Journal of Official Statistics, Vol. 22, No. 2, 2006, pp. 245–269

Use and Non-use of Clarification Features in Web Surveys

Frederick G. Conrad[1], *Mick P. Couper*[1], *Roger Tourangeau*[1], *and Andrey Peytchev*[2]

Survey respondents misunderstand questions often enough to compromise the quality of their answers. Web surveys promise to improve understanding by making definitions available to respondents when they need clarification. We explore web survey respondents' use of clarification features in two experiments. The first experiment demonstrates that respondents rarely request definitions but are more likely to do so when they realize definitions could be helpful (i.e., definitions are available for technical terms) and when requests involve relatively little effort (i.e., just one click); respondents who obtained a definition requested more subsequent definitions when the initial one proved useful (i.e., included counter-intuitive of surprising information). In the second experiment, definitions available via mouse roll-over were requested substantially more often than when available via clicking, suggesting that some respondents find even a click more effort than they are willing to expend. We conclude with a discussion of interactive features in web surveys in general, when they are likely to be used and when they are likely to be useful.

Key words: Web surveys; question clarification; interactivity; user interfaces; definitions.

1. Introduction

It has long been recognized that survey concepts are not always understood as intended (e.g., Belson 1981). When telephone respondents misunderstand the concepts about which they are questioned, the accuracy of their answers can suffer (Conrad and Schober 2000; Schober and Conrad 1997; Schober, Conrad, and Fricker 2004). This happens often enough to raise concerns about the quality of survey responses with ordinary standardized procedures in which clarification is not available. However, when respondents are able to obtain clarification about the intended meaning of questions their response accuracy can improve dramatically (Conrad and Schober 2000; Schober and Conrad 1997; Schober, Conrad, and Fricker 2004). It is a relatively simple matter to make definitions available in web surveys by linking them to the corresponding words in questions. Respondents need only click on a link to obtain a definition. But will they use this type of feature when it is made available? The current article explores the conditions under which web respondents use and do not use on-line definitions and examines the impact on their answers when they do consult them.

[1] University of Michigan Institute for Social Research, PO Box 1248, Ann Arbor, MI 48106, U.S.A. and the University of Maryland, Joint Program in Survey Methodology. Emails: fconrad@isr.umich.edu, mcouper@isr.umich.edu, and rtourang@survey.umd.edu
[2] University of Michigan, Institute for Social Research, PO Box 1248, Ann Arbor, MI 48106, U.S.A. Email: andrey@umich.edu
Acknowledgment: We thank Reg Baker, Mirta Galesic, Max Louwerse, Duston Pope, and Jenefer Willem for advice and assistance. We gratefully acknowledge support from The National Science Foundation (grant # SES-0106222) and The National Institutes for Health (grant # R01 HD041386-01A1). Please address correspondence to Frederick Conrad, Institute for Social Research, University of Michigan, 436 Thompson Street, Ann Arbor, MI, 48103 U.S.A. Email: fconrad@isr.umich.edu

In principle, web respondents' understanding of questions should benefit from the availability of definitions much like telephone respondents' understanding does when they are able to obtain definitions. In fact, in some ways, the web promises improvements in measurement over and above what has been observed in telephone interviews. Fowler and Mangione (1990) express concern that an interviewer might inaccurately define concepts (p. 21), and as a result they advocate the use of content-free or "neutral" replies to respondent requests for clarification, for example, "whatever it means to you." There is no danger of definitions being distorted when they are provided verbatim; in web questionnaires, all respondents who request definitions obtain identical textual definitions.

