More than Words in Medical Question-and-Answer Sites: A Content-Context Congruence Perspective

Chih-Hung Peng
College of Commerce
National Chengchi University
Taipei 11605, Taiwan
chpeng@nccu.edu.tw

Dezhi Yin
Muma College of Business
University of South Florida
Tampa, Florida 33620
dezhiyin@usf.edu

Han Zhang
Scheller College of Business
Georgia Institute of Technology
Atlanta, Georgia 30308
han.zhang@scheller.gatech.edu

Abstract: People are increasingly searching for health information from online sources such as question-and-answer (Q&A) services. The health information obtained this way can have significant impacts on people's health decisions and lives, and a natural question is: what constitutes a helpful answer in the medical domain? While prior studies examining antecedents of readers' helpfulness evaluation of answers have focused primarily on the independent impacts of content and source characteristics, we propose a content-context congruence perspective with a focus on the role of congruence between an answer's content and the answer's contextual cues. Specifically, we identify two types of contextual cues critical in the unique setting of medical Q&A sites—the language attributes (i.e., concreteness and emotional intensity) of the *question*'s content, and the acuteness of the *disease* to which the question is related. Building on the priming literature and construal-level theory, we hypothesize that an answer will be perceived more helpful if the language attributes of the answer's content are congruent with those of the preceding question, and if they are congruent with the disease's acuteness. Analyses of a unique data set from WebMD Answers provide empirical evidence for our theoretical model. This research deepens our understanding of readers' value judgment of online medical information, demonstrates the importance of considering the congruence of content with contextual cues, and opens up exciting opportunities for future research to explore the role of content-context congruence in all varieties of user-generated content. Our findings also provide direct practical implications for knowledge contributors and Q&A sites.

Keywords: medical Q&A, answer helpfulness, user-generated content, content-context congruence, fit, concreteness, emotional intensity, construal level

***** *Forthcoming at Information Systems Research* *****

INTRODUCTION

As the Internet enables people to easily access health information (Erdem and Harrison-Walker 2006; Kivits 2006), the general public increasingly turns to the Internet to seek help and advice over health concerns, medical questions and doctors' suggestions (Agarwal et al. 2010; Fichman et al. 2011). According to the "The Great American Search for Healthcare Information" survey of 1,700 American adults, 73% obtained health-related information from the Internet (Weber Shandwick 2018). In another nationally representative survey of over 1,300 U.S. teens and young adults (14- to 22-year-olds), 87% reported that they had searched online for health information (Rideout and Fox 2018). People seek online health information primarily because of their health conditions or illnesses, and they most often go to websites about specific health conditions (Cornejo 2018). In fact, more and more patients rely on the Internet as their first source to acquire knowledge about their own health conditions before pursuing a professional diagnosis (Tan and Goonawardene 2017). There is no doubt that the Internet and constantly evolving social technologies are transforming the healthcare industry as well as the way people seek health information and knowledge.

The most popular online resources for health-related information include WebMD, Facebook, YouTube and Twitter (Stricker 2014). As the most accessed website for health information, WebMD provides a question-and-answer (Q&A) platform, wherein people can search for answers and ask questions (Harper et al. 2009). While most people believe the medical information they obtain online (e.g., from medical Q&A sites) is trustworthy and of good quality (Taylor 2011), this is not always the case (McLeod 1998; Silberg et al. 1997). Because the health information that people find online has measurable impacts on their health-related decisions (Lawhon 2016), a deeper understanding of people's value judgment of online medical information is especially critical for patients, medical institutions, and society at large.

In this study, we examine what constitutes a helpful answer on medical Q&A sites. An important feature of many Q&A sites is a helpfulness voting system that relies on "wisdom of the crowd" to gauge the perceived value of information provided in an answer. To reduce information overload and highlight

more valuable content (Jones et al. 2004), many Q&A sites allow readers to cast votes on the helpfulness of answers and display helpful answers more prominently. When future readers decide which content to read, they also rely heavily on the judgment of others, such as helpfulness evaluation of the content (Otterbacher et al. 2011). Thus, a better understanding of the factors influencing answer helpfulness can assist future readers to find helpful information more easily and motivate contributors to create more useful information. This understanding is even more critical in the medical domain, as the medical answers that readers deem more valuable should be incorporated more into their health-related decisions and have critical implications for their lives.

In the medical setting, we define the helpfulness of an answer to a medical question as the extent to which the answer is perceived to facilitate readers' health-related judgment and decision-making process (see Mudambi and Schuff 2010; Yin et al. 2014). Answer helpfulness reflects information diagnosticity, as the answer can provide diagnostic value across *multiple stages* of readers' judgment and decision-making process (Mudambi and Schuff 2010). Prior studies on perceived value of answers in Q&A sites have examined the impact of answer characteristics, among which the most frequently studied is answer length that reflects the total amount of information contained in an answer (e.g., Edelman 2012; Shah and Pomerantz 2010). In addition, existing research in the Q&A setting has studied the role of various contextual cues—influential characteristics of environmental signals that surround the focal content (Murnane et al. 1999). Note that contextual cues are signals from the content's environment, not characteristics of the content. Studies have examined how a reader's value assessment of an answer could be shaped by the question's features (e.g., Harper et al. 2008; Shah and Pomerantz 2010; Zhang and Wang 2016), source characteristics (e.g., Edelman 2012; Lou et al. 2013; Oh and Worrall 2013), and the type of Q&A site (e.g., Chen et al. 2010; Fichman 2011; Jeon et al. 2010).

Our research deviates from earlier work in two important ways. First, we situate our study in the setting of *medical* Q&A sites and highlight the critical role of two unique types of contextual cues—the language attributes of a question, and the acuteness of the disease associated with the question. These cues are similar to a product's characteristics (such as its average rating and product type) in the online

review setting because they are all signals from the environment surrounding the target of users' value judgment (i.e., answers and reviews, respectively). However, the question and the disease are two unique aspects of medical Q&A sites that are not present in other types of user-generated content. These two types of contextual cues are also highly relevant for the judgment of answer helpfulness in medical Q&A sites, as an answer in this unique setting is always a response addressing a particular *question* and a particular *disease*. Second, despite significant progress made towards understanding the impacts of content characteristics and contextual cues on a reader's value assessment of an answer, the role of content-context congruence is relatively unexplored.[1] However, some emerging studies in user-generated content have revealed that readers' evaluation of content helpfulness is not context-free because contextual cues can shape readers' beliefs and mindsets (Huang et al. 2013; Yin et al. 2016). Drawing on the priming literature and construal-level theory (Trope and Liberman 2011; Winkielman and Cacioppo 2001), we propose that question-answer congruences and disease-answer congruences can increase a readers' perceived helpfulness of the answer. An empirical analysis of data collected from WebMD Answers provides evidence for our predictions.

This research demonstrates the critical role of content-context congruence in the helpfulness evaluation of answers in medical Q&A sites and makes two key contributions. First, this research complements and extends recent findings in other settings that value judgment of content is not context-free (e.g., Yin et al. 2016). Although it is intuitive to assume that perceived helpfulness of certain content is determined primarily by the content itself, we propose and find evidence for a more nuanced perspective emphasizing the importance of congruence between content and contextual cues. This content-context congruence perspective extends the proposals of value-from-fit from goal pursuit (Higgins 2000) and truth-from-fit from lie detection (Hansen and Wänke 2010) to reader evaluation of user-generated content, suggesting that readers also derive value from the congruence of content features

[1] For brevity, we use "content-context congruence" to refer to the congruence between content characteristics and contextual cues. Note that "content" in "content-context congruence" means content characteristics or language attributes, not substantive content or what is being conveyed.

with their context.[2] While our hypotheses are situated in the specific setting of medical Q&A sites, the broader content-context congruence perspective can be applied to other types of Q&A sites and all types of user-generated content. Thus, this paper opens up exciting opportunities for future research to explore the role of interplay between content and context in readers' judgment and decision-making. Second, we take advantage of the unique nature of the medical Q&A sites and identify two types of relevant and prominent contextual cues in this setting. In particular, our theoretical framework and findings highlight the critical role of *diseases* and deepen our understanding of readers' value judgment of *medical* answers that can have life-changing consequences. Given the huge presence and undisputable significance of medical Q&A sites in people's daily lives, we believe our findings have strong practical implications.

LITERATURE REVIEW

User-generated content (UGC) refers to content that is created by members of the general public and distributed over the Internet (Daugherty et al. 2008; Krumm et al. 2008), such as consumer reviews on e-commerce sites and answers on Q&A sites. Prior literature on the antecedents of perceived UGC diagnosticity has examined both content and contextual cues. In the setting of consumer reviews that has received the most academic attention (see Cheung and Thadani 2012 for a literature review), a large body of studies examined how a review's content characteristics such as length, readability and emotional expression can influence the perceived helpfulness of the review (e.g., Mudambi and Schuff 2010; Yin et al. 2014, 2017). In addition, some studies investigated the effect of contextual cues such as source and product characteristics (e.g., Cheung et al. 2012; Forman et al. 2008; Mudambi and Schuff 2010).

As the popularity of online Q&A sites grows, researchers have become increasingly interested in the perceived value of answer contributions (Lou et al. 2013; Oh and Worrall 2013; Shah and Pomerantz 2010). The literature on Q&A has also investigated the effect of content and contextual cues. Some studies focused on the characteristics of an answer, such as answer length (Edelman 2012; Shah and Pomerantz 2010), emotional support (Kim and Oh 2009), and politeness (Lee et al. 2019). Other studies

[2] We consider the terms of "congruence" and "fit" as interchangeable, and we use "congruence" throughout the paper.

examined the contextual cues available from the question, source, and even Q&A site. Question level factors that have been studied include question length (Shah and Pomerantz 2010; Zhang and Wang 2016), topic (e.g., technology, business, and entertainment) (Harper et al. 2008), and the number of the question's answers (Shah and Pomerantz 2010). Source characteristics include the answer contributor's expertise (Edelman 2012; Oh and Worrall 2013), experience (Oh and Worrall 2013; Shah and Pomerantz 2010), reputation (Chen et al. 2010), and motivation (Lou et al. 2013). Furthermore, some studies examined the types of a given Q&A site as contextual cues (Chen et al. 2010; Fichman 2011; Harper et al. 2008; Jeon et al. 2010). Appendix A reviews the studies conducted in the Q&A setting and summarizes the examined antecedents of the perceived value of answers.

Despite the ample UGC research into content and contextual cues, limited attention has been paid to the interaction between them or how their interplay affects reader perception of content diagnosticity (Cheung and Thadani 2012). Notably, a few emerging studies in online reviews suggest that the impact of review content on review helpfulness is not context-free, but instead dependent on contextual cues such as the average rating or type of product (Huang et al. 2013; Yin et al. 2016). Inspired by value-from-fit proposal in goal pursuit (Higgins 2000) and task-technology fit theory (Goodhue and Thompson 1995), we apply the fit perspective to medical Q&A sites, and argue that answer content may be perceived to be more helpful if its characteristics are congruent with the contextual cues available from the unique setting.

THEORY AND HYPOTHESES DEVELOPMENT

Congruence

The concept of congruence is receiving more attention in various academic disciplines, and a growing stream of research has demonstrated the importance of congruence between an entity and its contextual cues (e.g., Edwards 2008). For example, in goal pursuit, people experience regulatory fit when the manner of their engagement in an activity (e.g., eager or vigilant) fits their goal orientation or interests regarding that activity at the moment (e.g., promotion or prevention) (Higgins 2005). This regulatory fit can increase value judgments and evaluations, such as consumers' perceived value of their choices (Avnet and Higgins 2006). Similarly, task-technology fit theory posits that information technology is more likely

to be used and positively influence individual performance when the technology's capabilities match the tasks that the user must perform (Goodhue and Thompson 1995).

