (current-state) information exhibiting a fairly significant value in both tables. While the differences in $p$-values could have varied based on sample size, standard errors, and differences in means, I determined to further explore the differences by assessing the average word count of information types between answered and unanswered questions based on each contextual-information code.

When analyzing the word count of information types, I examined Q–Q plots that showed that the data did not distribute normally in all cases. As a result, I completed randomization tests using 10,000 random samples to determine whether answered and

<table>
<thead>
<tr>
<th>Desired state</th>
<th>Task</th>
<th>Specifications</th>
<th>Examples</th>
<th>Motivation</th>
<th>Thought</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3177</td>
<td>1400</td>
<td>1138</td>
<td>336</td>
<td>527</td>
<td>1446</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>12.71</td>
<td>5.60</td>
<td>4.55</td>
<td>1.34</td>
<td>2.11</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td>14.37</td>
<td>10.78</td>
<td>10.81</td>
<td>6.14</td>
<td>6.34</td>
<td>6.34</td>
</tr>
<tr>
<td></td>
<td>3540</td>
<td>1516</td>
<td>1233</td>
<td>314</td>
<td>668</td>
<td>1721</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>14.16</td>
<td>6.06</td>
<td>4.93</td>
<td>1.26</td>
<td>2.67</td>
<td>6.88</td>
</tr>
<tr>
<td></td>
<td>17.14</td>
<td>11.70</td>
<td>13.83</td>
<td>5.92</td>
<td>8.88</td>
<td>12.71</td>
</tr>
<tr>
<td></td>
<td>.304</td>
<td>.642</td>
<td>.741</td>
<td>.870</td>
<td>.426</td>
<td>.301</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Other</th>
<th>Gratitude/welcome</th>
<th>Other questions/comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>271</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>1.08</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>2.92</td>
<td>3.30</td>
</tr>
<tr>
<td></td>
<td>345</td>
<td>348</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td></td>
<td>1.38</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>3.46</td>
<td>2.84</td>
</tr>
<tr>
<td></td>
<td>.248</td>
<td>.195</td>
</tr>
</tbody>
</table>

| Total                  | 22,757             | 26,175                   |
unanswered types differed in their average word counts per contextual-information type. Notably, while table 5.6 shows that word counts do not differ significantly between the majority of the answered and unanswered question types, table 5.7 shows that answered \((M = 17.50, SD = 5.6)\) and unanswered questions \((M = 27.00, SD = 5.94)\) differed in average number of words in error information related to the current state \((p = .007)\). In addition, answered \((M = 12.57, SD = 7.19)\) and unanswered questions \((M = 11.76, SD = 6.96)\) differed in average number of words in situation information \((p = .059)\). Answered \((M = 7.50, SD = 7.93)\) and unanswered questions \((M = 10.24, SD = 6.94)\) also differed in average number of words in other questions/comments information \((p = .084)\). Finally, answered \((M = 16.85, SD = 15.46)\) and unanswered questions \((M = 20.63, SD = 18.34)\) differed in average number of words in examples related to the current state \((p = .091)\).

In three of the four cases, unanswered questions provided more words, suggesting that, on average, unanswered questions provided lengthier types related to errors, other questions/comments, and examples (current state). However, in the case of situation information, answered questions provided more words, suggesting that on average answered questions provided lengthier types that described what led to current circumstances or the current circumstances themselves. While unanswered questions included more t-units of situation information, as shown previously in 5.3, the types are shorter on average than situation types in answered questions. Askers of unanswered questions, therefore, appear to be writing situation information concisely at the word level, but they are providing more sentences of situation information overall.
In summary, the previous sections provided exploratory evidence that questions’ answerability related to the presence and absence of thought (current-state) information and gratitude/welcome information. In addition, results suggested that on average answered and unanswered questions differed in the number of sentences they contained, with unanswered questions containing more sentences related to thought (current-state) information, situation information, and gratitude/welcome information. In addition, answered questions included more sentences related to error messages.

Variation measures showed that on average answered and unanswered questions differed in the proportion of distinct information they contained, with answered questions providing proportionately more unique information. The efficiency of examples (desired state), situation information, and task information differed between answered and
unanswered questions. Answered questions were more efficient for each information type.

Finally, the results showed that answered questions included fewer words than unanswered questions. The word counts of both examples related to the current state and thought information differed between answered and unanswered questions, with answered questions providing fewer words of both information types. I also found that answered and unanswered questions did not differ in the length of t-units related to most contextual-information types. The exceptions included types related to error information, situation information, other information, and situation examples, with answered types including lower word counts in all cases except situation information.

**RQ3: Logical Coherence and Answerability**

Whereas the previous sections examined contextual information for answered and unanswered questions as a whole and by individual contextual-information types, in this second section, I examine whether answered and unanswered questions differed based on two broader categories: whether the questions included contextual information related to the current state and desired state. Building on the model that Farkas (1999) first articulated, I postulated in chapter 2 that logically coherent social how-to questions would include contextual information related to both the current and desired states. Because previous researchers have found that some unanswered questions exhibit a lack of coherence (Asaduzzaman et al., 2013; Shah et al., 2012), I proposed that answered and unanswered questions could differ in the frequency at which they include contextual information related to the two states. To explore my idea, I examined the contextual
information related to the current and desired states in the previously mentioned 500 unanswered and answered questions.

In this section, I examine whether answerability relates to the presence or absence of contextual information related to both current and desired states. I found that the presence or absence of the information does not relate to answerability. However, the analysis showed that the majority of questions included both desired and current-state information.

Because so many questions included information related to both states, I examine whether the proportion of desired and current-state information differs in unanswered and answered questions. The results show that answered and unanswered questions do not differ in their proportions of desired and current state t-units, but they do differ in their proportions of word count.

**Presence or absence of desired-and current-state information**

To examine whether answerability related to the presence or absence of contextual information related to both current and desired states, I constructed the contingency table in table 5.8 and conducted a chi-square test to determine whether answerability associated with providing or not providing contextual information related to the current and desired states: \( \chi^2(2, N = 500) = 2.41, \ p = .30 \). The results suggest that answerability did not depend on askers’ providing information related to either or both of the current and desired states.
Table 5.8. Contingency table of questions not including and including contextual information related to the current and desired states

<table>
<thead>
<tr>
<th></th>
<th>Questions with no current-state information</th>
<th>Questions with no desired-state information</th>
<th>Questions with both current- and desired-state information</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Answered</td>
<td>22</td>
<td>18</td>
<td>210</td>
<td>250</td>
</tr>
<tr>
<td>Unanswered</td>
<td>14</td>
<td>23</td>
<td>213</td>
<td>250</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>41</td>
<td>423</td>
<td>500</td>
</tr>
</tbody>
</table>

Notably, all questions in the data set included at least some contextual information related to either the current state, desired state, or both. In addition, of the 500 questions in the data set, 423 questions (84.60%) included contextual information related to both. The high frequency of questions including information related to both states suggests that a logical model for SQA questions should include current and desired states. Based on these results, I cannot conclude that following the model predicts answerability, but askers of both answered and unanswered questions clearly included information related to both states in the vast majority of questions.

**Proportion of desired- and current-state t-units and words**

Because relatively few questions did not include contextual information related to the current or desired states, I explored the proportion of desired-state information to current-state information to see if it differed between answered and unanswered questions. In essence, whereas the previous chi-square test examined the extremes in the data set of questions not providing information related to either state, in this analysis I explored whether a middle-ground approach might be more appropriate for assessing differences between answered and unanswered questions.

Answered questions included 1,567 t-units of current and desired-state information, with 528 (33.69%) being desired-state information. In contrast, unanswered
questions included 1,748 t-units of current- and desired-state information, with 552 (31.58%) being desired-state information. I conducted a test to assess differences in the proportion of desired-state information to total desired- and current-state information between answered and unanswered questions, and the proportion of desired-state information in the answered questions (33.69%) was not significantly higher than the proportion of the desired-state information in the unanswered questions (31.58%): $z = 1.30, p = .194$.