An additional reason why clarification on the web may be superior to the comparable process in interviews is that respondents might be more willing to request a definition from a web-based questionnaire than from a human interviewer. In an interview, the respondents must formulate and articulate a request; on the web, the respondents need only lower their index fingers to click a mouse. And in an interview, respondents may be reluctant to request clarification of ordinary concepts because, although the terms may be used in atypical or ambiguous ways, respondents may feel they should know what the terms mean. It seems unlikely that web respondents will be similarly embarrassed to ask a computer for a definition (although the body of literature on social presence, e.g., Nass and Reeves 1996, suggests computers do trigger similar self-presentation concerns and so could potentially inhibit requests for clarification; see, Tourangeau, Couper, and Steiger 2003 for a discussion of social presence in web surveys).

Whether due to differences in the cognitive demands (formulating the question) or social demands (admitting uncertainty about the meaning of ordinary words), it would seem that laboratory respondents interacting with computers are more inclined to request definitions than their telephone interview counterparts. In one study (Schober, Conrad, and Fricker 2004, Experiment 1), telephone interviewers told laboratory respondents that ordinary words sometimes mean something different to the researchers who write the questions than they mean to respondents and so it might be necessary to request clarification to be sure they understand as intended. Respondents asked interviewers for definitions 47% of the time that questions involved ordinary words like "job" and "furniture" that were ambiguous with respect to fictional scenarios, i.e., occasions when the definitions would help. In a comparable study (Conrad and Schober 1999), laboratory respondents entered answers into computers. Despite similar instructions, identical questions and identical scenarios, they clicked a mouse for definitions on 83% of the same ambiguous occasions, far more than in the interviews reported in Schober et al. (2004). However, when Conrad and Schober (1999) merely instructed respondents that definitions were available if they chose to request them, these respondents clicked for definitions on only 23% of the same ambiguous questions. The implication is that if respondents realize questions are potentially ambiguous and they can be obtained at low cost, then respondents are more likely to request clarification from a computer than from a human interviewer; however, if they do not realize this, there is no advantage from computer administration. In both interviewer- and computer-administered data collection, when respondents did obtain clarification for ambiguous scenarios, response accuracy improved dramatically.

What about web respondents outside the laboratory? Will they take advantage of definitions that can be obtained by clicking? The primary reason that respondents might do so is to be sure that they understand the question as intended. However, they are likely to weigh this against at least three reasons for not using hyperlinked definitions. First, clicking for a definition may require more effort than respondents are willing to expend, even though only a very minor motor action is involved. Second, respondents may not realize that definitions might be useful, i.e., they may not realize that without a definition they might misunderstand the question. Third, after requesting a definition, respondents may decide that it is not useful, e.g., turned out not to affect how they answered, thus inhibiting subsequent requests.

We explore these three factors (ease of use, awareness of potential usefulness, and discovery of actual usefulness) in two web survey experiments. We are interested in how often respondents request definitions and whether doing so affects their answers. In general, if respondents do not request definitions when ordinary terms are used with unconventional meaning, this is likely to leave misconceptions uncorrected; increased use of definitions is likely to promote the intended interpretation and reduce measurement error. We do not measure response accuracy in the current experiments but assume this general connection between use of definitions and question understanding. If respondents do request definitions, we are interested in how the content of the definitions affects the likelihood they will make subsequent requests. In particular, are those respondents who request one definition more likely to make follow-up requests when the definition is useful?

In both experiments, hyperlinked definitions were available for the key terms used in questions about the amount of food and nutrients that respondents consume. The questions were displayed in a grid with a different food or nutrient (we refer to these as "concepts") in each of four rows and a different response option in each of five columns (see, Figure 1a.). The definition for each concept could be requested by a simple mouse action using standard hyperlinks and an accompanying instruction. A question stem, which appeared immediately above the grid, read "How much of the following items do you typically consume?" The response options were "Much less than I should" "Somewhat less than I should" "As much as I should" "Somewhat more than I should" and "Much more than I should." Once respondents had selected answers for all the questions in the grid that they intended to answer, they submitted the entire page by pressing the "Next Screen" button at the bottom.