In this paper, we apply the notion of fit or congruence to user-generated content and argue that content-context congruence plays an indispensable role in readers' value evaluations. When people read and evaluate a piece of content, information contained in the content is in the focus of their attention. At the same time, they are also exposed to contextual cues—defined as information that is present in the content's environment but is peripheral to the focus of attention (Murnane et al. 1999). In the setting of medical Q&A sites, we propose that the congruence of an answer's content characteristics with contextual cues from the question and the disease topic can influence reader perception of answer helpfulness.

In the following, we introduce two characteristics of *answer* content: language concreteness and emotional intensity. Furthermore, we identify two classes of contextual cues unique in medical Q&A sites: concreteness and emotional intensity expressed in the content of a *question*, and the acuteness of a *disease*. We finally develop and propose our hypotheses regarding question-answer congruences and disease-answer congruences.[3] Our theoretical framework is illustrated in Figure 1.

[3] The outcome variable in our research is perceived helpfulness of an *answer*. Thus, we selected answer-related congruence variables as our independent variables, including question-answer congruences and disease-answer congruences. On the other hand, question-disease congruences might be more relevant for predicting question-level outcome variables, such as intention to read answers after seeing a question (which would be outside the scope of this research).

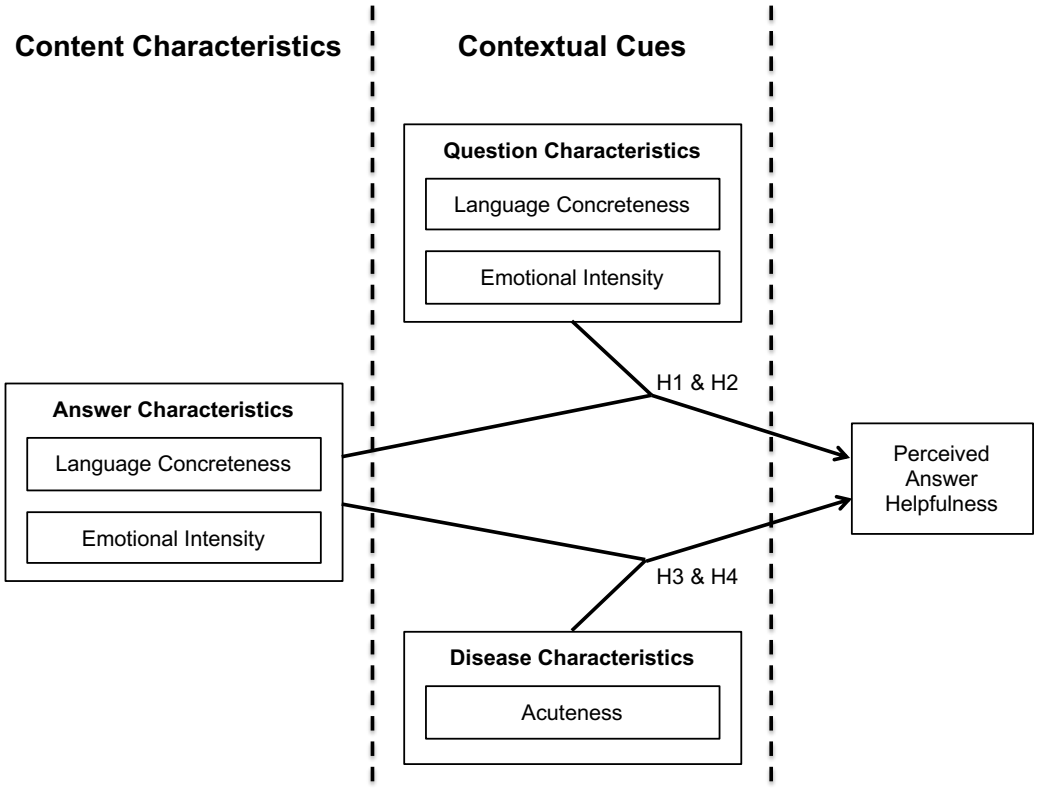

Figure 1: Theoretical Framework

Language Concreteness and Emotional Intensity

Answer contributors can write an answer in different ways (i.e., making use of different words) while expressing the same idea. In addition to the substantive content of a message (i.e., what is being conveyed), the attributes of its language can also influence the audience (Bradac et al. 1979; Miller et al. 2007). Even when the substantive content of two messages is similar, variation in the use of specific words may lead readers to perceive the messages in different ways, which can influence their evaluation of the messages (Berry et al. 1997). In this paper, we focus on two aspects of language used in the content of answers: one cognitive and the other emotional.

First, a cognitive aspect of language is its concreteness level, as answers on medical Q&A sites can be framed in either a concrete or abstract manner when addressing the same question and conveying a similar meaning. We define language concreteness as the extent to which words provide descriptive, specific, and vivid information about an object or situation (Hansen and Wänke 2010). For example, "a stethoscope" and "a medical device" can be used to describe the same equipment in our setting, but the

former is more concrete and less abstract than the latter. Compared with abstract language, concrete language is better recalled (Tse and Altarriba 2009), judged as more truthful (Hansen and Wänke 2010) and more persuasive in shaping attitudes and behaviors (Larrimore et al. 2011). In user-generated content, the concreteness of words used in online reviews increases consumers' evaluation of the reviews' helpfulness (Li et al. 2013). We expect language concreteness to also play a role in readers' helpfulness evaluation of answers on Q&A sites.

In addition to the cognitive aspect of the content, answer contributors may also express their feelings to varying degrees. We define emotions in our setting as subjective feelings expressed by the content contributor, and we define emotional intensity as the percentage of expressed emotions in the content, following the existing literature (see Fujita et al. 1991; Kahn et al. 2007). Emotional expressions are prevalent in user-generated content. Although emotions are typically not part of the substantive content, they can have critical implications for reader value perception of the content. For example, the intensity of expressed emotions in an online review presents a prominent signal for readers to make sense of the review (Jensen et al. 2013), and it can influence reader perception of review helpfulness in sophisticated manners (Yin et al. 2017). On community-based social Q&A sites where users form a community and support each other, emotional support is an important selection criteria for best answers (Kim et al. 2007; Kim and Oh 2009). Similarly, on Q&A sites primarily used for knowledge exchange, answer contributors can express emotions at different intensity levels. Next, we develop hypotheses about the impact of congruence in content characteristics between a question and its answer on reader evaluation of that answer's helpfulness.

Question-Answer Congruences

Contextual cues present within the questions can shape reader evaluation of answers (Harper et al. 2008; Shah and Pomerantz 2010). After users input keywords to search for answers to a question on a Q&A site, they typically see a list of previously asked questions. When they click on a specific question, that question is universally displayed on top of the page followed by answers to that question. Readers are very likely to first read the question and judge its relevance to their inquiry before proceeding to read the

answers. In the setting of online reviews, consumers would observe a product's average rating and other summary rating statistics before reading individual reviews, and empirical evidence suggests that their evaluation of an individual review's helpfulness is shaped by the contextual cues such as that product's average rating (Yin et al. 2016). It is reasonable to expect the same to occur on Q&A sites, whereby a reader's evaluation of an individual answer's helpfulness can be shaped by contextual cues such as the language attributes of the question's content.

While answer content may vary in language concreteness and emotional intensity, question content can also vary in regard to these language attributes. In our first two hypotheses, we argue that question-answer congruences could contribute to the perceived value of the answer. A probable reason for this effect is that priming facilitates processing and enhances fluency (Winkielman et al. 2012). Fluency is broadly defined as an individual's subjective experience of ease with which they process information (Oppenheimer 2008). Fluency arises from a wide variety of factors, such as font clarity, visual contrast, and repeated exposure (Alter and Oppenheimer 2009). Most relevant to our purposes, fluency can arise from a priming effect of the question's content characteristics. Specifically, the observation of a stimulus (called the prime) can reduce reaction time to subsequently encountered related concepts, because the prime sends activation to related concepts in memory and this pre-activation leads to faster identification (Collins and Loftus 1975). In particular, prior exposure to certain stimuli has been found to enhance fluency of matching stimuli when compared to non-matching stimuli at a later time (Winkielman and Cacioppo 2001).

The way in which a question is written might prime readers and lead them to be more fluent with answers having matching content characteristics. Reading questions associated with concrete thinking can activate one's concrete mindsets and predispose them to concrete words in the memory (Freitas et al. 2004; Wakslak and Trope 2009). When readers are primed by a concrete question and put into a concrete mindset, they should find concrete answers easier to process than abstract answers, and vice versa. Similarly, the mere observation of emotionally laden stimuli (i.e., affect primes) can elicit emotions in the observers by pre-activating emotion-related words and concepts in their memory (Chartrand et al. 2006).

When readers are primed by an emotion-laden question, they should find emotional answers more fluent than non-emotional answers, and vice versa.

Moreover, fluency is typically associated with positive evaluations (Winkielman et al. 2012). The ease of processing indicates the likelihood of an external stimulus being good or bad. In fact, fluency is usually a learned and readily available heuristic for identifying better choices (Gigerenzer 2007). A wide range of studies have provided evidence that fluency obtained through various means (such as repetition and priming) enhances liking, even for an initially neutral stimulus (see Reber et al. 2004 for a comprehensive review). Physiological evidence also shows that fluency triggers stronger activity over the "smiling" region of the brain (Winkielman and Cacioppo 2001). Applied to our setting, answers that readers find more fluent should be evaluated more positively. Taken together, we propose that the congruence in language concreteness or emotional intensity between a question and its answer, namely the *question-answer concreteness congruence* and *question-answer emotion congruence*, can enhance the perceived helpfulness of the answer:

H1: The congruence in language concreteness between an answer and its question is positively related to the perceived helpfulness of the answer.

H2: The congruence in emotional intensity between an answer and its question is positively related to the perceived helpfulness of the answer.

Disease-Answer Congruences

Our next set of hypotheses argue that disease-answer congruences can also contribute to the perceived value of the answer. In addition to question characteristics, the unique setting of *medical* Q&A sites creates another prominent contextual cue. A distinctive aspect of various diseases is that they differ in treatment duration. Diseases are typically classified into acute diseases that last a limited period of time and chronic diseases that are lengthy in duration (Murrow and Oglesby 1996; Perrin et al. 1993). Because disease topics that users inquire about on medical Q&A sites vary significantly in terms of urgency and duration, the acute nature of a disease could influence a user's mindset as they seek medical information and advice (Holman and Lorig 2004).

In this paper, we define disease acuteness as the extent to which a disease is urgent and short-term in nature (Parkes and Jewell 2001; Zochling et al. 2006). Since acute diseases can be cured in a short time, readers consulting questions related to an acute disease tend to think about how to treat it in a timely manner. On the other hand, chronic diseases are typically long-lasting conditions, thus readers concerned with a chronic disease tend to reflect on how to deal with its long-term consequences that may be unclear until sometime in the distant future. Notably, disease acuteness reflects temporal distance, defined as the perception of some event being close to or far away from the reference point of the present (Maglio et al. 2013).

According to construal-level theory, temporal distance has direct implications on the level of mental construal—the extent to which people's *thinking* about an event is at a concrete or abstract level (Liberman and Trope 1998; Trope and Liberman 2003). As temporal distance of an event increases, the *thinking* of the event would be more abstract in people's minds (Trope and Liberman 2010). This is because high-level abstract mental representations are more likely than low-level concrete mental representations to remain unchanged as people get closer to the events in the distant future (Trope and Liberman 2011). For example, contacting a friend is more abstract than sending that friend a WhatsApp message. If this action happens in the distant future (e.g., in a year), then people are more likely to think about contacting the friend because this abstract notion is more stable over time than sending the WhatsApp message; in fact, the platform of WhatsApp might not be available when one is trying to contact the friend in a year's time (e.g., Yahoo Messenger was discontinued). As a result, it is more useful for people to think about distant events with a more abstract (and thus stable) mindset, and to think about urgent events with a more concrete mindset. On medical Q&A sites, readers of an answer addressing an acute (or chronic) disease should be predisposed to *think* about the disease in a concrete (or abstract) manner.