To further explore the proportion of current- and desired-state information, I examined it based on word count. Answered questions included 22,276 words of current and desired-state information, with 8,024 words (36.02%) being desired-state information. In contrast, unanswered questions included 25,482 words of current state and desired-state information, with 8,992 words (34.90%) being desired-state information. I conducted a test to assess the equality of the proportions of desired-state information to total desired- and current-state information, and the proportion of desired-state information in answered questions (36.02%) was higher than the proportion of distinct information in unanswered questions (34.90%): $z = 2.57, p = .010$.

The results suggest that while the proportion of desired-state information based on t-units did not differ between answered and unanswered questions, the proportion of words within those t-units differed. Answered questions included on average a relatively higher proportion of words relating to the desired state than unanswered questions included. Conversely, answered questions included on average a relatively lower proportion of words relating to the current state than unanswered questions included. As shown in table 5.6, situation information comprised a relatively large part of the
difference in word counts between answered and unanswered questions. Notably, current-
state information includes situation information, so situation information may be one of
the driving factors for the proportional differences between the word counts of current
and desired states.

In summary, answerability did not relate to the presence or absence of
information related to the current and desired states. However, the majority of questions
included information related to both states, suggesting that while Farkas’s (1999) logical
model does not relate to answerability, the model accurately describes the information
askers provided. Exploratory proportion tests showed that the proportion of desired-state
information in answered and unanswered questions did not differ based on total t-units,
but the proportion differed based on total words. Answered questions provided a higher
proportion of desired-state information and thus a relatively lower proportion of current-
state information than unanswered questions provided.

**RQ4: Contextual Information Answerers Requested**

In this fourth section, I transition from analyzing asker-provided contextual information
to analyzing answerer-requested contextual information: I analyze the contextual
information that answerers asked for in their comments on unanswered questions. As
discussed in chapter 2, coherence manifests both in textual microstructures and in
audience perception, so this section focuses on the contextual-information types
answerers needed to make questions more coherent according to Farkas’s (1999) logical
model. As noted in chapter 3, only unanswered questions in this study included
comments, so I examined all comments that answerers posted on the 250 unanswered
questions in the data set.
Table 5.9 shows the results of how frequently answerers made other comments and how frequently answerers requested each type of contextual information. The findings show that answerers made other comments most frequently, representing 525 of 690 (76.09%) of the t-units in answerers’ comments. Other comments represented t-units wherein answerers made no requests for contextual information, and I will discuss them in greater depth later in this chapter.

Table 5.9. Contextual information requested by commenters by t-unit totals by code

<table>
<thead>
<tr>
<th>Code description</th>
<th>Total t-units</th>
<th>% of total</th>
<th>% of subtotal</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current state</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>86</td>
<td>12.46</td>
<td>68.25</td>
</tr>
<tr>
<td>Specifications</td>
<td>15</td>
<td>2.17</td>
<td>11.90</td>
</tr>
<tr>
<td>Examples</td>
<td>10</td>
<td>1.45</td>
<td>7.94</td>
</tr>
<tr>
<td>Knowledge</td>
<td>1</td>
<td>0.14</td>
<td>0.79</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>12</td>
<td>1.74</td>
<td>9.52</td>
</tr>
<tr>
<td>Error</td>
<td>2</td>
<td>0.29</td>
<td>1.59</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td>126</td>
<td>18.26</td>
<td></td>
</tr>
<tr>
<td><strong>Desired state</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>16</td>
<td>2.32</td>
<td>41.03</td>
</tr>
<tr>
<td>Thought</td>
<td>23</td>
<td>3.33</td>
<td>58.97</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td>39</td>
<td>5.65</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other comments</td>
<td>525</td>
<td>76.09</td>
<td>100.00</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td>525</td>
<td>76.09</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>690</td>
<td>100.00</td>
<td></td>
</tr>
</tbody>
</table>

Despite other comments representing 525 of 690 (76.09%) of the t-units in answerers’ comments, answerers still requested contextual information related to the current and desired states in 165 (23.91%) t-units. Of the contextual information answerers requested most frequently, answerers requested the information in this order of frequency: situation (12.46%), thought (3.33%), task (2.32%), specifications (2.17%), previous attempts (1.74%), examples (1.45%), error (0.29%), and knowledge (0.14%). With the exception of thought information, which I will discuss later in this chapter, the frequency of information requested by answerers in table 5.9 approximately matches the relative frequency of information provided by askers in table 4.1. In table 4.1, askers
provided information in this order of frequency: situation (30.72%), specifications (15.13%), task (14.28%), examples (6.63%), previous attempts (5.81%), knowledge (1.22%), and error (0.51%). In both tables, situation information represented the highest frequency, task and specifications information represented moderately high frequency, examples and previous attempts represented moderate frequency, and knowledge and error represented low frequency. These similar frequencies suggested that inasmuch as supply equals demand, answerers valued situation, specifications, and task information most highly.

In the next two sections, I describe and discuss each type of requested information within the two states. Following that discussion, I discuss other comments in more detail.

**Requested current-state information**

Answerers requested contextual information related to the current state in 126 (18.26%) of the 690 analyzed t-units. In comparison, answerers requested 39 (5.65%) t-units related to the desired state and 525 (76.09%) t-units related to other comments. Thus of the t-units wherein askers actually requested contextual information, contextual information related to the current state dominated by almost three times. In the following sections, I detail the specific types of contextual information answerers requested that related to the current state in order of frequency.

**Situation**

Of the 126 information t-units wherein answerers requested information related to the current state, answerers most frequently requested situation information. Answerers requested situation information in 86 (68.25%) of the 126 t-units. Of the total 165 t-units related to the current and desired states wherein answerers requested contextual
information, situation t-units represented 52.12%, surpassing the next most frequent code type by five times.

Answerers requested situation information to ask for information related to the asker’s present circumstances or environment, as AN7 did when asking for formatting information: “What does your Page Number Format show in Header & Footer Tools > Design > Header & Footer?” In addition, answerers requested situation information when they asked what led to the asker’s present circumstances or environment. AN76 exemplified this type of request: “How has the document been created?” The relatively high frequency with which answerers requested situation information suggests that situation information assists answerers in providing answers.

However, previous findings in this chapter show that unanswered questions included more situation t-units than answered questions on average, and unanswered questions also included situation t-units less efficiently on average than answered questions. Despite an abundance of situation information in unanswered questions, answerers still requested situation information as the most frequent information type. At least two causes might have generated this result: (1) askers did not provide the specific situation information answerers needed to answer questions and (2) the specific questions answerers commented on provided less than average situation information.

Many of the answerers’ comments included very specific requests for situation information, such as A19’s request: “Also, did you paste the text in from another document or did you type it in?” Askers may have simply failed to provide the specific situation information that answerers needed to provide an answer. Further, answerers have specific background and knowledge that could influence the specific situation
information they need to provide answers; if the asker-supplied situation information does not match the situation information that the answerer needed, then answerers may be unable to help.

To assess whether askers provided less situation information in questions for which answerers requested more situation information, I returned to the original question data in table 5.4. The table showed that the 250 answered questions provided an average of 1.94 situation t-units, whereas the 250 unanswered questions provided an average of 2.40 situation t-units. In the 43 unanswered questions for which answerers requested additional situation information, askers provided an average of only 2.19 t-units of situation information. (When I disregarded those questions and recalculated an average for the remaining 207 unanswered questions for which answerers did not request more situation information, and the average rose to 2.44 situation t-units per question.) Therefore, askers may not have provided the right situation information in the unanswered questions for which answerers requested additional situation information, but they also provided less situation information on average when compared to the average of other unanswered questions in the sample. Answerers thus appear to request situation information more frequently when askers provided less situation information in their questions.