2. Ease of Use and Usefulness of Definitions

The first experiment was embedded within a web-based questionnaire consisting of about 40 questions on "lifestyle" topics: health, food, travel, cars, etc. The grid of four items with definitions appeared on about the fifth page of the questionnaire for most respondents. The grid was designed to appear virtually the same regardless of browser or browser version. However, the user interface for obtaining definitions, the concepts for which definitions were available, and the type of definitions were varied experimentally between respondents. The four food and nutrition concepts always appeared in the same order.

2.1. User Interface

To explore the impact of ease of requesting definitions on the frequency with which they are obtained we designed the user interface so that one, two or a variable number of clicks was required. More specifically, to obtain a definition with the One-click interface, respondents needed to click on a food or nutrient concept; the corresponding definition would then appear in a small browser window (see, Figure 1b). In the Two-click interface, an initial click displayed a list of terms for which definitions were available (Figure 1c) from which respondents could select one by clicking. Note this is not the same as "double-clicking." In this interface, respondents must single-click on separate links in separate windows. Finally, in the Click-and-scroll interface, clicking on a hyperlinked term displayed a glossary – an alphabetic list – of all definitions, though only the first few were initially visible (Figure 1d). If the corresponding definition was not initially visible, the respondent could scroll to it by clicking in the scroll bar, dragging the slider until the appropriate definition was displayed or by using hardware scrolling devices like a wheel mouse. This last interface led to a variable number of clicks – though at least two – depending on how much scrolling was necessary and whether respondents clicked or dragged[3].

One reason respondents might find even a click to involve more effort than they are willing to expend is because it is not necessary to obtain a definition in order to answer the question, i.e., getting a definition is not on the "critical path" (e.g., Gray, John, and Atwood 1993). If respondents viewed their task as answering accurately – as opposed to just answering – they would be inclined to obtain a definition to be sure they understood the question as the author intended it. However, we suspect they simply view their task as answering the question: it either does not occur to them that they might misunderstand the question or they just are not motivated to check the meaning (see, Schober, Conrad, Ehlen, and Fricker 2003).

In addition, by many analyses of human-computer interaction, a click entails more than just clicking. In particular, each overt user action, of which clicks are one type, is immediately preceded and followed by mental actions that take time. For example, prior to clicking, a respondent needs to decide that the definition might actually help achieve his or her goal and, after retrieving the definition, the respondent might evaluate the results, i.e., "did it move me closer to the goal?" (The reality of such invisible decision making along side overt user actions has been demonstrated numerous times with the GOMS family of techniques developed by Card, Moran, and Newell 1983; see, for example Gray, John, and Atwood 1993.[4]) Thus, in deciding whether or not to request a definition, respondents may consider actions that are purely mental, i.e., not accompanied by overt behavior, as well as overt motor actions.

The time and effort involved in reading a definition and possibly rethinking one's conception of the question's topic could deter requests over and above the effort involved in clicking but this should be constant across the three user interfaces. Thus, in the current

[3] The three user interfaces we used were designed to systematically vary the amount of effort required to obtain a definition. They do not vary the effort involved in reading and interpreting the definitions as this text was the same across interfaces. The Two-click and Click-and-scroll interfaces were intentionally less efficient than the One-click interface and so were not intended as viable, production interfaces.

[4] GOMS stands for Goals (what the person is trying to accomplish), Operators (the set of basic procedures that can be combined to accomplish the task), Methods (particular sequences of particular operators assembled to accomplish the current goal) and Selection Rules (criteria for choosing one method out of many that are applicable for attaining the current goal). GOMS techniques can be used to model and describe human performance in virtually any task, though in practice these have tended to be computer-based tasks, so the application of GOMS techniques extends well beyond web surveys.