Because the differences in temporal distance of events can lead to mindsets at different construal levels, greater congruence between temporal distance and the concreteness level of information in a message can make the message more persuasive and thus more diagnostic (Kim et al. 2009). This effect

can be explained using the same fluency arguments from before: when temporal distance (and the corresponding, activated mental construal of a person's mindset—*thinking* at a concrete or abstract level) is congruent with the concreteness of information, they tend to experience increased fluency, which in turn leads to a more positive evaluation of the information (Oppenheimer 2006; Reber and Schwarz 1999). For example, a number of experimental studies in message framing literature find compelling evidence that higher congruence between a message's concreteness level and the psychological distance (such as temporal distance, social distance, etc.) inherent in the decision-making context can positively influence attitudes and behaviors (Freling et al. 2014; White et al. 2011).

The above reasoning suggests that the congruence between disease acuteness and information concreteness can increase perceived helpfulness of the information, while the concreteness level of a message can manifest in different ways. First, the concreteness level of words used in the message is the most straightforward reflection of that message's concreteness level. The use of more concrete words indicates a more concrete message, while the use of more abstract words indicates a more abstract message. In the medical setting, when readers encounter answer information whose concreteness level matches disease acuteness (for example, a concrete answer to a question on an acute disease, or an abstract answer to a question on a chronic disease), they are likely to experience greater fluency (see Reber et al. 2004) and perceive the answer to be more helpful. In contrast, when readers encounter answer information whose concreteness does not match disease acuteness, they are less likely to experience fluency and thus may perceive the answer to be less helpful. Therefore, the congruence between the disease acuteness and the answer's concreteness level should positively influence readers' perceived helpfulness of the answer, and we propose the following hypothesis regarding this congruence (hereafter *disease-answer concreteness congruence*).

H3: The congruence between an answer's concrete level and disease acuteness is positively related to the perceived helpfulness of the answer.

Second, a message's concreteness level can manifest in the intensity of expressed emotions. Although emotionally charged words are distinct from either concrete or abstract words (Altarriba and

Bauer 2004; Altarriba et al. 1999), experimental studies have provided evidence that emotional words are more vivid than neutral or non-emotional words (Dewhurst and Parry 2000; Kensinger and Corkin 2003). Survey data also supported that emotionally charged words are rated as more concrete and vivid (Campos 1989). Thus, even if the concreteness level of words used in substantive content is held constant, more intense expressions of emotion should result in a more concrete answer. Extending the same reasoning from construal-level theory to our setting, readers are likely to experience greater fluency when they encounter answers whose emotional intensity matches disease acuteness (for example, an emotional answer responding to a question on an acute disease, or a non-emotional answer is responding to a question on a chronic disease). The positive experience of fluency can result in a more positive evaluation of the answer. In contrast, readers are less likely to experience fluency when the answers' emotional intensity does not match disease acuteness (for example, an emotional answer addressing a question on a chronic disease, or a non-emotional answer addressing a question on an acute disease). This reduced fluency should lead readers to rate the answer as less helpful. Therefore, we propose the following hypothesis regarding the congruence between disease acuteness and answer emotional intensity (hereafter *disease-answer emotion congruence*):

H4: The congruence between an answer's emotional intensity and disease acuteness is positively related to the perceived helpfulness of the answer.

To summarize, we have presented a theoretical framework in this section that focuses on the critical role of congruence between unique contextual cues (i.e., language attributes of the question and acuteness of diseases) and language attributes of answer content in the setting of medical Q&A sites. Based on the priming literature and construal-level theory (Trope and Liberman 2011; Winkielman and Cacioppo 2001), we hypothesize that question-answer congruence and disease-answer congruence can positively influence the perceived value of answers.

METHOD AND RESULTS

To test these hypotheses, we collected and analyzed a unique data set from WebMD Answers. According to ComScore's Media Metrix report (comScore 2013), WebMD.com is the leading healthcare portal with over 50 million unique visitors per month. As one of its provided services, WebMD Answers allows users to ask medical questions under a range of diverse categories such as bipolar disorder, high blood pressure, pregnancy and others. Because medical questions are classified into diverse disease topics, data from WebMD Answers enables us to quantify the key independent variables (see more details below) and is ideal for testing the above outlined hypotheses. In January 2018, we collected the content of all the 26,318 questions and their answers categorized under 99 common disease topics that WebMD Answers organized for users to explore. We collected 35,098 answers in total, 16,802 of which (over 46%) had received at least one vote. We also collected data about 3022 distinctive authors who provided the answers.

Measures

The dependent variable is answer helpfulness (*Helpfulness*). For each answer from WebMD Answers, we collected the following data in addition to the answer content: the number of "helpful" votes and the number of total votes (that is the total number of users who voted either "Yes" or "No" to the question "Was this helpful?"). Following prior literature (Mudambi and Schuff 2010; Yin et al. 2014), we measured *Helpfulness* at the answer level using the ratio of the number of helpful votes to the total number of votes for an answer.

The independent variables of interest are four cognitive and emotional congruence variables—two of these are factors measuring congruence between a question and its answer (*question-answer concreteness congruence* and *question-answer emotion congruence*) while the other two are factors measuring congruence between the acuteness of disease topic discussed in the question and the answer (*disease-answer concreteness congruence* and *disease-answer emotion congruence*). Before we can quantify these congruence variables, we need to measure the concreteness and emotional intensity of each question and answer, and acuteness of the relevant disease.

To measure content concreteness, we relied on the dictionary of concreteness ratings of nearly 40,000 generally known English words and expressions obtained through Internet crowdsourcing (Brysbaert et al. 2014). The dictionary provides the ratings of these words and expressions on a 5-point scale ranging from abstract to concrete. We adopted this dictionary in our study because (1) it is specifically designed for measuring language concreteness, (2) its reliability and validity have been demonstrated by its creators, and (3) it has been used to quantify language concreteness in the medial setting (e.g., Cousins et al. 2018) as well as in online reviews (e.g., Ransbotham et al. 2019). Notably, this dictionary includes words from the medical domain. For example, "disinfection" has a concreteness rating of 2.67, whereas "aorta" has a concreteness raring of 4.61. To calculate the average concreteness of an answer, we divided the sum of concreteness ratings of words in the answer by the total number of words in the answer. We calculated the average concreteness of a question in a similar way.

Second, we used Linguistic Inquiry and Word Count (LIWC) software to compute the emotional intensity of each question and answer. Pennebaker et al. (2007) developed a psychometrically validated dictionary comprised of over 4,000 words and word stems assigned to multiple categories. This tool has been widely adopted in various fields (Tausczik and Pennebaker 2010) and extensively validated (Pennebaker et al. 2007; Pennebaker and Francis 1996). Most relevant for our purposes, this dictionary includes a list of words that indicate positive or negative emotions. LIWC has been found to be a valid tool for measuring emotional discourse (Bantum and Owen 2009; Kahn et al. 2007), and it is increasingly used by information systems and marketing scholars to quantify emotional expression in various types of user-generated content (e.g., Ransbotham et al. 2019; Schweidel and Moe 2014; Yin et al. 2014). Following common practice and suggestions from the LIWC inventors (Pennebaker et al. 2007), we measured emotional intensity through calculating the number of emotional words (identified by LIWC dictionary) divided by the total number of words.

Third, we measured the acuteness of each disease topic through the professional evaluations obtained from family physicians. WebMD Answers categorized the questions and their answers under 99 common disease topics. We recruited three American family physicians and instructed them to

independently rate the acuteness of 99 health topics listed by WebMD Answers. Specifically, each physician read the definitions for acute and chronic diseases, and then evaluated each disease along a 5-point Likert scale (1 = Always Chronic, 2 = Often Chronic, 3 = Sometimes Chronic and Sometimes Acute, 4 = Often Acute, 5 = Always Acute).[4] To assess the reliability of this measure, we calculated Cohen's kappa among the three physicians for each disease (Cohen 1960). The average index was 0.54, suggesting moderate agreement among the physicians (McHugh 2012).[5] As a result, we aggregated these ratings for each disease and used the mean values as a measure for disease acuteness.

Finally, we measured the four congruence variables through the commonly used difference score approach (David et al. 1989; Keller 1994; Oh and Pinsonneault 2007; Tilcsik 2014). This approach is deemed appropriate for examining congruence concepts when congruence is defined theoretically as a match between two component variables that are independent of the outcome variable (supposedly predicted by the congruence) (Venkatraman 1989). Under this circumstance, the difference score measurement provides a variety of advantages over other approaches (including residuals, interaction, and split-sample analysis): it reduces the measurement error and interpretation problems associated with residuals from regression equations, alleviates the multicollinearity concern from the use of interaction terms, and addresses the problems of restricted range, reduced sample size, and interpretation of congruence magnitude that occur in a split-sample analysis (Keller 1994; Venkatraman 1989). Therefore, we applied the difference score approach in the present study, as content concreteness, emotional intensity, and disease acuteness are all independent from the helpfulness of answer content at the theoretical level.

[4] Among the 99 disease topics, 83 topics were rated by all the physicians, while the rest were rated by one or two physicians. As a robustness check, we did the same analyses excluding the observations from disease topics that were rated by only one or two physicians, and we found that the results did not qualitatively differ. To be comprehensive, we report results with answers from all the 99 disease topics in the main text.

[5] Our results were robust to using a 3-point Likert scale (i.e., collapsing the points of 1 and 2 to represent "chronic", and collapsing 4 and 5 to represent "acute"; the average index of Cohen's kappa was 0.63 in this case, suggesting substantial agreement among the physicians).

For each congruence variable, we first standardized both components making up the congruence because some component variables (e.g., answer emotional intensity and disease acuteness) were measured along different scales. Then we calculated the absolute difference between these two standardized component variables. To ease interpretation, we multiplied the absolute difference by -1 to create the measure for the congruence variables: a higher value of the congruence variable indicates greater congruence between its two components. We used this same procedure to operationalize the two question-answer congruence variables and the two disease-answer congruence variables. As an example, disease-answer concreteness congruence was measured as the extent to which the concrete level of the answer's content matches the acuteness of the disease topic associated with the question.

We also included a series of control variables to account for the influence of source, answer, and question characteristics. First, source characteristics have been found to contribute to information helpfulness (e.g., Cheung et al. 2012; Forman et al. 2008). For the Q&A setting, we controlled for author expertise (*Author Expertise*), which is a dummy variable indicating whether the contributor of an answer is an expert (e.g., M.D., which stands for "Doctor of Medicine"). We also accounted for author credibility (*Author Credibility*), operationalized as the total number of helpful votes an author has received divided by the total number of questions that author has answered.

Next, we controlled for characteristics of answer content that may influence its information helpfulness. First, the amount of information was measured by answer length and defined as the number of words in an answer (*Answer Length*). Second, the difficulty of reading an answer (*Answer Reading Difficulty*) has direct implications for its perceived helpfulness (see Korfiatis et al. 2012). Thus, we calculated the Coleman–Liau Index as a proxy for reading difficulty, which is an estimate of the U.S. grade level that a student would need to read and understand a text sample (DuBay 2004). Third, because of a potential relationship between total votes and content helpfulness (Mudambi and Schuff 2010), we included the total number of votes an answer received (*Answer Total Votes*) as a control. Fourth, an older answer may have had more time to accumulate more helpful votes. Thus, we also controlled for the number of days (*Answer Days*) since the answer was posted. Fifth, we controlled for an answer's

concreteness (*Answer Concreteness*) and emotional intensity (*Answer Emotional Intensity*) that might directly impact answer helpfulness. Sixth, the first few answers to a question are likely to receive more votes than the latter answers, so we controlled for the sequence of a focal answer (*Answer Sequence*). Finally, we included a dummy variable to indicate whether there are any other answers marked as helpful (*Other Helpful Answers*): 1 if yes, 0 otherwise.