**Specifications**

Of the 126 information t-units wherein answerers requested information related to the current state, answerers requested specifications information second most frequently. Answerers requested specifications information in 15 (11.90%) of the 126 t-units. Of the total 165 t-units related to the current and desired states wherein answerers requested
contextual information, specifications t-units represented 9.10%, placing requests for specifications information at almost equal frequency to requests for task information (9.70%).

Answerers requested specifications information to request information related to the current state along with hardware or hardware settings (4 times), software version (4 times), file types (3 times), software (2 times), operating system (1 time), and software settings (1 time). Most frequently, answerers requested information related to hardware, as exemplified by AN41: “what’s your printer and printer driver?” In addition, answerers also most frequently asked for information about software versions, as AN33 requested: “What version of Word are you referring to?” Given how frequently new versions of Word appear on the market and the differences in capabilities between those versions, answerers apparently needed version information to troubleshoot or to provide appropriate instructions for the askers’ situations.

Previous attempts

Of the 126 information t-units wherein answerers requested information related to the current state, answerers requested specifications information third most frequently. Answerers requested previous attempts information in 12 (9.52%) of the 126 t-units. Of the total 165 t-units related to the current and desired states wherein answerers requested contextual information, previous attempts t-units represented 7.27%.

Answerers requested previous attempts information to request information related to the current state that also requested information about the asker’s past attempts to satisfactorily accomplish or research the task. Askers most frequently communicated these requests by asking about what askers had “tried” (7 times), what askers had
“checked” (2 times), whether askers had “ensured” (1 time), and whether askers had “rebooted” (1 time). Askers also frequently included the word “try” when communicating previous attempts in their questions, suggesting that both askers and answerers expect askers to try to solve their problems to at least some degree.

Examples

Of the 126 information t-units wherein answerers requested information related to the current state, answerers requested examples information fourth most frequently. Answerers requested examples in 10 (7.94%) of the 126 t-units. Of the total 165 t-units related to the current and desired states wherein answerers requested contextual information, examples t-units represented 6.10%.

Answerers requested examples information to request information related to the current state as well as screenshots (4 times), example documents (3 times), example programming code (1 time), example drawings (1 time), or examples in general (1 time). Notably, answerers requested screen shots and example documents most frequently, suggesting that answerers preferred to actually see what askers were seeing or experiencing as opposed to reading about examples. A139’s request exemplified such requests: “A screen shot would help us understand your question.”

Error

Of the 126 information t-units wherein answerers requested information related to the current state, answerers requested error information fifth most frequently. Answerers requested error information in (1.59%) of the 126 t-units. Of the total 165 t-units related to the current and desired states wherein answerers requested contextual information, error t-units represented 1.21%.
Answerers requested error information to request the error messages askers received. A129’s request exemplified this type of request: “What does the error message say?” While answerers requested error information infrequently, error information could provide answerers with valuable information that would give them insight into problems and thus enable them to provide answers.

**Knowledge**

Of the 126 information t-units wherein answerers requested information related to the current state, answerers requested knowledge information least frequently. Answerers requested knowledge information in 1 (0.79%) of the 126 t-units. Of the total 165 t-units related to the current and desired states wherein answerers requested contextual information, knowledge t-units represented 0.61%.

Answerers requested knowledge information to assess what askers knew. A44’s lone request provided the only example: “Do you know the difference.” As noted previously in the chapter, knowledge demanded by answerers matched the relative knowledge supplied by askers, suggesting that what askers know is less important than other types of contextual information.

**Requested desired-state information**

Answerers requested contextual information related to the desired state in 39 (5.65%) of the 690 analyzed t-units. In comparison, answerers requested 126 (18.26%) t-units related to the current state and 525 (76.09%) t-units related to other comments. In the following sections, I detail the specific types of contextual information answerers requested that related to the current state in order of frequency.
Thought

Of the 39 information t-units wherein answerers requested information related to the desired state, answerers requested thought information most frequently. Answerers requested thought information in 23 (58.97%) of the 39 t-units. Of the total 165 t-units related to the current and desired states wherein answerers requested contextual information, thought t-units represented 13.94%, placing thought requests second only to situation requests in overall frequency (52.12%).

Answerers requested thought information to request that askers evaluate a possible answer or proposed solution of which the answerer thought. Answerers used two strategies to request that askers evaluate a thought-of solution. First, answerers specifically asked askers if a solution worked, as A81 requested: “Does entering those words as custom fields work?” Second, answerers ended phrased their solution as a question with question mark syntax at the end, as A63’s request exemplified: “Why don’t you just use the recent document feature of Word to located the last location Word thought the file was at?” In both cases, answerers clearly intended for askers to respond back. If the solution didn’t work, then the answerers could potentially post the solution as an official answer to the question, or they could provide further solutions until askers found that one worked.

The relatively high frequency at which answerers posted these requests suggests that answerers engage in some level of answer verification before posting official answers. Such solution development and refinement was observed by Swarts (2015) in his analysis of complex tasks in forum posts. Similar to thought information that askers provided about the cause of problems, the question structure of answerers’ thought
requests indicate some level of answerers talking about solutions in “tentative, speculative, and conditional ways” (p 170). For example, even A63’s request of, “Why don’t you just use the recent document feature of Word to located the last location Word thought the file was at?” suggests that A63 recognizes that a reason might exist for the asker’s not using Word’s built-in functionality. As I discuss later in this chapter, these thought requests contrast from many of the other comments where answerers provided solutions without explicitly seeking confirmation back from the asker.

**Task**

Of the 39 information t-units wherein answerers requested information related to the desired state, answerers requested task information second most frequently. Answerers requested task information in 16 (41.03%) of the 39 t-units. Of the total 165 t-units related to the current and desired states wherein answerers requested contextual information, specifications t-units represented 9.70%, placing requests for task information as the third most frequently requested contextual information type.

Answerers requested thought information to clarify the asker’s goal (9 times), to learn what the asker’s goal was (4 times), or to assess the asker’s motives for achieving the goal (3 times).

- When asking for clarification, askers checked to make sure they understood the desired state, as AN127 did: “But just to clarify, your generated table would include one row for *each of the other tables*?” Notably, answerers requested examples in two t-units in order to clarify what the asker needed, and one asker requested clarification of specifications in relation to the goal.
• When asking about what the asker’s goal was, askers requested about the wants or needs of the asker; for example, AN51 wrote this request: “Are you trying to create a form in Word?”

• When asking about the asker’s motivation for achieving the goal, askers requested for reasons or purposes, as AN58 did: “What might be the purpose of such changes?”

Most frequently, then, answerers requested task information in order to acquire clarity regarding the asker’s goal.

**Other information**

The findings show that answerers made other comments most frequently, representing 525 of 690 (76.09%) t-units. In comparison, answerers requested contextual information related to the current state in 126 (18.26%) of the analyzed t-units and in 39 (5.65%) t-units related to the desired state. In contrast to comments related to the current and desired states, other comments did not request contextual information from askers, despite SQA sites seeking to minimize needless commentary and to focus on just questioning and answering. However, far from being needless, the commentary still communicated important information to both askers and answerers.

My informal analysis of other comments suggested that many answerers posted solutions in the comments rather than formally posting their solutions as official answers. For example, A55 suggested, “You may try reinstalling the software.” They seemed to be testing out possibly solutions by posting them in the comments and then waiting for answerers to confirm that the solution worked. Answerers similarly posted comments when they believed no solutions existed, such as when A86 posted, “I can’t really think
of an easy solution there.” Or if answerers could not duplicate the asker’s problem, they communicated their findings, as A51 did: “Can’t duplicate the issue in Word 2013.” Other answerers communicated their analysis of the problem, such as when AN25 stated, “I noticed that you’re document is in tracked changes mode.” In addition, some comments corrected askers, as AN44 exemplified: “You are describing an index as opposed to a table of contents.” While all of this information did not specifically request information back from askers, other comments clearly did not represent needless chatter on the site.