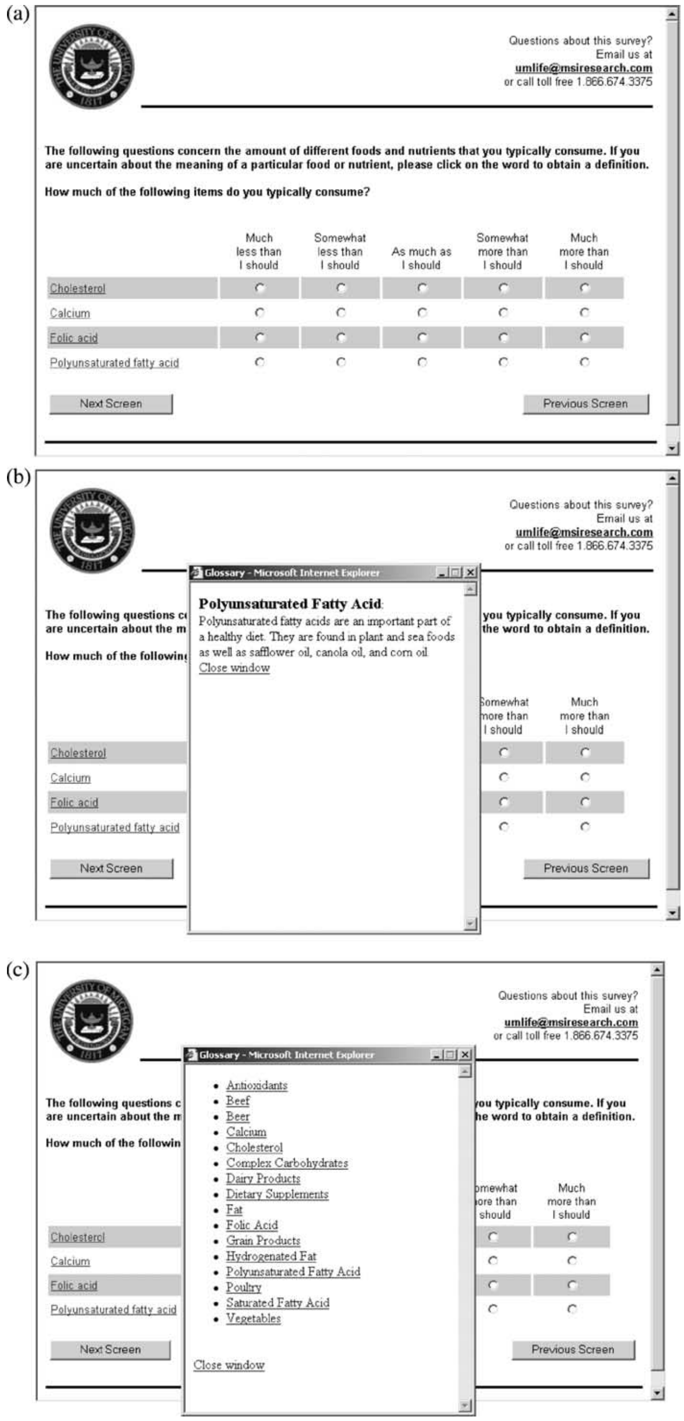

Fig. 1. (a) Grid of four items for which definitions are available. (b) Definition made available by clicking on term in grid (Figure 1a) for One-click interface or on term in list (Figure 1c) for Two-click interface. (c) List of terms for which definitions are made available by clicking on term in grid (Figure 1a) for Two-click interface. (d) Glossary (all definitions for all terms) made available by clicking on term in grid (Figure 1a) in Click-and-scroll interface. If definition is not visible, respondent must scroll to it by clicking in or dragging scroll bar

(d)