We also included relevant control variables at the question level. We accounted for the generality of the question by counting the number of keywords identified by WebMD for a specific question (*Question Keywords*). A question with more keywords is considered a more general question as it can be answered from more perspectives, potentially increasing the difficulty of creating helpful answers. We also included the number of words in each question (*Question Length*) in the models.

Finally, we used disease-level fixed effects to control for disease topic heterogeneity. These fixed effects are algebraically equivalent to including a dummy for every disease in our sample, and so they enabled us to control for differences in the average helpfulness of answers across diseases. Moreover, we controlled for the temporal trend in answers. It is likely that helpful ratings change as a Q&A website matures. Therefore, we included month-year pair dummies and the dummies for the days of a week. Table 1 reports the descriptive statistics for the variables in our analysis, and Table 2 reports the correlations.

Table 1: Descriptive Statistics

	Mean	S.D.	Min	Median	Max
1.Helpfulness	0.59	0.17	0.03	0.55	1.00
2.Question-Answer Concreteness Congruence	-1.32	1.20	-12.76	-1.06	0.00
3.Question-Answer Emotion Congruence	-1.33	1.29	-23.23	-1.02	0.00
4.Disease-Answer Concreteness Congruence	-1.25	1.11	-12.03	-1.05	0.00
5.Disease-Answer Emotion Congruence	-1.35	1.02	-24.04	-1.20	0.00
6.Author Expertise	0.21	0.41	0.00	0.00	1.00
7.Author Credibility	28.85	134.38	0.00	12.34	8812.00
8.Answer Length	95.48	97.67	1.00	64.00	853.00
9.Answer Reading Difficulty	10.68	5.15	-16.10	10.20	137.00
10.Answer Total Votes	48.11	424.33	1.00	12.00	23698.00
11.Answer Days	1874.20	868.78	187.00	1701.00	4956.00
12.Answer Concreteness	2.21	0.31	0.00	2.23	4.89
13.Answer Emotional Intensity	5.60	4.81	0.00	5.00	100.00
14.Answer Sequence	1.38	1.44	1.00	1.00	24.00
15.Other Helpful Answers	0.30	0.46	0.00	0.00	1.00
16.Question Keywords	3.00	1.52	1.00	3.00	5.00
17.Question Length	31.97	36.49	0.00	15.00	112.00

Table 2: Correlations

	1	2	3	4	5	6	7	8
1.Helpfulness	1.00							
2.Question-Answer Concreteness Congruence	0.11	1.00						
3.Question-Answer Emotion Congruence	0.14	0.12	1.00					
4.Disease-Answer Concreteness Congruence	0.10	0.61	0.11	1.00				
5.Disease-Answer Emotion Congruence	0.06	0.11	0.34	0.24	1.00			
6.Author Expertise	0.10	0.08	0.01	0.08	-0.06	1.00		
7.Author Credibility	0.06	0.01	0.05	0.01	0.00	0.04	1.00	
8.Answer Length	0.26	0.14	0.17	0.16	0.16	-0.01	0.11	1.00
9.Answer Reading Difficulty	0.12	-0.12	0.05	-0.13	0.02	0.06	0.04	0.10
10.Answer Total Votes	0.06	0.00	0.03	0.00	-0.01	0.01	0.55	0.08
11.Answer Days	0.29	0.15	0.26	0.11	0.09	-0.01	0.08	0.27
12.Answer Concreteness	0.04	0.48	0.02	0.45	-0.02	0.07	-0.01	-0.02
13.Answer Emotional Intensity	-0.04	0.04	-0.12	0.06	0.00	-0.13	-0.03	-0.06
14.Answer Sequence	-0.01	-0.09	-0.04	-0.07	-0.03	-0.12	0.02	-0.05
15.Other Helpful Answers	-0.06	-0.11	-0.09	-0.07	-0.02	-0.16	0.02	-0.09
16.Question Keywords	-0.23	-0.07	-0.28	-0.04	-0.05	0.02	-0.05	-0.15
17.Question Length	-0.21	-0.06	-0.27	-0.04	-0.01	-0.05	-0.04	-0.10

	9	10	11	12	13	14	15	16
9.Answer Reading Difficulty	1.00							
10.Answer Total Votes	0.01	1.00						
11.Answer Days	0.31	0.02	1.00					
12.Answer Concreteness	-0.16	0.00	0.03	1.00				
13.Answer Emotional Intensity	0.00	-0.02	-0.09	0.00	1.00			
14.Answer Sequence	-0.03	0.02	-0.14	-0.01	0.00	1.00		
15.Other Helpful Answers	-0.06	0.06	-0.21	-0.02	0.03	0.41	1.00	
16.Question Keywords	-0.23	-0.03	-0.48	0.00	0.05	0.06	0.14	1.00
17.Question Length	-0.23	-0.01	-0.50	-0.04	0.07	0.06	0.16	0.69

Data Analysis and Results

The dependent variable, the ratio of the number of helpful votes over the number of total votes for an answer, is bounded between 0 and 1. Given that a bounded dependent variable leads to inconsistent parameter estimates using ordinary least squares (OLS), we used the two-limited Tobit model (Greene 2003; Kennedy 2008). This model has been widely used in diverse disciplines, including marketing, strategy, finance, management, and information systems (Barthélemy 2008; Ferreira et al. 2010; Luo and Homburg 2007; Mudambi and Schuff 2010; Sosa 2009).

Table 3 presents the results of the main regressions used to test our hypotheses. We standardized all continuous independent variables to ease the comparison of effect sizes. We first entered the control variables in Model 1 (baseline model), and then added the four congruence variables in Model 2. We

conducted the likelihood ratio (*LR*) test to compare Model 2 to Model 1 and found that the addition of the four congruence variables improves the model fit significantly ($p < 0.001$). Furthermore, to assess any potential multicollinearity, we calculated variance inflation factor (VIF) scores for all independent and control variables in Model 2. The scores were all below the rule-of-thumb value of 10 (Kennedy 2008), indicating that multicollinearity was not a concern.

Table 3: Tobit Regressions

	Model 1	Model 2
Author Expertise	0.043***	0.041***
	(0.004)	(0.004)
Author Credibility	-0.001	-0.001
	(0.001)	(0.001)
Answer Length	0.026***	0.024***
	(0.001)	(0.001)
Answer Reading Difficulty	0.004***	0.006***
	(0.001)	(0.001)
Answer Total Votes	0.002	0.002
	(0.001)	(0.001)
Answer Days	0.049	0.038
	(0.139)	(0.138)
Answer Concreteness	0.004***	0.0001
	(0.001)	(0.001)
Answer Emotional Intensity	0.003***	0.003**
	(0.001)	(0.001)
Answer Sequence	0.006***	0.006***
	(0.001)	(0.001)
Other Helpful Answers	0.002*	0.002**
	(0.001)	(0.001)
Question Keywords	-0.019***	-0.019***
	(0.002)	(0.002)
Question Length	-0.001	-0.001
	(0.002)	(0.002)
Question-Answer Concreteness Congruence		0.003***
		(0.001)
Question-Answer Emotion Congruence		0.003**
		(0.001)
Disease-Answer Concreteness Congruence		0.006***
		(0.001)
Disease-Answer Emotion Congruence		0.004***
		(0.001)
Constant	0.481	0.511
	(0.412)	(0.411)
N	16726	16726
Log-likelihood	5,519.81	5,558.64
p-value, LR test		0.000

All continuous independent variables standardized; Disease topic dummies, days of week dummies, and month*year dummies included; Standard errors in parentheses; ** p < 0.05, *** p < 0.01

Hypotheses 1 and 2 propose a positive effect of the congruence in concreteness and emotional intensity between a question and its answer on the perceived helpfulness of that answer. In Model 2 of Table 4, *Question-Answer Concreteness Congruence* was positively related to perceived answer helpfulness ($\beta = 0.003$, $p < 0.01$), and *Question-Answer Emotion Congruence* was positively related to perceived answer helpfulness ($\beta = 0.003$, $p < 0.05$). Controlling for other factors, an answer whose concreteness or emotional intensity is more similar to that of its corresponding question is expected to be more helpful. Therefore, the first two hypotheses are supported.

Hypotheses 3 and 4 propose a positive effect of the congruence between disease acuteness and an answer's concreteness or emotional intensity on the perceived helpfulness of the answer. In Model 2 of Table 4, *Disease-Answer Concreteness Congruence* was positively related to perceived answer helpfulness ($\beta = 0.006$, $p < 0.01$), and *Disease-Answer Emotion Fit* was positively related to perceived answer helpfulness ($\beta = 0.004$, $p < 0.01$). Controlling for other factors, an answer with a higher level of congruence between its concreteness or emotional intensity and the acuteness of its corresponding disease topic is expected to be more helpful. Therefore, the last two hypotheses are supported.[6]

As robustness checks, we conducted a number of additional tests (see Appendix B) to address the following potential issues: a sample selection bias, the interdependence of answers within the same topic, the bounded nature of our ratio-based dependent variable, extreme values of our dependent variable when the total number of helpfulness votes is small, and endogeneity caused by reverse causality or omitted variables. All our results still hold when using alternative models in these robustness checks.

[6] Because our congruence variables are composite indices, it may not be reasonable to directly interpret their coefficients. On the other hand, we can compare the effect sizes of congruence variables with other variables known to influence information helpfulness. For example, the intensity of emotions expressed in UGC has been revealed to play an important role in reader's perception of the content (Jensen et al. 2013), and it is also a critical component comprising two of our congruence variables. As shown in Model 2 of Table 3, the effect sizes of the two emotion-related congruence variables (i.e., *Question-Answer Emotion Congruence* and *Disease-Answer Emotion Congruence*) are comparable to the effect sizes of *Answer Emotional Intensity*.

GENERAL DISCUSSIONS

Situated in the critical and unique setting of medical Q&A sites, we supplement an emerging stream of research studying factors that influence the perceived helpfulness of user-generated content (e.g., Jensen et al. 2013; Mudambi and Schuff 2010; Yin et al. 2014), and explore what constitutes a helpful answer to a medical question. Focusing on two unique types of contextual cues—the language attributes of the question and the acuteness level of the associated disease, we propose a content-context congruence perspective: an answer that is congruent with its question or disease will be rated by readers as more helpful. An empirical analysis of data from WebMD Answers provides evidence for our proposed theoretical framework. At a broader level, the results illustrate the important role of content-context congruence in studying the perceived value of user-generated content and demonstrate its important implications for both theory and practice.

Theoretical Implications

This paper offers a number of theoretical contributions. First and foremost, this study proposes a congruence perspective in explaining users' value perception of user-generated content. While previous studies on antecedents of information helpfulness generally assumed that content needs to be written in a certain way or placed in a certain context in order to be considered helpful, emerging evidence in online reviews suggests that contextual cues can influence the beliefs and mindsets of consumers as they read and make sense of reviews (Huang et al. 2018; Yin et al. 2016). Building on and going beyond recent studies that found "context matters," we advance a congruence perspective inspired by fit theories in organizational behavior, individual goal pursuit, and individual performance in IT use. In organizations, an employee whose individual needs, values and goals are congruent with the organization's culture, values and norms (namely person-organization fit) performs better and has higher job satisfaction (Hoffman and Woehr 2006). When people pursue a goal, a regulatory fit between the way they engage in an activity and their goal orientation increases their perceived value of the activity (Higgins 2005). In the use of information systems, a task-technology fit between the technology's capabilities and the associated tasks increases technology adoption and individual performance (Goodhue and Thompson 1995). All of

these theories share the perspective that the congruence between someone or something and its environment leads to positive outcomes.