In summary, this section provided frequencies and descriptions of the contextual-information types answerers requested in comments on unanswered questions. Results show that answerers requested contextual information in approximately the same proportion that askers provided it in their questions. The majority of information in comments did not take the form of requests. In other comments, answerers frequently provided solutions to askers without posting them as official answers. Besides these other comments, answerers requested situation information most frequently. While previous findings in this chapter suggest that unanswered questions provided an overabundance of situation, the frequency of answerers’ requests for situation information suggests that answerers still need situation information to answer questions.

**Conclusion**

This chapter analyzed contextual information and answerability in social how-to questions. Results showed that answered questions contained fewer t-units and fewer words than answered questions. These results add additional support to SQA research that has found that longer questions associate more frequently with unanswered questions.
(Chua & Banerjee, 2015). Results also showed that answered questions exhibited more variation based on their type-token ratio, suggesting that answered questions included more unique content than unanswered questions. This finding augment’s Kitzie et al.’s (2013) study of social questions that found that higher-quality questions tended to include more unique information while concurrently expressing that unique content in fewer words.

Results also suggested that questions’ answerability related to the presence and absence as well as the number of tokens of thought information related to the current state. The presence of more speculative thought information in unanswered questions adds support to Swarts’s (2015) findings that more complex tasks in software forums included speculative language. The results also showed that the presence and absence as well as magnitude of more gratitude/welcome information related to answerability. More gratitude/welcome information in unanswered questions confirms previous SQA research that found that more politeness associated with less answerability (Chua & Banerjee, 2015; Yang et al., 2011). Findings also showed that answered questions included more tokens of error messages, suggesting that including specific error messages may aid answerers in providing answers.

Compared to unanswered questions, answered questions included fewer situation-information tokens. However, answered questions included slightly more words in situation tokens than unanswered questions. Therefore, situation tokens in unanswered questions were more concise at the word level, but they were more abundant at the token level. Consequently, answered questions included situation information more efficiently, meaning that on average situation tokens appeared together in questions less frequently
and in smaller numbers. Answered questions also communicated examples related to the
desired state and task information more efficiently.

Regarding word count, answered and unanswered questions did not differ in the
number of words for most contextual-information types. However, answered tokens
related to error information, situation information, other information, and examples
(current state) included fewer words on average. Notably, my informal analysis of
situation examples suggested that answered questions provided examples in the form of
images more frequently than unanswered questions, possibly allowing images to
communicate in place of words.

Variation measures showed that on average answered and unanswered questions
differed in the proportion of distinct information they contained, with answered questions
providing proportionately more unique information. The efficiency of certain contextual-
information types related to answerability: examples related to the desired state, situation
information, and task information related to answerability. Answered questions were
more efficient for each type.

The findings in this chapter showed that answerability did not relate to the
presence or absence of information related to the current and desired states. However, the
majority of questions included information related to both states. Therefore, Farkas’s
(1999) logical model appears to accurately describe the information askers provided.
Exploratory proportion tests showed that the proportion of desired-state information to
desired- and current-state information between answered and unanswered questions did
not differ based on total types. However, the proportion differed based on total words,
with answered questions providing proportionately more desired-state information and a lower proportion of situation information.

In this chapter, I also described and analyzed answerers’ requests for information in unanswered questions. My analysis showed that answerers included requests for information in only about a quarter of all t-units in the comments. Of those requests, however, answerers most frequently requested situation information, underscoring the importance of situation information in social questions. The other information types that askers requested approximately reflected the frequency at which askers provided the information in the questions in the data set.

This chapter explored answerability as it related to contextual information and logical coherence. In addition, it exemplified and described answerers’ requests for contextual information. In the final chapter, I conclude by reviewing findings, discussing limitations of the study, and providing closing thoughts.
CHAPTER VI: CONCLUSION

I began this dissertation by arguing that social questions and answers represent an emerging form of documentation. However, for SQA documentation to be complete, askers must post questions, and answerers must provide answers. To describe and analyze the information that might contribute to complete documentation, I conducted a content analysis of the contextual information that askers provided in 250 answered and 250 unanswered social how-to questions and examined differences between the two sets of questions; in addition, I analyzed the contextual information that answerers requested in the comments on the 250 unanswered questions in the data set. To conclude this study, I first review major findings related to the study’s four research questions; discuss possibilities for future research; and explore implications of the findings for askers, SQA site administrators, and documentation providers. Second, I discuss the limitations of the study. Finally, I provide closing thoughts.

RQ1: Multiple Types of Contextual Information

SQA researchers have described multiple types of contextual information in social questions without reliably coding for the presence or absence of the types. In their analysis of contextual information on 1,400 questions from Yahoo Answers!, Suzuki et al. (2011) reported five groups of contextual information with some subtypes that I described in chapter 2. Results from the present study reliably confirm the presence of five types or subtypes of information that Suzuki et al. described: situation, knowledge, thought related to the desired state, task, and limit information. In his analysis of opening messages on software forums, Steehouder (2002) observed seven types of contextual information as I described in chapter 2. The present study reliably confirms the presence
of these seven types in social how-to questions: situation, specifications, task, previous attempts, thought related to the current state, error, and previous attempts information. Askers of social how-to questions, therefore, provide numerous types of contextual information.

In addition to reliably confirming the 12 information types that previous researchers already described, this study contributed examples as an additional type of contextual information that askers provided in social how-to questions. Other research has described programmers providing code examples (Nasehi, Sillito, Maurer, & Burns, 2012), but other types of examples have not, according to my research, been described. Future research could further explore the formats and functions of examples information in social how-to questions. In this study, askers provided examples as descriptions, specific text examples, extensive text examples, duplication steps, sample files, and comparisons. Previous experimental research shows that the way in which software documentation presents screen shots—full image, partial image, or full image with instructional material—influences the learning outcomes of users (van der Meij, 2000). While the findings apply to traditional print documentation, further research might similarly explore whether the format in which askers provide examples in social questions influences answerers’ understanding of questions.

In practice, administrators of SQA sites might provide examples and the other 12 information types as writing prompts to help new askers who are unsure about how to articulate their questions. Research shows that even professional documentation writers use heuristics and other formulaic prompts to help them come up for ideas when writing (Zhang & Kitalong, 2015). My analysis of the Super User site’s question-asking user
interface shows that askers are prompted with the direction of “Provide details” and “Share your research” in relation to the types of contextual information they should provide (“How to ask,” 2017). By providing more specific contextual-information types in their directions to askers, site administrators could help askers facilitate ideas that could help them come up with ways to phrase their questions that are common to the community.

In addition to analyzing what contextual information askers provided, this study also reported how much contextual information askers provided. Technical communication researchers have argued that documentation writers must be sensitive to the users’ environments and circumstances. In her discussion of documentation for complex, computer-driven tasks, Mirel (1994) argued that “for complex work, users have to consider circumstances and contingencies on- and off-line” (p. 15). The present study showed that askers provided situation information most frequently of all contextual-information types, suggesting that askers in the study faced complex, computer-driven tasks and, therefore, considered their situations when accomplishing and seeking help on their tasks. Given users’ needs in such complex situations, Mirel prescriptively advocated for documentation writers to “bring context to the foreground of documentation” (p. 20), and the present study appears to provide support for her prescription by showing that users themselves foreground their situation information in their questions.