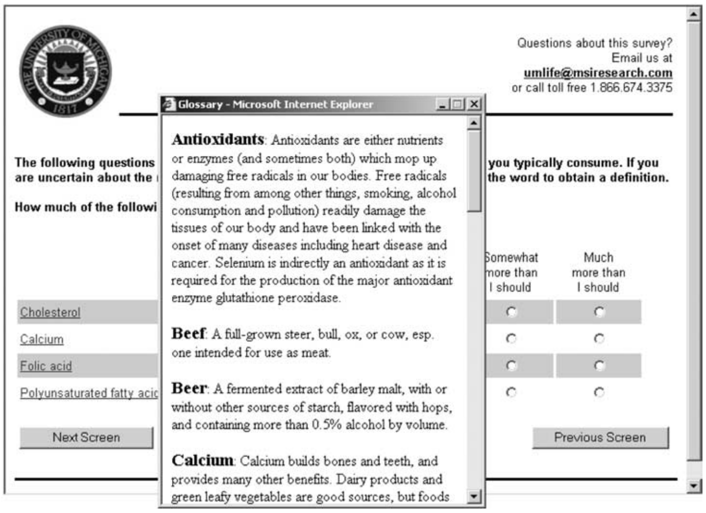

Fig. 1. Continued

study, differences in the number of requests as a function of the user interface can be attributed to the effort involved in clicking.

2.2. Concepts

A precondition for requesting definitions, irrespective of the associated effort, is respondents' belief that the definition will help them achieve some goal, most likely understanding the question as intended in order to answer as accurately as possible. If respondents believe that definitions will not help move them toward their goal(s) – or are not aware of how they might help – there is little reason for them to expend any effort obtaining them. In order to vary respondents' awareness that definitions might help, we asked about their consumption of either technical (e.g., *polyunsaturated fatty acid*) or ordinary (e.g., *vegetables*) food and nutrition. The concepts are listed along with corresponding definitions in Appendix A. If respondents encounter a concept that seems unfamiliar, as is likely to be the case with technical foods and nutrition terms, they should be more likely to request a definition. If this is the case, it suggests that ordinary concepts used in unexpected ways (i.e., the definition departs from ordinary usage) are less likely to be clarified and thus more likely to be misunderstood with respect to the survey objectives than are technical concepts.

For example, in the Current Population Survey question, "How many hours per week do you usually work at your job?" the word "usually" is defined as "50% of the time or more, or the most frequent schedule during the past 4 or 5 months" (U.S. Department of Commerce, 1994). "Usually" is such a common term that there is little reason for respondents to expect it has a technical or unconventional meaning and thus request a definition. A respondent might reasonably assume that the question authors have chosen this word because they believe the respondent will understand it as intended (Clark and Schober 1992 refer to this as the "presumption of interpretability"). For more technical concepts, the respondent might reason that the author of the question would only use such a term if he or she believes the respondent is familiar with the concept so the meaning that comes to mind must be what is intended. Of course if nothing comes to mind, the respondent might assume it is not necessary to know how the concept is defined. In a question of the form "Have you ever had a myocardial infarction?" the respondent might conclude that because the phrase is very unfamiliar the answer must be "no," reasoning that he or she would know what a myocardial infarction was if he or she had had one (Gentner and Collins 1981).

In general, we assume that technical concepts used in survey questions are familiar enough that respondents are likely to conclude the terms are potentially applicable and so will need to understand them in order to answer correctly. However, we selected the technical concepts to be less familiar than the ordinary concepts using frequency in English texts and concreteness as measures of familiarity: more frequent words are more familiar (see Graesser, Cai, Louwerse, and Daniel 2006, for an application of this idea to designing survey questions) and more concrete words are more familiar. See Appendix A for details.

2.3. *Definitions*

Finally, we varied the actual usefulness of definitions by designing half of them to include some counterintuitive or surprising information and half to confirm what most respondents already know or provide no information that would be likely to change their answers. This manipulation, based on our intuitions, makes it possible to test whether requesting a useful definition leads to more follow-up requests than requesting a definition that is not useful. Consider the definition for vegetables. This is intended to be useful in the sense that it includes French fries and potato chips, foods we believed would not often have spontaneously come to mind when respondents think of vegetables ("In general, vegetables include the edible stems, leaves, and roots of a plant. Potatoes, including French fries, mashed potatoes, and potato chips are vegetables").