Adapting this perspective to user-generated content (UGC), we derive a congruence-focused theoretical framework based on the priming literature and construal level theory. Specifically, we argue and find empirical evidence that a greater congruence between content and contextual cues can lead readers to perceive the content as more helpful. In addition to the independent impacts of content characteristics and contextual cues revealed in the UGC literature, the level of congruence between content and context may also play a nontrivial role. For example, concrete information may not always be desirable; rather, its impact can be determined by its congruence with environmental cues such as those that prime people's concrete (vs. abstract) thinking or activate their concrete (vs. abstract) mindsets. Although question and disease characteristics in our setting are very different types of contextual cues compared with a product's average rating or product type in online reviews, they are all fundamentally environmental signals around the target of users' value assessment. Our proposed hypotheses involve characteristics of questions and diseases that may not be applicable in other settings, but our proposed congruence perspective has broader appeals and can be extended to other types of user-generated content such as online reviews. For example, a match between review content and product type may contribute to perceived value of online reviews (Huang et al. 2013). Our findings highlight the importance of taking this more nuanced congruence perspective when studying the perceived value of user-generated content, and they open up exciting new avenues for future research in this area.

Second, our disease-related findings illustrate the nontrivial role of event timing in influencing readers' thinking and their subsequent evaluation of UGC. Studies in online word-of-mouth provided evidence that the temporal distance of events (e.g., authoring a review, consuming a product) and *other psychological distances* (such as spatial distance and social distance) can jointly affect consumer evaluation and preferences (Huang et al. 2016; Zhao and Xie 2011). Our work complements and extends these earlier studies by demonstrating that the temporal distance of events and *user-generated content* can also jointly influence readers' value judgment of the content. This study also illustrates the possibility of

applying construal-level theory in online, interpersonal settings where multiple parties (such as content contributors and readers) are involved. Construal-level theory was originally developed to explain how people mentally construe *their own* personal events that will occur in their near or distant future (Liberman and Trope 1998), and its applications have started to appear in the information systems field over recent years (e.g., Ho et al. 2015). This paper extends its application from intra-personal to inter-personal situations wherein readers make sense of online information produced by others. While the majority of evidence for construal-level theory comes from laboratory experiments (see Trope and Liberman 2010), this paper is also among the first to provide empirical evidence for the theory based on an analysis of a large-scale archival data set (see also Huang et al. 2016).

Third, we explore readers' value evaluation of online health information from medical Q&A sites, a *knowledge-focused* form of user-generated content that has critical implications for people's *health-related* decisions and lives. The majority of studies examining the perceived diagnosticity of content were carried out in the setting of online word-of-mouth (Chevalier and Mayzlin 2006; Mudambi and Schuff 2010; Willemsen et al. 2011). However, critical aspects of online reviews, such as ratings that indicate reviewers' opinions about a product or service, are not relevant on Q&A websites where the focus is on the effective exchange of information and knowledge. In addition, a product's average rating represents a critical contextual cue that may influence consumers' initial beliefs and their interpretation of reviews (Yin et al. 2016), but this cue is again not relevant on Q&A sites. We argue that the unique nature of medical Q&A sites warrants its own examination. Answers from medical Q&A sites are posted to address specific medical *questions* and specific *diseases* that vary in their acuteness. Our findings indicate that these unique contextual cues significantly influence the way that prospective readers make sense of and evaluate answers written in certain ways. For example, our question-related hypotheses and findings suggest that *questions* are an integral part of Q&A sites, and they should not be overlooked in explaining readers' value judgment of answers. Thus, we not only add to a growing literature examining factors that drive the perceived helpfulness of user contributions, but also contribute to a better understanding of *knowledge* evaluation—how to identify and promote more helpful answers—in the *medical* setting.

Practical Implications

This research offers useful insights for medical sites that intend to educate and help patients looking for answers to their health-related inquires. Perceived value of online health information is critical as it determines the extent to which the information is incorporated into people's health decisions and influences health outcomes. Websites providing medical information should be aware that in the healthcare domain, diseases or health conditions are a primary reason for people to search online information in the first place (Cornejo 2018), and they also represent a prominent contextual cue that can shape readers' thinking and mindsets. While different diseases are necessary in organizing the vast quantity of medical information and advices available online, our findings extend the significance of disease categories from information organization to information evaluation. If the acuteness of two disease categories is very different, then the experience of crafting or selecting the most helpful information for one category of diseases might not be directly applicable for the other category.

This study also offers practical insights for content contributors. Content contributors have a basic human need for competence (Deci and Ryan 2000). Such need is most likely to be satisfied when their contributions are valued and perceived as helpful. Most contributors probably hold the simplistic assumption that answers written in a certain way (e.g., longer, more concrete, etc.) are more valuable. They are typically not aware of the subtle role of contextual cues in prospective readers' value judgments. Our findings suggest that such awareness is critical for content contributors. For example, concrete information is typically considered more credible and valuable than abstract information (Hansen and Wänke 2010), but the reverse might occur if the question is written in an abstract or non-emotional manner or if the disease is chronic. The mere realization that readers interested in different medical questions may differ in their thinking and mindsets can go a long way in guiding contributor's writing efforts.

Our findings also have valuable implications for the design of Q&A sites. Content is essential for a Q&A website; if readers perceive the content from the site to be useful for their decision-making, they will "stick" to the site longer. Thus, most user-generated content platforms offer guidelines for content

contributors in order to guide their writing in ways more conducive to being helpful. These guidelines almost universally promote certain characteristics of content without regard to context, such as providing more details, avoid being emotional, etc. As an example, guidelines such as "provide details and be specific" may prompt contributors to write more concrete answers regardless of whether the question is written concretely or abstractly, or whether the disease is acute or chronic. Instead of prescribing a simplistic formula centered around characteristics of the "ideal" content, Q&A sites are advised to educate content contributors about the pitfalls of following simple formulas and the diversity of readers looking for answers of different questions. For example, the awareness of "readers interested in chronic diseases are different from those interested in acute diseases" can help contributors reduce their natural tendency to create more concrete answers when they strive to write more valuable answers.

In addition, the helpfulness rating of answers is the primary means by which Q&A sites determine the value of different answers in order to highlight more helpful or the most helpful ones. However, the accumulation of helpfulness votes takes considerable time after an answer is posted. Our work offers useful insights for Q&A sites attempting to develop ways of estimating and predicting, *a priori*, the extent to which an answer will be perceived helpful by future readers. In addition to content characteristics most frequently associated with information helpfulness such as the length of content, our results suggest that the congruence between content and its unique contextual cues should also be taken into consideration in the estimation and prediction of answer helpfulness. Readily available software and dictionaries such as those used in this work can help automate the measurement of content-context congruence and the prediction of answer helpfulness.

Limitations and Future Directions

A number of exciting opportunities present themselves for future studies. First, it would be valuable for further investigation to explore the mechanisms underlying the impact of content-context congruence on perceived information helpfulness. When readers make sense of certain content, their beliefs and mindsets play a critical role in their helpfulness evaluation of the content (e.g., Yin et al. 2016). However, a limitation of this research is the lack of data on readers and voters, which inhibits the

possibility of investigating the mechanisms. Our empirical study is also limited in the establishment of causality. While our hypotheses are described in associational terms, the logic behind them implies that congruence between content and its contextual cues causes greater perceived helpfulness of answers. Although the cross-sectional nature of our data makes it challenging to directly examine this causal influence, we conducted various robustness checks (including an analysis using the instrument variable approach) that effectively alleviate concerns of endogeneity. Future research should explore both the underlying mechanisms and causal relationships using complementary methodologies, such as surveys and experiments.

Second, while our theoretical framework focuses on the congruence between content characteristics and contextual cues unique in the medical Q&A setting, content-context congruence may play similar roles in other settings of user-generated content involving different types of contextual cues. In online reviews, for example, product type is a prominent contextual cue, and consumers may have a certain expectation for emotional expression in the reviews depending on the nature of the product (such as hedonic versus utilitarian products). Thus, it stands to reason that a review in which emotional intensity is congruent with product type may be evaluated more favorably. Future research should explore the generalizability of our proposed theoretical framework in other kinds of user-generated content.

Third, our paper examines the consequences of content-context congruence—how this congruence influences readers' perceived value of answers. It would also be fascinating to explore the sources of content-context congruence. For example, the concreteness level or emotional intensity of content in a question may not only prime readers, but also prime some answer contributors. More studies are needed to uncover the source of congruence as well as the types of contributors who are more likely to write congruent answers.

Fourth, we focus on two content characteristics—language concreteness and emotional intensity—as the basis for our question-answer congruence and disease-answer congruence variables. We selected these variables because they are theory-driven, have been demonstrated as critical in users' evaluation of information diagnosticity, and can yield meaningful theoretical implications. However, a question and an

answer can be congruent in other aspects, and future research can explore whether similarity in these other aspects (e.g., emotional valence) can also contribute to answer helpfulness.

Fifth, this research focuses on disease acuteness—one dimension of psychological distance (i.e., temporal distance)—in our latter two hypotheses and archival data analysis. In addition to the temporal dimension, other dimensions of psychological distance (such as spatial distance or social distance) may also play important roles in a variety of user-generated content settings. For example, regarding online reviews, some platforms such as Amazon reveal the geographical locations of reviewers (Forman et al. 2008), which may influence how a consumer evaluates concrete versus abstract reviews depending on his/her spatial distance from the reviewers. In order to fully understand the role of psychological distance, more studies are needed to explore the effects of its other dimensions, and how different dimensions interact to influence user judgment (Huang et al. 2018; Zhao and Xie 2011).

Conclusion

Complementing the emerging interest in the role of contextual cues in reader helpfulness assessment of certain content, we propose a content-context congruence perspective in order to explain the perceived value of answers on medical Q&A sites. Our research provides real-world evidence that answer content is evaluated as more helpful if it is congruent with the contextual cues unique to medical Q&A sites. We believe that this work extends our current understanding of the interplay between content and context in user-generated content, and we look forward to future research exploring the role of different kinds of content-context congruence in various user-generated content settings.

ACKNOWLEDGEMNT

We thank the senior editor, the associate editor, and three anonymous reviewers for their constructive guidance during the review process. We are grateful to Sabyasachi Mitra and Chih-Ping Wei for their insightful feedback on earlier versions of this paper. This work was partially supported by the Ministry of Science and Technology of Taiwan under the grant [MOST 108-2410-H-004-227-MY2] and by E.SUN Bank.

REFERENCES

Adamic, L. A., Zhang, J., Bakshy, E., and Ackerman, M. S. 2008. "Knowledge Sharing and Yahoo Answers: Everyone Knows Something," *Proceedings of the 17th international conference on World Wide Web*: ACM, pp. 665-674.

Agarwal, R., Gao, G., DesRoches, C., and Jha, A. K. 2010. "Research Commentary—the Digital Transformation of Healthcare: Current Status and the Road Ahead," *Information Systems Research* (21:4), pp. 796-809.

Altarriba, J., and Bauer, L. M. 2004. "The Distinctiveness of Emotion Concepts: A Comparison between Emotion, Abstract, and Concrete Words," *American Journal of Psychology* (117:3), pp. 389-410.

Altarriba, J., Bauer, L. M., and Benvenuto, C. 1999. "Concreteness, Context Availability, and Imageability Ratings and Word Associations for Abstract, Concrete, and Emotion Words," *Behavior Research Methods, Instruments and Computers* (31:4), pp. 578-602.

Alter, A. L., and Oppenheimer, D. M. 2009. "Uniting the Tribes of Fluency to Form a Metacognitive Nation," *Personality and Social Psychology Review* (13:3), pp. 219-235.

Anderson, J. D., and Core, J. E. 2018. "Managerial Incentives to Increase Risk Provided by Debt, Stock, and Options," *Management Science* (64:9), pp. 4408-4432.

Arellano, M., and Bond, S. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies* (58:2), pp. 277-297.

Avnet, T., and Higgins, E. T. 2006. "How Regulatory Fit Affects Value in Consumer Choices and Opinions," *Journal of Marketing Research* (43:1), pp. 1-10.