Yet foregrounding situation information contradicts minimalist documentation principles that Carroll (1990) articulated and that still pervade the thinking of many documentation writers today. As Carroll noted in relation to training manuals, “People seem to be more interested in action, in working on real tasks, than in reading” (p. 8).
Given this observation, he constructed minimalist guidelines that presented training and documentation materials as succinctly as possible and that highlighted tasks over situation information (Brockmann, 1990; Carroll & van der Meij, 1996). The present study does not contradict the importance of tasks since the majority of questions included task-related information in addition to situation information; however, the results suggest that within the context of online social questioning, situation information occurs most frequently and, therefore, appears to be salient to askers.

As a result, technical and professional communicators who work with social questions should be aware that the task focus of minimalism may need to be balanced with a similar focus on situation in order to accommodate the needs of askers and answerers. In their discussion of myths surrounding minimalism, Carroll and van der Meij (1996) observed that minimalism is not the panacea for all documentation needs, and thus “increased insights into the specific needs and deficiencies of users” should cause documentation creators to adapt their documentation accordingly (p. 82). The results of this study suggest that both askers and answerers need situation information. I have mentioned that askers provided situation information most frequently in their questions, yet, as I discuss more fully below, answerers requested situation most frequently in their comments. Therefore, documentation creators should not hesitate to adapt their minimalist task orientation to take into account askers’ and answerers’ needs for increased situation information in social how-to questions.

**RQ2: Question Content and Answerability**

One of the recurring questions in SQA research is how the length of questions relates to answerability. As I described in chapter 2, researchers generally agree that length of
questions predicts and correlates with answerability (Li et al., 2012; Saha et al., 2013; Shah et al., 2012); however, researchers disagree on how whether more or less information is best (Agichtein et al., 2008; Asaduzzaman et al., 2013; Chua & Banerjee, 2015; Yang et al., 2011). The results of the present study showed that answered questions included fewer t-units and fewer words, providing support for previous research that suggested that answered questions tend to be shorter on average (Chua & Banerjee, 2015).

Yet one of the major contributions of this study is that it did not stop at examining how much content askers provided, but it also analyzed what content askers provided. Findings showed that no difference existed in the average number of types of content that answered and unanswered questions included. However, additional exploratory analysis showed that answered and unanswered questions differed in their type-token ratios. This measure constituted a measure of how varied the contextual information was that askers provided in their questions. In their study of unanswered questions, Choi et al. (2013) used an inverse document-frequency (idf) score, or a ratio of a word’s frequency in their corpus to the frequency of individual documents containing the word, to assess the uniqueness of words in questions. The idf score significantly contributed to the researchers’ prediction of unanswered questions, suggesting that the 200 answered and 200 unanswered questions in their study differed based on their amount of unique words. While proportional differences in my study examined uniqueness based on the amount of t-units and not words, the findings in the present study appear to support Choi et al.’s (2013) findings that unique content, whether in words or t-units, associates with answerability.
To increase answerability, therefore, askers of social how-to questions should attempt to limit the length of their questions in both word and sentence count. While limiting the length of their questions, they should also seek to provide varying types of information that shed differing and unique perspectives on their questions. In Kitzie et al.’s (2013) study of question quality and answerability, experts revised 129 low-quality, unanswered questions, and then researchers compared the original questions to the new questions. They found that the revised questions tended to include more unique information while concurrently expressing that unique content in fewer words. Thus askers could increase the quality of their questions by limiting length and increasing unique content, while also increasing their chances of receiving answers.

The previous findings related to the content in questions as a whole, yet I also divided the whole content into individual contextual-information types and found that answerability varied based on these types. In the remainder of this first section, I examine three characteristics of unanswered questions that my analysis of contextual information provided.

**Complex**

Citing Mirel (1998) earlier in this chapter, I discussed reasons why situation information is salient for askers who are completing computer-driven tasks. Findings of this study supported her claims based on the high frequency of situation information in all questions. Additional findings provided further insight, however, as askers of both answered and unanswered questions included situation information most frequently of all contextual-information types. In addition, findings showed that answered and unanswered questions on average did not differ in the presence or absence of situation
information. Therefore, situation information appears to be salient for askers of both answered and unanswered questions.

However, findings suggested that providing more situation information decreased the likelihood of questions receiving answers: answered questions included fewer situation tokens than unanswered questions. Notably, situation tokens in answered questions included more words on average than situation tokens in unanswered questions, showing that answered questions were more concise in relation to sentences but not words. One possible reason for askers of answered questions to provide fewer situation tokens is that their situations were simply less complex and, thus, communicating simpler situations required fewer tokens. Further research might explore this finding by determining ways to evaluate task complexity and then analyzing the distribution of situation information in the resulting questions.

Experimental research could also prime askers with predefined, complex tasks and then analyze the resulting questions that askers compose. For example, Robinson (2001) presented simple and complex speaking tasks to non-native English speakers in Japan and evaluated their resulting utterances. The researcher broke utterances into c-units, similar to the t-units used in the present study, and found that speakers of simple tasks included higher lexical variation as measured by type-token ratio and included more words per c-unit. The present study found that answered questions included a similarly higher type-token ratio than unanswered questions at the situation t-unit level, suggesting that task complexity could be influencing the way askers write questions and, therefore, the questions’ answerability. The present study also showed that on average answered questions included more words in their situation tokens, again suggesting that answered
questions may have described simpler tasks. Clearly, Robinson’s study and the present study differ in the mode of communication, so future research might more extensively explore the relationships between task complexity, askers’ lexical variation, and answerability.

Because of the uncertain nature of complex problems and the need for users to work out that uncertainty through back-and-forth communication with other users, Swarts (2015) argued that forums are a possible venue for users who are seeking help on complex problems. He observed that many forum users who provided solutions did so in speculative and tentative ways. Similar hesitant communication appeared in the present study as askers expressed their thoughts about possible causes of their software problems. Thought information was present in more unanswered questions than answered questions and also manifested itself in more t-units and in more words in unanswered questions. Given Swarts’s observation that users write using speculative language when discussing complex tasks, these findings provide additional support that askers of unanswered questions faced more complex tasks than askers of answered questions. Future research exploring task complexity and lexical variation might also analyze askers’ use of speculative language in simple and complex tasks.

The findings suggest that in practice askers may need to provide fewer thought t-units and fewer situation t-units in their questions in order to facilitate an increased likelihood of receiving an answer. Inasmuch as askers’ questions are more complex and demand a lengthy description of the situation or the asker’s thoughts, askers may need to seek alternative methods for receiving answers, such as the user forums Swarts (2015) described. Because users of forums do not hold the same expectations of minimizing
back-and-forth communication that users of SQA sites hold, forums may facilitate greater answerability of complex questions than SQA sites.

**Gratitude filled**

Previous research in SQA suggests that posting expressions of gratitude in social questions associates with lower answerability (Chua & Banerjee, 2015; Yang et al., 2011). Confirming these studies, findings of the present study showed that a higher proportion of unanswered questions than answered questions included gratitude, welcoming information, and the politeness strategy of indirectness. Unanswered questions also included more gratitude/welcome t-units than answered questions included. These findings clearly show that answered questions were less likely to include gratitude/welcome information, yet the reason why they were less likely to receive answers remains unclear. Three possible reasons could exist for this difference.

First, Yang et al.’s (2011) informal analysis of politeness in social questions suggested that polite language could be associated with more complex tasks, providing additional support for the possibility that unanswered questions described more complex tasks. When askers post complex questions that will take significant time of answerers to respond, askers may feel a greater need to express gratitude for askers’ help. Future research exploring the relationship between task complexity and lexical variation could explore whether askers provide gratitude more frequently when communicating complex tasks.

Second, Harper et al.’s (2008) findings suggested that users of each SQA site might react differently to language expressing gratitude. Various SQA sites might develop norms that, if violated, could result in askers not receiving answers. However,
simply providing gratitude information seems unlikely to directly influence answerability and, therefore, gratitude probably signals some other question characteristic that is influencing answerability. For example, not providing gratitude/welcome information could be a norm of the Super User site. New askers may not be aware of this norm and may violate it when they post their first questions. As askers spend more time on the site posting questions and answers, they would likely learn the norms and expectations of the Super User community. As a result, over time these askers would be less likely to include gratitude/welcome information and would also adapt their questions in other ways that better conformed to community expectations. Therefore, answerability may not necessarily hinge solely upon the presence of gratitude/welcome information, but also upon the level of experience of the asker or upon other question characteristics that gratitude/welcome information might signal. The present study did not analyze askers’ experience with the Super User site, so additional studies might explore whether askers new to the site are more likely to post gratitude/welcome information.

Third, Mackiewicz and Riley (2003) suggested that overly polite language, such as the use of indirectness, could obscure the clarity of editors’ directives to authors. Similar to editors’ directives, askers’ requests for answers attempted to motivate answerers to act by providing answers. Yet some askers communicated these requests indirectly, and gratitude/welcome information included these indirect requests since indirectness is a type of politeness. Yet these polite, indirect requests could have obscured the clarity of the questions, resulting in answerers who were unable to understand the questions well enough to provide answers. Future research could more
closely analyze politeness in the discourse of answered and unanswered questions to investigate the relationship between indirectness and answerability.

Given these three possible reasons, new askers on SQA sites might benefit from analyzing other askers’ questions to learn whether expressions of gratitude and politeness are welcomed or expected. By analyzing the way experienced askers post questions, new askers would be able to better formulate their questions in ways accepted by the community and that would be clearer to the community. For the Super User site specifically, the findings in the present study suggest that askers may see an increased likelihood of answerability if they do not include gratitude/welcome information in their questions.

**Example filled**

As I discussed above, examples constituted a new type of contextual information in social questions that previous researchers have not yet described. While examples related to the current state constituted only 6.63% of all t-units askers provided and examples related to the desired state constituted only 3.29% of all t-units askers provided, these t-units appeared in 30.20% and 19.60% of questions, respectively. Additional findings showed that unanswered questions provided examples related to the desired state less efficiently, suggesting that unanswered questions communicated the examples in larger groups of tokens than askers of answered questions. In addition, results show that unanswered questions provided example types related to the current state in fewer words on average than answered questions. These findings suggest that unanswered questions communicated examples less efficiently and less concisely than answered questions.
My additional informal analysis explored whether answered and unanswered questions differed in the format in which they provided examples. My findings suggested that answered questions included more images than unanswered questions. For example, when providing examples related to the desired state, askers of unanswered questions included 5 images in 57 questions (.09 examples per t-unit), whereas askers of answered questions included 19 images in 59 questions (.32 examples per t-unit). When providing examples related to the current state, askers of unanswered questions included 29 images in 126 t-units (.23 examples per t-unit), and askers of answered questions included 42 images in 108 t-units (.39 examples per t-unit). Askers’ use of images to provide examples might, therefore, explain much of the difference in efficiency and word count between answered and unanswered questions. Additional research could verify the results of my informal analysis to further explore the influence of examples on answerability.

When posting questions, askers could see an increased likelihood of answerability if they communicate answers more efficiently and in fewer words. One way to achieve increased efficiency and conciseness could be to use images to communicate examples. Van der Meij (2000) showed that different types of images can produce different learning outcomes in users of documentation, however, so further research might explore whether different types of image examples influence answerability most positively.

RQ3: Logical Coherence and Answerability

Farkas (1999) argued that all documentation followed the same logical structure, made up of states and actions. In chapter 2, I described how askers provided current and desired states in their questions, and then answerers posted answers to explain how to bridge the gap between the states. I therefore asserted that questions that did not include current or
desired states might break down the coherence of Farkas’s logical structure, resulting in less answerability. However, I found no association between answerability and the presence or absence of desired- and current-state information in questions. In addition, answered and unanswered questions did not differ in the proportion of tokens relating to the current and desired states. The proportions of words related to the current and desired states did differ between answered and unanswered questions, with unanswered questions providing proportionately more current-state information and less desired-state information than answered questions.

While the findings show that Farkas’s (1999) logical structure does not necessarily relate to answerability, they do suggest that his logical model accurately describes the contextual information that askers provided in social how-to questions. The findings showed that the majority of questions included both desired- and current-state information. Of the 500 questions in the study, 423 (84.60%) included contextual information related to both the current and desired states. Answered and unanswered questions both included similar proportions, with 210 of 250 (84.00%) answered questions and 213 of 250 (85.20%) of unanswered questions including information related to both states. Notably, in both answered and unanswered questions, approximately one-third of t-units related to the desired state and two-thirds of t-units related to the current state. Given the high and similar percentages of both answered and unanswered questions that included information related to both states, Farkas’s (1999) logical model appears to accurately describe the general structure of information within social questions.
Within this general structure, the location of specific contextual-information types also provides additional insight into the structure of social questions. Additional exploratory analysis suggests that the various contextual-information types appear most frequently at the points shown in table 6.1.

**Table 6.1. Contextual-information types by percentage in t-unit location**

<table>
<thead>
<tr>
<th>Contextual-information type</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Current state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Situation</td>
<td>8.76</td>
<td>16.24</td>
<td>15.87</td>
<td>13.28</td>
<td>10.79</td>
<td>8.12</td>
<td>5.72</td>
<td>5.72</td>
<td>3.60</td>
</tr>
<tr>
<td>Specifications</td>
<td>37.83</td>
<td>17.79</td>
<td>9.36</td>
<td>8.61</td>
<td>5.43</td>
<td>3.93</td>
<td>4.49</td>
<td>2.81</td>
<td>1.69</td>
</tr>
<tr>
<td>Sources</td>
<td>11.90</td>
<td>4.76</td>
<td>9.52</td>
<td>9.52</td>
<td>14.29</td>
<td>9.52</td>
<td>9.52</td>
<td>9.52</td>
<td>4.76</td>
</tr>
<tr>
<td>Frusturation</td>
<td>57.14</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>14.29</td>
<td>0.00</td>
<td>0.00</td>
<td>14.29</td>
<td>0.00</td>
</tr>
<tr>
<td>Knowledge</td>
<td>11.63</td>
<td>6.98</td>
<td>16.28</td>
<td>16.28</td>
<td>11.63</td>
<td>6.98</td>
<td>4.65</td>
<td>6.98</td>
<td>4.65</td>
</tr>
<tr>
<td>Thought</td>
<td>5.88</td>
<td>7.35</td>
<td>8.82</td>
<td>10.29</td>
<td>14.71</td>
<td>7.35</td>
<td>8.82</td>
<td>10.29</td>
<td>2.94</td>
</tr>
<tr>
<td>Previous attempts</td>
<td>2.44</td>
<td>8.29</td>
<td>11.22</td>
<td>10.24</td>
<td>11.22</td>
<td>12.20</td>
<td>11.71</td>
<td>5.37</td>
<td>4.88</td>
</tr>
<tr>
<td>Error</td>
<td>16.67</td>
<td>16.67</td>
<td>16.67</td>
<td>16.67</td>
<td>0.00</td>
<td>16.67</td>
<td>16.67</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>Desired state</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>9.52</td>
<td>18.45</td>
<td>15.48</td>
<td>11.51</td>
<td>10.91</td>
<td>10.12</td>
<td>7.14</td>
<td>4.96</td>
<td>2.38</td>
</tr>
<tr>
<td>Specifications</td>
<td>35.29</td>
<td>17.06</td>
<td>10.59</td>
<td>7.06</td>
<td>6.47</td>
<td>5.88</td>
<td>4.71</td>
<td>1.76</td>
<td>4.71</td>
</tr>
<tr>
<td>Examples</td>
<td>18.10</td>
<td>16.38</td>
<td>18.97</td>
<td>13.79</td>
<td>12.93</td>
<td>7.76</td>
<td>1.72</td>
<td>2.59</td>
<td>2.59</td>
</tr>
<tr>
<td>Motivation</td>
<td>3.57</td>
<td>17.86</td>
<td>32.14</td>
<td>10.71</td>
<td>17.86</td>
<td>10.71</td>
<td>3.57</td>
<td>0.00</td>
<td>3.57</td>
</tr>
<tr>
<td>Thought</td>
<td>1.43</td>
<td>4.29</td>
<td>7.14</td>
<td>14.29</td>
<td>8.57</td>
<td>10.00</td>
<td>7.14</td>
<td>10.00</td>
<td>7.14</td>
</tr>
<tr>
<td>Limit</td>
<td>4.69</td>
<td>8.85</td>
<td>12.50</td>
<td>11.46</td>
<td>11.46</td>
<td>12.50</td>
<td>8.33</td>
<td>7.29</td>
<td>8.85</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gratitude/welcome</td>
<td>1.97</td>
<td>0.66</td>
<td>4.61</td>
<td>9.87</td>
<td>10.53</td>
<td>12.50</td>
<td>13.16</td>
<td>9.87</td>
<td>9.21</td>
</tr>
<tr>
<td>Other questions/comments</td>
<td>1.61</td>
<td>3.23</td>
<td>9.68</td>
<td>11.29</td>
<td>9.68</td>
<td>8.06</td>
<td>4.84</td>
<td>6.45</td>
<td>11.29</td>
</tr>
</tbody>
</table>