Bantum, E. O. C., and Owen, J. E. 2009. "Evaluating the Validity of Computerized Content Analysis Programs for Identification of Emotional Expression in Cancer Narratives," *Psychological Assessment* (21:1), pp. 79-88.

Barthélemy, J. 2008. "Opportunism, Knowledge, and the Performance of Franchise Chains," *Strategic Management Journal* (29:13), pp. 1451-1463.

Berry, D. S., Pennebaker, J. W., Mueller, J. S., and Hiller, W. S. 1997. "Linguistic Bases of Social Perception," *Personality and Social Psychology Bulletin* (23:5), pp. 526-537.

Bradac, J. J., Bowers, J. W., and Courtright, J. A. 1979. "Three Language Variables in Communication Research: Intensity, Immediacy, and Diversity," *Human Communication Research* (5:3), pp. 257-269.

Brysbaert, M., Warriner, A. B., and Kuperman, V. 2014. "Concreteness Ratings for 40 Thousand Generally Known English Word Lemmas," *Behavior Research Methods* (46:3), pp. 904-911.

Campos, A. 1989. "Emotional Values of Words: Relations with Concreteness and Vividness of Imagery," *Perceptual and Motor Skills* (69:2), pp. 495-498.

Chartrand, T. L., van Baaren, R. B., and Bargh, J. A. 2006. "Linking Automatic Evaluation to Mood and Information Processing Style: Consequences for Experienced Affect, Impression Formation, and Stereotyping," *Journal of Experimental Psychology: General* (135:1), pp. 70-77.

Chen, Y., Ho, T.-H., and Kim, Y.-M. 2010. "Knowledge Market Design: A Field Experiment at Google Answers," *Journal of Public Economic Theory* (12:4), pp. 641-664.

Cheung, C. M.-Y., Sia, C.-L., and Kuan, K. K. Y. 2012. "Is This Review Believable? A Study of Factors Affecting the Credibility of Online Consumer Reviews from an Elm Perspective," *Journal of the Association for Information Systems* (13:8), pp. 618-635.

Cheung, C. M. K., and Thadani, D. R. 2012. "The Impact of Electronic Word-of-Mouth Communication: A Literature Analysis and Integrative Model," *Decision Support Systems* (54:1), pp. 461-470.

Chevalier, J. A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), pp. 345-354.

Cohen, J. 1960. "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* (20), pp. 37-46.

Collins, A. M., and Loftus, E. F. 1975. "A Spreading-Activation Theory of Semantic Processing," *Psychological Review* (82:6), pp. 407-428.

comScore. 2013. "Comscore Media Metrix Ranks Top 50 U.S. Web Properties for February 2013." Retrieved April 28, 2013, 2013, from http://www.comscore.com/Insights/Press_Releases/2013/3/comScore_Media_Metrix_Ranks_Top_50_U.S._Web_Properties_for_February_2013

Cornejo, C. 2018. "Social Life of Health Information." Retrieved July 16, 2019, from https://www.wegohealth.com/2018/02/05/social-life-of-health-information/

Cousins, K. A. Q., Ash, S., Olm, C. A., and Grossman, M. 2018. "Longitudinal Changes in Semantic Concreteness in Semantic Variant Primary Progressive Aphasia (Svppa)," *eNeuro* (5:6), pp. e0197-0118.2018.

Daugherty, T., Eastin, M. S., and Bright, L. 2008. "Exploring Consumer Motivations for Creating User-Generated Content," *Journal of Interactive Advertising* (8:2), pp. 16-25.

David, F. R., Pearce, J. A., and Randolph, W. A. 1989. "Linking Technology and Structure to Enhance Group Performance," *Journal of Applied Psychology* (74:2), pp. 233-241.

Deci, E. L., and Ryan, R. M. 2000. "The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior," *Psychological Inquiry* (11:4), pp. 227-268.

Dewhurst, S. A., and Parry, L. A. 2000. "Emotionality, Distinctiveness, and Recollective Experience," *European Journal of Cognitive Psychology* (12:4), pp. 541-551.

DuBay, W. H. 2004. "The Principles of Readability." *Impact Information*, from http://www.nald.ca/library/research/readab/readab.pdf

Edelman, B. 2012. "Earnings and Ratings at Google Answers," *Economic Inquiry* (50:2), pp. 309-320.

Edwards, J. R. 2008. "Person–Environment Fit in Organizations: An Assessment of Theoretical Progress," *The Academy of Management Annals* (2:1), pp. 167-230.

Erdem, S. A., and Harrison-Walker, L. J. 2006. "The Role of the Internet in Physician–Patient Relationships: The Issue of Trust," *Business Horizons* (49:5), pp. 387-393.

Ferreira, M. A., Massa, M., and Matos, P. 2010. "Shareholders at the Gate? Institutional Investors and Cross-Border Mergers and Acquisitions," *Review of Financial Studies* (23:2), pp. 601-644.

Fichman, P. 2011. "A Comparative Assessment of Answer Quality on Four Question Answering Sites," *Journal of Information Science* (37:5), pp. 476-486.

Fichman, R. G., Kohli, R., and Krishnan, R. 2011. "Editorial Overview—the Role of Information Systems in Healthcare: Current Research and Future Trends," *Information Systems Research* (22:3), pp. 419-428.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19:3), pp. 291-313.

Freitas, A. L., Gollwitzer, P., and Trope, Y. 2004. "The Influence of Abstract and Concrete Mindsets on Anticipating and Guiding Others' Self-Regulatory Efforts," *Journal of Experimental Social Psychology* (40:6), pp. 739-752.

Freling, T. H., Vincent, L. H., and Henard, D. H. 2014. "When Not to Accentuate the Positive: Re-Examining Valence Effects in Attribute Framing," *Organizational Behavior and Human Decision Processes* (124:2), pp. 95-109.

Fujita, F., Diener, E., and Sandvik, E. 1991. "Gender Differences in Negative Affect and Well-Being: The Case for Emotional Intensity," *Journal of Personality and Social Psychology* (61:3), pp. 427-434.

Gazan, R. 2006. "Specialists and Synthesists in a Question Answering Community," *Proceedings of the American Society for Information Science and Technology* (43:1), pp. 1-10.

Gigerenzer, G. 2007. *Gut Feelings: The Intelligence of the Unconscious*. New York: Viking Press.

Goodhue, D. L., and Thompson, R. L. 1995. "Task-Technology Fit and Individual Performance," *MIS Quarterly* (19:2), pp. 213-236.

Greene, W. H. 2003. *Econometric Analysis*. Upper Saddle River, NJ.: Prentice Hall.

Hansen, J., and Wänke, M. 2010. "Truth from Language and Truth from Fit: The Impact of Linguistic Concreteness and Level of Construal on Subjective Truth," *Personality and Social Psychology Bulletin* (36:11), pp. 1576-1588.

Harper, F. M., Moy, D., and Konstan, J. A. 2009. "Facts or Friends? Distinguishing Informational and Conversational Questions in Social Q&a Sites," in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Boston, MA, USA: ACM, pp. 759-768.

Harper, F. M., Raban, D., Rafaeli, S., and Konstan, J. A. 2008. "Predictors of Answer Quality in Online Q&a Sites," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*: ACM, pp. 865-874.

Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error," *Econometrica* (47:1), pp. 153-161.

Higgins, E. T. 2000. "Making a Good Decision: Value from Fit," *American Psychologist* (55:11), pp. 1217-1230.

Higgins, E. T. 2005. "Value from Regulatory Fit," *Current Directions in Psychological Science* (14:4), pp. 209-213.

Ho, C. K. Y., Ke, W., and Liu, H. 2015. "Choice Decision of E-Learning System: Implications from Construal Level Theory," *Information & Management* (52:2), pp. 160-169.

Hoffman, B. J., and Woehr, D. J. 2006. "A Quantitative Review of the Relationship between Person–Organization Fit and Behavioral Outcomes," *Journal of Vocational Behavior* (68:3), pp. 389-399.

Holman, H., and Lorig, K. 2004. "Patient Self-Management: A Key to Effectiveness and Efficiency in Care of Chronic Disease," *Public Health Reports* (119:3), pp. 239-243.

Hong, Y., Hu, Y., and Burtch, G. 2018. "Embeddedness, Prosociality, and Social Influence: Evidence from Online Crowdfunding," *MIS Quarterly* (42:4), pp. 1211-1224.

Huang, L., Tan, C.-H., Ke, W., and Wei, K.-K. 2013. "Comprehension and Assessment of Product Reviews: A Review-Product Congruity Proposition," *Journal of Management Information Systems* (30:3), pp. 311-343.

Huang, L., Tan, C. H., Ke, W., and Wei, K. K. 2018. "Helpfulness of Online Review Content: The Moderating Effects of Temporal and Social Cues," *Journal of the Association for Information Systems* (19:6), pp. 503-522.

Huang, N., Burtch, G., Hong, Y., and Polman, E. 2016. "Effects of Multiple Psychological Distances on Construal and Consumer Evaluation: A Field Study of Online Reviews," *Journal of Consumer Psychology* (26:4), pp. 474-482.

Jensen, M. L., Averbeck, J. M., Zhang, Z., and Wright, K. B. 2013. "Credibility of Anonymous Online Product Reviews: A Language Expectancy Perspective," *Journal of Management Information Systems* (30:1), pp. 293-324.

Jeon, G. Y., Kim, Y.-M., and Chen, Y. 2010. "Re-Examining Price as a Predictor of Answer Quality in an Online Q&a Site," *Proceedings of the SIGCHI conference on human factors in computing systems*: ACM, pp. 325-328.

Jeon, J., Croft, W. B., Lee, J. H., and Park, S. 2006. "A Framework to Predict the Quality of Answers with Non-Textual Features," *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, Seattle, Washington, USA: ACM, pp. 228-235.

Jones, Q., Ravid, G., and Rafaeli, S. 2004. "Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration," *Information Systems Research* (15:2), pp. 194-210.

Kahn, J. H., Tobin, R. M., Massey, A. E., and Anderson, J. A. 2007. "Measuring Emotional Expression with the Linguistic Inquiry and Word Count," *The American Journal of Psychology* (120:2), pp. 263-286.

Kashyap, V., and Murtha, B. R. 2017. "The Joint Effects of Ex Ante Contractual Completeness and Ex Post Governance on Compliance in Franchised Marketing Channels," *Journal of Marketing* (81:3), pp. 130-153.

Keller, R. T. 1994. "Technology-Information Processing Fit and the Performance of R&D Project Groups: A Test of Contingency Theory," *Academy of Management Journal* (37:1), pp. 167-179.

Kennedy, P. 2008. *A Guide to Econometrics*, (6 ed.). Cambridge, MA: Wiley-Blackwell.

Kensinger, E. A., and Corkin, S. 2003. "Memory Enhancement for Emotional Words: Are Emotional Words More Vividly Remembered Than Neutral Words?," *Memory & Cognition* (31:8), pp. 1169-1180.

Kim, H., Rao, A. R., and Lee, A. Y. 2009. "It's Time to Vote: The Effect of Matching Message Orientation and Temporal Frame on Political Persuasion," *Journal of Consumer Research* (35:6), pp. 877-889.

Kim, S., Oh, J. S., and Oh, S. 2007. "Best-Answer Selection Criteria in a Social Q&a Site from the User-Oriented Relevance Perspective," *Proceedings of the 70th annual meeting of the American Society for Information Science and Technology*, pp. 1-15.

Kim, S., and Oh, S. 2009. "Users' Relevance Criteria for Evaluating Answers in a Social Q&a Site," *Journal of the American Society for Information Science & Technology* (60:4), pp. 716-727.

Kivits, J. 2006. "Informed Patients and the Internet: A Mediated Context for Consultations with Health Professionals," *Journal of Health Psychology* (11:2), pp. 269-282.

Klein, K. J., and Kozlowski, S. W. J. 2000. *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*. San Francisco, CA: Jossey-Bass.