Note: This table does not take into account the diminishing likelihood of t-units appearing in each location as a result of varying question length.

The percentage data in table 6.1 suggest that askers frequently provided contextual information in these ways:

- When providing specifications information, askers most frequently included it as the first t-unit.
- With specifications established, askers then provided more general situation and task information in the second t-unit.
• With specifications, task, and situation information provided, askers then provided additional contextual information in the forms of examples, knowledge, error messages, motivation, and limits.

• Askers then provided additional contextual information in the forms of sources and thought, followed by previous attempts and more limits.

• Finally, toward the end of questions, answers provided gratitude/welcome information.

A231 provided an example of this pattern:

A231: <t1>Word shows me a dashed border around every paragraph which is pretty confusing.</t1><t2> It is definitely not a border, but rather a border that marks the dimensions of the paragraph.</t2><t3> It looks like this: https://i.stack.imgur.com/XzDGj.png</t4><t5> Does anyone have an idea of how to remove this dashed border?</t5><t6> Thanks!</t6>

In t1, A231 provided his specifications, Word, and other information related to his current state. With specifications established, he then elaborated on general situation information in t2. To more concretely communicate the situation in t4, he included an example image of what he saw on his computer. With the current state fully described, he then expressed his desired state in t5 by stating his task of removing the unwanted border. Finally, to close, he expressed gratitude in t6. Notably, out of 500 questions, this pattern of specifications, task or situation, and examples occurred in 24 (4.80%) questions in the data set. A similar pattern of specifications, task or situation, and motivation or limits occurred in 21 (4.20%) questions. In general, askers appear to write their questions from specific to more general, ensuring that they communicate both situation and task
information early in the question. In this way, askers communicate both the current and desired states within the first few t-units of their questions.

To further explore the patterns in table 6.1, future research could examine how-to questions on other SQA sites and compare the resulting patterns in the contextual information. These future studies of other sites could similarly examine whether Farkas’s (1999) logical structure appears to accurately model the contextual information askers provide. By examining these patterns, researchers would help askers better understand how to ask social questions in ways common to the conventions of the communities that exist on these sites.

In practice, askers might begin to follow Farkas’s (1999) logical model when constructing their questions. To follow the pattern set by askers in the present study, askers would provide contextual information in relation to the present state in roughly two-thirds of their t-units and contextual information in relation to the desired state in roughly one-third of their t-units. They would provide specifications early in the question along with descriptions of their situations and tasks. Following these descriptions, they would then provide additional contextual information as shown in table 6.1.

**RQ4: Comments**

Findings showed that answerers requested contextual information in this order of frequency: situation, specifications, previous attempts, examples, error, and knowledge. Notably, askers provided the information in the same relative frequency as answerers requested it. While this finding might suggest that supply meets demand, the presence of comments on unanswered questions suggests that some askers are not supplying the information answerers need in order to provide answers.
Askers’ ability to communicate salient situation information could explain why askers of answered questions provided fewer t-units of situation information, and why askers of unanswered questions provided more t-units of situation information. Albers (2003) argued that documentation providers need to consider users’ knowledge, users’ ability to comprehend information, and the level of detail users want. He noted that for complex situations, “rather than a lack of information, the failure to anticipate people’s needs forms the basis of most information problems” (Albers, 2012, p. 174). While the present analysis of answerers’ comments provides some insight regarding the specific information needs of answerers, additional research might more fully investigate answerers’ needs through other methods to help askers determine which situation information is requisite to answer specific questions.

For example, Yang et al. (2014) analyzed edits to social questions that immediately preceded an answer. They called these “important edits” since the edits apparently provided the information an answerer needed in order to provide an answer (p. 181). When they analyzed these important edits, the researchers found that they cascaded into these categories: source code refinement, context, hardware and software details, examples, problem clarification, previous attempts, solutions, and formatting. Notably, the comments in the present study requested information related to many important edits, including situation, specifications, previous attempts, and examples information. Additional research might more fully explore the information answerers find salient so that askers can make better decisions about what information to include in their questions.
Yet equally important to analyzing what contextual information answerers requested is analyzing what contextual information answerers did not request. Answerers did not request thought information related to the current state. As I discussed previously in this chapter, unanswered questions included a higher frequency of thought information than answered questions, and the analysis of comments seems to confirm that answerers do not find salient information related to askers providing their thoughts about causes of their problems.

In practice, askers might benefit from carefully considering the needs of answerers to determine what information would be most salient. As I discussed in chapter 3, Traxler and Gernsbacher (1995) found in their experimental study that writers who received feedback from their audiences wrote more coherently. This analysis suggested that askers provided feedback related to situation, specifications, previous attempts, and examples most frequently, possibly signaling the salience of this contextual information. Notably, Yang et al.’s (2014) study corroborated these findings. Askers, therefore, should carefully consider including these information types in their social how-to questions.

**Limitations**

This study focused solely on how-to questions, so findings may not be applicable to other types of questions. However, given the large number of social questions that askers post online, these findings may apply to a large number of social questions. Additional research could also test for the presence of these contextual-information types on different types of questions. In their study of social questions, Harper et al. (2009) coded questions broadly as either information seeking or conversation seeking, and social how-to questions are a subtype of information-seeking questions. Future research could
explore the types of contextual information askers provide in conversational questions and compare them with the findings of the present study.

In order to achieve mutual exclusivity in the codebook related to contextual information that askers provided, I had to define codes in ways that would sacrifice some precision. For example, whereas the specifications code included information related to only specifications and situation, the frustration code could or could not include specifications information. In one attempt to quantify how much precision I lost by defining the codes this way, I observed that in the final study I coded 42 out of 2539 t-units (1.65%) as including sources information. To assess how many other codes could have included sources that were not coded as sources, I completed a search based on the abbreviation “http” that would have been included in any online source that askers included. Note that this search would not have included references of sources not included as web addresses. The search returned 49 instances that did not relate to linked images, which showed that at least 7 (16.66%) more sources existed in the data set than I captured in my codes. Thus this lack of precision likely underestimated a number of codes, predominantly sources, specifications, and examples codes. Future studies could focus in on these codes specifically to more accurately portray their presence in and relationship to answered and unanswered questions.