Korfiatis, N., García-Bariocanal, E., and Sánchez-Alonso, S. 2012. "Evaluating Content Quality and Helpfulness of Online Product Reviews: The Interplay of Review Helpfulness Vs. Review Content," *Electronic Commerce Research and Applications* (11:3), pp. 205-217.

Kreft, I. G., and Leeuw, J. d. 1998. *Introducing Multilevel Modeling*. Thousand Oaks, CA: Sage.

Krumm, J., Davies, N., and Narayanaswami, C. 2008. "User-Generated Content," *Pervasive Computing, IEEE* (7:4), pp. 10-11.

Kuan, K. K. Y., Hui, K.-L., Prasarnphanich, P., and Lai, H.-Y. 2015. "What Makes a Review Voted? An Empirical Investigation of Review Voting in Online Review Systems," *Journal of the Association for Information Systems* (16:1), pp. 48-71.

Larrimore, L., Jiang, L., Larrimore, J., Markowitz, D., and Gorski, S. 2011. "Peer to Peer Lending: The Relationship between Language Features, Trustworthiness, and Persuasion Success," *Journal of Applied Communication Research* (39:1), pp. 19-37.

Lawhon, L. 2016. "New Health Union Survey Reveals Importance of Online Health Communities." Retrieved July 16, 2019, from https://health-union.com/news/online-health-experience-survey/

Lee, S.-Y., Rui, H., and Whinston, A. B. 2019. "Is Best Answer Really the Best Answer? The Politeness Bias," *MIS Quarterly* (43:2), pp. 579-600.

Lewbel, A. 2012. "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business & Economic Statistics* (30:1), pp. 67-80.

Li, M., Huang, L., Tan, C.-H., and Wei, K.-K. 2013. "Helpfulness of Online Product Reviews as Seen by Consumers: Source and Content Features," *International Journal of Electronic Commerce* (17:4), pp. 101-136.

Liberman, N., and Trope, Y. 1998. "The Role of Feasibility and Desirability Considerations in near and Distant Future Decisions: A Test of Temporal Construal Theory," *Journal of Personality and Social Psychology* (75:1), pp. 5-18.

Lou, J., Fang, Y., Lim, K. H., and Peng, J. Z. 2013. "Contributing High Quantity and Quality Knowledge to Online Q&a Communities," *Journal of the American Society for Information Science and Technology* (64:2), pp. 356-371.

Luo, X., and Homburg, C. 2007. "Neglected Outcomes of Customer Satisfaction," *Journal of Marketing* (71:2), pp. 133-149.

Maglio, S. J., Trope, Y., and Liberman, N. 2013. "The Common Currency of Psychological Distance," *Current Directions in Psychological Science* (22:4), pp. 278-282.

McHugh, M. L. 2012. "Interrater Reliability: The Kappa Statistic," *Biochemia Medica* (22:3), pp. 276-282.

McLeod, S. D. 1998. "The Quality of Medical Information on the Internet: A New Public Health Concern," *Archives of Ophthalmology* (116:12), pp. 1663-1665.

Miller, C. H., Lane, L. T., Deatrick, L. M., Young, A. M., and Potts, K. A. 2007. "Psychological Reactance and Promotional Health Messages: The Effects of Controlling Language, Lexical Concreteness, and the Restoration of Freedom," *Human Communication Research* (33:2), pp. 219-240.

Mudambi, S. M., and Schuff, D. 2010. "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com," *MIS Quarterly* (34:1), pp. 185-200.

Murnane, K., Phelps, M. P., and Malmberg, K. 1999. "Context-Dependent Recognition Memory: The Ice Theory," *Journal of Experimental Psychology: General* (128:4), pp. 403-415.

Murrow, E. J., and Oglesby, F. M. 1996. "Acute and Chronic Illness: Similarities, Differences and Challenges," *Orthopaedic Nursing* (15:5), pp. 47-51.

Oh, S., and Worrall, A. 2013. "Health Answer Quality Evaluation by Librarians, Nurses, and Users in Social Q&A," *Library & Information Science Research* (35:4), pp. 288-298.

Oh, W., and Pinsonneault, A. 2007. "On the Assessment of the Strategic Value of Information Technologies: Conceptual and Analytical Approaches," *MIS Quarterly* (31:2), pp. 239-265.

Oppenheimer, D. M. 2006. "Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly," *Applied Cognitive Psychology* (20:2), pp. 139-156.

Oppenheimer, D. M. 2008. "The Secret Life of Fluency," *Trends in Cognitive Sciences* (12:6), pp. 237-241.

Otterbacher, J., Hemphill, L., and Dekker, E. 2011. "Helpful to You Is Useful to Me: The Use and Interpretation of Social Voting," *Proceedings of the American Society for Information Science and Technology* (48:1), pp. 1-10.

Parkes, M., and Jewell, D. P. 2001. "The Management of Severe Crohn's Disease," *Alimentary Pharmacology & Therapeutics* (15:5), pp. 563-573.

Pennebaker, J. W., Booth, R. J., and Francis, M. E. 2007. *Linguistic Inquiry and Word Count (Liwc2007)*. Austin, TX: LIWC.

Pennebaker, J. W., and Francis, M. E. 1996. "Cognitive, Emotional, and Language Processes in Disclosure," *Cognition and Emotion* (10:6), pp. 601-626.

Perrin, E. C., Newacheck, P., Pless, I. B., Drotar, D., Gortmaker, S. L., Leventhal, J., Perrin, J. M., Stein, R. E. K., Walker, D. K., and Weitzman, M. 1993. "Issues Involved in the Definition and Classification of Chronic Health Conditions," *Pediatrics* (91:4), p. 787.

Ransbotham, S., Lurie, N. H., and Liu, H. 2019. "Creation and Consumption of Mobile Word of Mouth: How Are Mobile Reviews Different?," *Marketing Science* (38:5), pp. 733-912.

Raudenbush, S. W., and Bryk, A. S. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.

Reber, R., and Schwarz, N. 1999. "Effects of Perceptual Fluency on Judgments of Truth," *Consciousness and Cognition* (8:3), pp. 338-342.

Reber, R., Schwarz, N., and Winkielman, P. 2004. "Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience?," *Personality and Social Psychology Review* (8:4), pp. 364-382.

Rideout, V., and Fox, S. 2018. "Digital Health Practices, Social Media Use, and Mental Well-Being among Teens and Young Adults in the U.S." Retrieved July 16, 2019, from https://www.hopelab.org/reports/pdf/a-national-survey-by-hopelab-and-well-being-trust-2018.pdf

Schweidel, D. A., and Moe, W. W. 2014. "Listening in on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research* (51:4), pp. 387-402.

Shah, C. 2011. "Measuring Effectiveness and User Satisfaction in Yahoo! Answers," *First Monday* (16:2).

Shah, C., and Pomerantz, J. 2010. "Evaluating and Predicting Answer Quality in Community Qa," in: *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. Geneva, Switzerland: ACM, pp. 411-418.

Shaver, J. M. 1998. "Accounting for Endogeneity When Assessing Strategy Performance: Does Entry Mode Choice Affect Fdi Survival?," *Management Science* (44:4), pp. 571-585.

Silberg, W. M., Lundberg, G. D., and Musacchio, R. A. 1997. "Assessing, Controlling, and Assuring the Quality of Medical Information on the Internet: Caveant Lector Et Viewor—Let the Reader and Viewer Beware," *JAMA* (277:15), pp. 1244-1245.

Sosa, M. L. 2009. "Application-Specific R&D Capabilities and the Advantage of Incumbents: Evidence from the Anticancer Drug Market," *Management Science* (55:8), pp. 1409-1422.

Stricker, P. 2014. "Social Media: An Emerging Communication Modality." Retrieved July 16, 2019, from http://www.naylornetwork.com/cmsatoday/articles/index-v2.asp?aid=296436&issueID=30171

Tan, S. S.-L., and Goonawardene, N. 2017. "Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review," *Journal of medical Internet research* (19:1), p. e9.

Tausczik, Y. R., and Pennebaker, J. W. 2010. "The Psychological Meaning of Words: Liwc and Computerized Text Analysis Methods," *Journal of Language and Social Psychology* (29:1), pp. 24-54.

Taylor, H. 2011. "The Growing Influence and Use of Health Care Information Obtained Online," Harris Poll.

Tilcsik, A. 2014. "Imprint–Environment Fit and Performance: How Organizational Munificence at the Time of Hire Affects Subsequent Job Performance," *Administrative Science Quarterly* (59:4), pp. 639-668.

Trope, Y., and Liberman, N. 2003. "Temporal Construal," *Psychological Review* (110:3), pp. 403-421.

Trope, Y., and Liberman, N. 2010. "Construal-Level Theory of Psychological Distance," *Psychological Review* (117:2), pp. 440-463.

Trope, Y., and Liberman, N. 2011. "Construal Level Theory," in *Handbook of Theories of Social Psychology,* P.A.M.V. Lange, A.W. Kruglanski and E.T. Higgins (eds.). Thousand Oaks, CA: SAGE Publications Inc., pp. 118-134.

Tse, C.-S., and Altarriba, J. 2009. "The Word Concreteness Effect Occurs for Positive, but Not Negative, Emotion Words in Immediate Serial Recall," *British Journal of Psychology* (100:1), pp. 91-109.

Venkatraman, N. 1989. "The Concept of Fit in Strategy Research: Toward Verbal and Statistical Correspondence," *Academy of Management Review* (14:3), pp. 423-444.

Wakslak, C., and Trope, Y. 2009. "The Effect of Construal Level on Subjective Probability Estimates," *Psychological Science* (20:1), pp. 52-58.

Weber Shandwick. 2018. "The Great American Search for Healthcare Information." from https://www.webershandwick.com/wp-content/uploads/2018/11/Healthcare-Info-Search-Report.pdf

White, K., MacDonnell, R., and Dahl, D. W. 2011. "It's the Mind-Set That Matters: The Role of Construal Level and Message Framing in Influencing Consumer Efficacy and Conservation Behaviors," *Journal of Marketing Research* (48:3), pp. 472-485.

Willemsen, L. M., Neijens, P. C., Bronner, F., and de Ridder, J. A. 2011. ""Highly Recommended!" The Content Characteristics and Perceived Usefulness of Online Consumer Reviews," *Journal of Computer-Mediated Communication* (17:1), pp. 19-38.

Winkielman, P., and Cacioppo, J. T. 2001. "Mind at Ease Puts a Smile on the Face: Psychophysiological Evidence That Processing Facilitation Elicits Positive Affect," *Journal of Personality and Social Psychology* (81:6), pp. 989-1000.

Winkielman, P., Huber, D. E., Kavanagh, L., and Schwarz, N. 2012. "Fluency of Consistency: When Thoughts Fit Nicely and Flow Smoothly," in *Cognitive Consistency: A Fundamental Principle in Social Cognition,* B. Gawronski and F. Strack (eds.). New York: Guilford Press, pp. 89-111.

Yin, D., Bond, S. D., and Zhang, H. 2014. "Anxious or Angry? Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews," *MIS Quarterly* (38:2), pp. 539-560.

Yin, D., Bond, S. D., and Zhang, H. 2017. "Keep Your Cool or Let It Out: Nonlinear Effects of Expressed Arousal on Perceptions of Consumer Reviews," *Journal of Marketing Research* (54:3), pp. 447-463.

Yin, D., Mitra, S., and Zhang, H. 2016. "When Do Consumers Value Positive Vs. Negative Reviews? An Empirical Investigation of Confirmation Bias in Online Word of Mouth," *Information Systems Research* (27:1), pp. 131-144.

Zhang, Y., and Wang, P. 2016. "Interactions and User-Perceived Helpfulness in Diet Information Social Questions & Answers," *Health Information & Libraries Journal* (33:4), pp. 295-307.