This study also ignored any contextual information provided in question titles. When answerers view social questions, they generally first read the title of the question and then click on it to learn additional contextual information. Because this study’s focus was on contextual information, I decided not to code contextual information that might appear in question titles. However, by not coding this information, I may have excluded
some questions as not how-to questions when askers’ indicated the how-to nature of the question in the question title. Further, these titles may have included some contextual information, such as specifications, that I did not take into account in my coding. Although the bulk of contextual information appears in the question body that I analyzed for this study, additional studies might consider any contextual information that askers provide in their question titles.

**Closing Thoughts**

In the introduction of this dissertation, I observed that for knowledge workers who depend on answers to questions to complete their professional work, understanding how to write social how-to questions to encourage answers is critical. This dissertation highlighted specific differences between the answered and unanswered questions in the data set that may be helpful to these workers: answered questions included fewer contextual-information tokens, answered questions included fewer words, and answered questions had a higher type-token ratio. In addition, while the study did not produce evidence suggesting that askers’ adhering to Farkas’s (1999) logical structure would see an increase in answerability, the structure appeared to accurately model the contextual information in a majority of questions.

Yet my analysis of specific contextual-information types provided a more nuanced view. For example, askers’ contextual information provided evidence of varying levels of task complexity that future research needs to explore. As a result, askers may need to consider the complexity of their questions and post simpler questions to SQA sites and more complex questions to forums. While the presence of gratitude/welcome information was lower in answered questions, answerers’ acceptance of gratitude might
vary based on the site where askers post their questions. Askers may benefit from analyzing others’ questions on SQA sites to see what the accepted norms are concerning gratitude. Although askers of answered questions included situation information more efficiently, answerers’ follow-up comments requesting even more situation information suggested that askers need to carefully weigh the salience of the situation information they provide. In summary, this study of contextual-information types elevates the importance of context.
REFERENCES


http://www.jstor.org/stable/43090319


doi: 10.1177/1050651909333141


APPENDIX A: SQL CODE FOR DATA DOWNLOAD

SELECT
    p1.Id as QuestionID,
    p1.Title as QuestionTitle,
    p1.Body as QuestionText,
    p1.CreationDate as QuestionDate,
    p1.OwnerUserId as QOwnerID,
    p1.LastEditDate as Edited,
    p1.Score as QScore,
    p1.ViewCount as QViewCount,
    p1.Tags as QTags,
    p1.AnswerCount as QAnswerCount,
    p1.CommentCount as QCommentCount,
    p1.FavoriteCount as QFavoriteCount,
    p1.ClosedDate as QClosedDate,
    p2.ParentId,
    p2.Id,
    p2.CreationDate as AnswerDate,
    c1.PostId,
    c1.Id as CommentID,
    c1.Text as CommentText,
    c1.CreationDate as CommentDate,
    c1.UserID as CommentUserID,
    u1.Id as userID,
    u1.AccountId as accountId,
    u1.Reputation,
    u1.CreationDate,
    u1.UpVotes,
    u1.DownVotes,
    u1.Views
FROM Posts p1
LEFT JOIN Posts p2
    ON p1.Id=p2.ParentId
LEFT JOIN Comments c1
    ON p1.Id=c1.PostId
LEFT JOIN Users u1
    ON p1.OwnerUserID=u1.AccountId
WHERE p1.PostTypeId=1
    AND p1.CreationDate < '2016-10-01'
    AND p1.CreationDate > '2013-03-31'
    AND p1.Tags LIKE '%microsoft-word%'
APPENDIX B: RANDOMIZATION UNDER CONDITIONS OF HETEROGENEOUS VARIANCES

Randomization tests assume that data is independent and that an experimenter randomly assigned the data to treatment and control groups (Edgington & Onghena, 2007). As I noted in chapter 3, I randomly sampled data to increase the likelihood of independence in the study’s data set. Because my study is not a traditional experiment where random assignment would be within my control, I assumed random assignment of questions into answered and unanswered categories. While my use of a randomization test meets necessary assumptions (Edgington & Onghena, 2007), previous research has suggested that permutation tests (of which randomization tests are a type) could underestimate \( p \)-values when variances of group samples are heterogeneous (Huang, Zhu, Calian, & Hsu, 2006). Therefore, I determined to explore whether unequal variance significantly altered the results and conclusions of the study.

I used an \( F \) test to compare the variances in the 54 randomization tests that I described in chapter 5. In 25 (46.30%) of 54 tests, the results suggested that the variances between samples were similar enough that I could not reject the null hypothesis that they were different: \( p > .05 \). Therefore, almost half of the randomization tests in the study exhibited homogenous variance and did not require transformation to stabilize variance.

For the remaining 29 (53.70%) of 54 tests, I transformed sample data using a square-root transformation that is recommended for stabilizing variance in count data (Maingonald & Braun, 2010). Of these 29 tests, 22 had estimated \( p \)-values from untransformed data that were greater than .10. Of the 22 tests, the square-root
transformation stabilized the data’s variance for 14, which I assessed with $F$ tests: $p > .05$. Using transformed data, I again conducted the 14 randomization tests with 10,000 samples, and the 14 tests returned $p$-values greater than .10. Thus stabilizing the variance of these 14 tests did not meaningfully alter the results and conclusions of the study.

For 8 of the 22 tests with $p$-values greater than .10, the square-root transformation did not stabilize their variance, which I assessed with $F$ tests: $p < .05$. These tests included differences in the average number of contextual-information types between answered and unanswered questions for the sources (current-state) and frustration (current-state) types, with study results showing in table 5.3. In addition, these tests included differences in average word count between answered and unanswered questions for the thought (desired-state), sources (current-state), frustration (current-state), knowledge (current-state), error (current-state), and other questions/comments types, with study results showing in table 5.6. Previous research suggests that heterogeneous variance could cause randomization tests to underestimate $p$-values for these types, yet all 8 types returned $p$-values greater than .10 when I used untransformed data. Results and conclusions of all exploratory tests in the study focused primarily on tests that returned $p$-values less than .10; therefore, the heterogeneous variance exhibited in these tests did not consequentially impact the results and conclusions in the study.

Of the 29 tests whose data exhibited heterogeneous variance, only 7 returned $p$-values from untransformed data that were less than .10. For 4 of these tests, a square-root transformation did not stabilize the variance of their sample data. These tests included differences in the total tokens between answered and unanswered questions of the thought (current-state) and error (current-state) contextual-information types as shown in
table 5.3. In addition, these tests included differences in average word count between answered and unanswered questions of the examples (current-state) and thought (current state) contextual-information types as shown in table 5.6. Because the square-root transformation did not stabilize the variance in these four tests, the heterogeneous variance present in the untransformed data could have underestimated their $p$-values.

While I discussed results from these tests in chapter 5, major conclusions in the study do not hinge exclusively on the results from these tests.

For the remaining 3 tests, square-root transformations stabilized the variance, which I assessed with $F$ tests: $p > .05$. The first of these tests included the difference in average number of tokens between answered and unanswered questions. Using transformed data, I conducted the randomization test with 10,000 samples: $p = .036$. The $p$-value from the untransformed data was .026, suggesting that the test with heterogeneous variance underestimated the $p$-value in this case. The second of these tests included the difference in average number of tokens between answered and unanswered questions for the situation contextual-information type. Using transformed data, I conducted the randomization test with 10,000 samples: $p = .055$. The $p$-value from the untransformed data was .024, as shown in table 5.3, suggesting that the test with heterogeneous variance underestimated the $p$-value in this case as well. The third of these tests included the difference in average word count between answered and unanswered questions. Using transformed data, I conducted the randomization test with 10,000 samples: $p = .007$. The $p$-value from the untransformed data was .01, suggesting that the test with heterogeneous variance overestimated the $p$-value in this case. Notably, while Huang et al. (2006) generally observed that heterogeneous variance underestimated $p$-
values, they did observe overestimation in one slightly skewed set of data. Yet despite the differences between the $p$-values of untransformed and transformed data in this study, the underestimate and overestimates appear to be small and do not consequentially alter the results or conclusions of the study.