Zhao, M., and Xie, J. 2011. "Effects of Social and Temporal Distance on Consumers' Responses to Peer Recommendations," *Journal of Marketing Research* (48:3), pp. 486-496.

Zochling, J., Grill, E., Scheuringer, M., Liman, W., Stucki, G., and Braun, J. 2006. "Identification of Health Problems in Patients with Acute Inflammatory Arthritis, Using the International Classification of Functioning, Disability and Health (Icf)," *Clinical and Experimental Rheumatology* (24:3), pp. 239-246.

APPENDICES

Appendix A: Literature Review on Antecedents of the Perceived Value of Answers

Authors and Year	Answer Characteristics	Question Characteristics	Source Characteristics	Q&A Site Characteristics
Gazan (2006)			answerer expertise	
Jeon et al. (2006)	answer length, the number of times the answer is clicked, the number of times the answer is recommended by other users, the number of times that users print the answer		answerer expertise, answerer activity	
Kim et al. (2007)	content value, cognitive value, and socio-emotional value			
Adamic et al. (2008)	answer length		answerer expertise	
Harper et al. (2008)		question topic		paid-based vs. free-based sites
Kim and Oh (2009)	content, cognition, utility, socio-emotion		answerer expertise, answerer experience	
Chen et al. (2010)			answerer reputation	paid-based vs. free-based sites
Jeon et al. (2010)				paid-based vs. free-based sites
Shah and Pomerantz (2010)	answer length	question length, number of answers for the question, number of comments for the question	answerer profile	
Fichman (2011)				site popularity
Shah (2011)	answer time lapse			
Edelman (2012)	answer length, answer time lapse		answerer experience	
Lou et al. (2013)			internal motivation, external motivation	
Oh and Worrall (2013)	answer length		answerer expertise, answerer experience	
Zhang and Wang (2016)		question length		
Lee et al. (2019)	answer politeness			

Appendix B: Robustness Checks

We conducted several additional tests to examine the robustness of our results. First, a potential sample selection bias exists in the data, as not all answers have received helpfulness votes. More importantly, the likelihood of an answer being voted on may be correlated with the explanatory variables that predict answer helpfulness. To account for this potential bias, we employed Heckman's (1979) two-stage selection model as a robustness check. The first stage is a Probit "selection" equation that identifies the determinants of whether an answer was voted on or not. A vivid answer that draws readers' attention is more likely to receive votes, while a pallid answer that fails to draw attention is less likely to receive votes. We included author and answer characteristics that are related to information vividness in this stage, including author expertise, author credibility, answer length, answer readability, answer concreteness and answer emotional intensity (see Kuan et al. 2015). We also included the number of days since the answer was posted. In the second stage, the determinants of an answer's helpfulness are estimated using only voted answers, conditional on the first stage. As shown in Table B1, the inverse Mills ratio was significant ($p < 0.01$) and its inclusion in the stage 2 model alleviates potential bias due to sample selection and endogeneity (Shaver 1998). As shown in Model 2 of Table B1, the results of the second stage of Heckman's model did not qualitatively differ from the results reported earlier.

Table B1: Robustness Check with Heckman's Two-Stage Selection Model

	Model 1 First-stage selection equation	**Model 2** Second-stage outcome equation
Author Expertise	0.511***	0.034***
	(0.052)	(0.003)
Author Credibility	5.523***	-0.001
	(0.089)	(0.001)
Answer Length	0.263***	0.022***
	(0.015)	(0.001)
Answer Reading Difficulty	-0.050***	0.006***
	(0.015)	(0.001)
Answer Total Votes		0.002**
		(0.001)
Answer Days	-0.886***	0.118
	(0.020)	(0.133)

Answer Concreteness	-0.082***	0.001
	(0.012)	(0.001)
Answer Emotional Intensity	-0.081***	0.004***
	(0.011)	(0.001)
Answer Sequence	-0.162***	0.007***
	(0.011)	(0.001)
Other Helpful Answers	0.624***	-0.002
	(0.021)	(0.001)
Question Keywords	0.531***	-0.022***
	(0.022)	(0.002)
Question Length	-0.128***	0.0003
	(0.021)	(0.002)
Question-Answer Concreteness Congruence		0.003**
		(0.001)
Question-Answer Emotion Congruence		0.002**
		(0.001)
Disease-Answer Concreteness Congruence		0.006***
		(0.001)
Disease-Answer Emotion Congruence		0.004***
		(0.001)
Inverse Mills Ratio		-0.056***
		(0.005)
Constant	0.347***	0.338
	(0.016)	(0.394)
N	34894	16726
Chi2		4,612.25

All continuous independent variables standardized; Disease topic dummies, days of week dummies, and month*year dummies included; Standard errors in parentheses; ** $p < 0.05$, *** $p < 0.01$

Second, we utilized a number of alternative models to see if our findings would still hold. One alternative model was a multilevel regression model, chosen because most disease topics in WebMD Answers have more than one posted question. It is expected that answers addressing the same topic will be more similar to each other than to answers not addressing the same topic. Such similarity can cause intra-class correlation (ICC), which results in the standard errors of regression coefficients being underestimated (Klein and Kozlowski 2000; Kreft and Leeuw 1998; Raudenbush and Bryk 2002). Therefore, we utilized a random coefficient multilevel model to account for the interdependence of

individual answers within the same topic (Model 1 of Table B2). Another chosen alternative was a fractional logit model. This approach was deemed appropriate because our dependent variable is a ratio bounded in the range of 0 to 1 (Baum 2008). We utilized fractional logit regressions to analyze the data. As shown in Model 1-2 of Table B2, we found consistent results by using these alternative models.

Table B2: Additional Robustness Checks

	Model 1	Model 2	Model 3	Model 4	Model 5
	Multilevel	Fractional Logit	Tobit	Tobit	Tobit
			#votes >= 3	>= 5	>= 10
Author Expertise	0.041***	0.170***	0.042***	0.042***	0.045***
	(0.003)	(0.016)	(0.003)	(0.003)	(0.004)
Author Credibility	-0.001	-0.004	-0.0002	-0.0001	-0.001
	(0.001)	(0.007)	(0.001)	(0.001)	(0.001)
Answer Length	0.024***	0.108***	0.024***	0.024***	0.022***
	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)
Answer Reading Difficulty	0.006***	0.026***	0.007***	0.008***	0.006***
	(0.001)	(0.006)	(0.001)	(0.001)	(0.001)
Answer Total Votes	0.002*	0.011	0.002**	0.002**	0.002*
	(0.001)	(0.008)	(0.001)	(0.001)	(0.001)
Answer Days	-0.004	0.046	-0.115	-0.118	0.036
	(0.133)	(0.557)	(0.130)	(0.133)	(0.155)
Answer Concreteness	0.0003	0.001	0.0002	0.001	0.0002
	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)
Answer Emotional Intensity	0.003**	0.012***	0.003**	0.002*	0.002
	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)
Answer Sequence	0.006***	0.023***	0.004***	0.002**	0.004***
	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)
Other Helpful Answers	0.003**	0.011**	0.005***	0.006***	0.005***
	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)
Question Keywords	-0.017***	-0.074***	-0.012***	-0.012***	-0.016***
	(0.002)	(0.008)	(0.002)	(0.002)	(0.002)
Question Length	-0.001	-0.004	-0.001	-0.001	0.001
	(0.001)	(0.006)	(0.001)	(0.001)	(0.002)
Question-Answer Concreteness Congruence	0.003***	0.013**	0.002*	0.002*	0.004**
	(0.001)	(0.005)	(0.001)	(0.001)	(0.002)
Question-Answer Emotion Congruence	0.003**	0.011**	0.004***	0.004***	0.003**
	(0.001)	(0.005)	(0.001)	(0.001)	(0.001)
Disease-Answer Concreteness Congruence	0.006***	0.023***	0.006***	0.007***	0.006***

	(0.001)	(0.005)	(0.001)	(0.001)	(0.002)
Disease-Answer Emotion Congruence	0.004***	0.017***	0.004***	0.004***	0.004***
	(0.001)	(0.005)	(0.001)	(0.001)	(0.002)
Disease Topic Dummies	N.A.	Included	Included	Included	Included
Days of Week Dummies	Included	Included	Included	Included	Included
*Month*Year Dummies*	Included	Included	Included	Included	Included
Constant	0.648	0.362	0.946**	0.955**	0.597
	(0.395)	(1.601)	(0.384)	(0.394)	(0.367)
N	16726	16726	15362	13757	9960
Log-likelihood	7,350.61	-7,646.56	7,374.56	7,170.59	5,504.65

All continuous independent variables standardized; Disease topic dummies included in all models except Model 1; Days of week dummies and month*year dummies included in all models; Standard errors in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Third, in the main analyses, we measured our dependent variable by dividing the number of helpful votes by the total number of votes for a given answer. Although this measure was commonly adopted in prior research, its accuracy may suffer when measuring extreme values. For example, assume that two answers both get zero helpful votes, while one answer has ten not-helpful votes, and the other answer has one not-helpful vote. These two answers received the same score for answer helpfulness, but the latter answer may be perceived as more helpful. To reduce this concern, we conducted a set of additional analyses, in which we included only the answers whose total number of votes were at least 3, 5, or 10. As shown in Model 3-5 of Table B2, the results were again consistent with those reported earlier.

Fourth, we alleviate further concerns about possible endogeneity by employing the instrumental variable approach proposed by Lewbel (2012). This approach identifies instruments as simple functions of the observed regression variables when external instrumental variables are not available. Although similar to Arellano and Bond (1991)'s approach using panel data estimators, Lewbel (2012)'s method can be implemented in cross-sectional data sets. Identification hinges on finding regressors that are uncorrelated with the product of heteroskedastic errors. Lewbel (2012)'s approach has been used in many fields such as information systems, marketing, and accounting (e.g., Anderson and Core 2018; Hong et al. 2018; Kashyap and Murtha 2017).

We followed the approach in Lewbel (2012) to generate instruments for our four congruence variables ($Question - Answer\ \widehat{Concreteness\ Congruence}$, $Question - Answer\ \widehat{Emotion\ Congruence}$, $Disease - Answer\ \widehat{Concreteness\ Congruence}$, $Disease - Answer\ \widehat{Emotion\ Congruence}$). We used Hansen's J-statistic of over-identifying restrictions to examine the exogeneity of our four generated instruments and were unable to reject the null hypothesis that our instruments are uncorrelated with the errors (p-value > 0.10). This result suggests that our generated instrumental variables are valid and appropriate. As shown in Table B3, we found consistent results by using these generated instrument variables. Therefore, our earlier results do not appear to be driven by reverse causality or omitted variables.

Table B3: Robustness Check with Lewbel (2012)'s Instrument Variable Approach

	Model 1
Author Expertise	0.041***
	(0.003)
Author Credibility	-0.001
	(0.001)
Answer Length	0.024***
	(0.001)
Answer Reading Difficulty	0.006***
	(0.001)
Answer Total Votes	0.002*
	(0.001)
Answer Days	0.007
	(0.133)
Answer Concreteness	0.0002
	(0.001)
Answer Emotional Intensity	0.003**
	(0.001)
Answer Sequence	0.006***
	(0.001)
Other Helpful Answers	0.003**
	(0.001)
Question Keywords	-0.018***
	(0.002)
Question Length	-0.001
	(0.001)

$Question - Answer\ \widehat{Concreteness}\ Congruence$	0.003**
	(0.001)
$Question - Answer\ \widehat{Emotion}\ Congruence$	0.003**
	(0.001)
$Disease - Answer\ \widehat{Concreteness}\ Congruence$	0.006***
	(0.001)
$Disease - Answer\ \widehat{Emotion}\ Congruence$	0.004***
	(0.001)
Constant	0.599
	(0.395)
N	16726
Log-likelihood	7496.44

All continuous independent variables standardized; Disease topic dummies, days of week dummies, and month*year dummies included; Standard errors in parentheses; ** $p < 0.05$, *** $p < 0.01$