The definition for beer is designed to be not very useful in that it is consistent with what most respondents already think of when they think of beer, though perhaps somewhat more precise ("A fermented extract of barley malt, with or without other sources of starch, flavored with hops, and containing more than 0.5% alcohol by volume"). The point is that it is likely respondents' answers will be the same whether they read this kind of definition or they do not. We refer to the two kinds of definitions as "useful" and "not useful" respectively. The idea was that, after viewing a definition with relatively surprising information, respondents would recognize that definitions could affect their answers, for example having read the vegetable definition respondents might include their French fry consumption which they might not otherwise have included. The result would be more requests for definitions than when the definitions seem unlikely to affect their answers as in the beer example.[5]

The concepts for which useful and non-useful definitions were available were chosen to be of similar familiarity to respondents so that requests for definitions could not be attributed to differences on this dimension. Thus the concepts for which useful and

[5] A definition could inform the respondent without being useful: if the definition contained information that was new to the respondent but not relevant to the respondent's circumstances, it would not be useful. For example the U.S. Census definition of "residence" goes into detail about boarders and children living away from home in the armed forces, when it is possible these situations do not apply to a particular respondent. This kind of definition might actually discourage subsequent requests for definitions despite being informative.

non-useful definitions were available were chosen to be of equal frequency and concreteness on average (see Appendix A for details). In addition, the average word length per definition was matched across the four sets of definitions.

2.4. *Experimental Design*

Respondents were randomly assigned to one of 12 combinations of interfaces (One-click, Two-clicks and Click-and-scroll), concepts (technical or ordinary) and definitions (useful or not useful). Respondents were presented one of four groups of concepts/definitions (technical/useful, technical/not useful, ordinary/useful, ordinary/ not useful)[6] and the definitions were made available with one of the three user interfaces.

2.4.1. Respondents

Respondents from two commercial opt-in panels were invited by e-mail to answer a questionnaire administered on the web concerning a variety of "lifestyle" topics. Those who completed the questionnaire were entered into a sweepstakes enabling them to win up to $20,000. Of those who were invited to participate, 2,871 respondents completed the questionnaire for a response rate of 18%.[7] Respondents' ages were distributed relatively uniformly: 8.3% were 24 years and under, 20.4% were between 25 and 34 years, 25.1% were between 35 and 44 years, 26.1% were between 45 and 54 years, 14.6% were between 55 and 64 years and 5.5% were 65 years or older. For 25.2% of respondents, the highest level of education was high school or less, 43.5% had completed some college or an associate's degree, and 31.2% had earned a college degree or advanced degree. 43.9% of respondents were male and 56.1% were female.

2.4.2. Questionnaire

Respondents were asked 43 questions about health, nutrition, travel, cars, shopping and other lifestyle topics, as well as a battery of demographic and debriefing items at the end. The items concerned both the respondents' opinions and their behaviors. In general, one item (including grids of items) was presented per page.

2.4.3. Instructions

A standard set of instructions appeared above the four food and nutrition items presented to all respondents: "The following questions concern the amount of food and nutrients that you typically consume. If you are uncertain about the meaning of a particular food or nutrient, please click on the word to obtain a definition. How much of the following items do you typically consume?" These were designed to be relatively neutral in that they do not encourage respondents to obtain definitions nor do they discuss the potential benefits of definitions but simply point out they are available.

[6] In practice, respondents in actual web surveys would likely be presented a mix of technical and ordinary concepts with some useful and some less useful definitions.
[7] This is a low response rate by conventional standards. However, our current goal was simply to be sure that participants were randomly assigned to the different experimental conditions and not to establish their representativeness of any particular population.

2.4.4. Results

One very clear result is that these respondents rarely requested clarification. Overall, only 13.8% of the respondents who answered the experimental questions with definitions (17.4% of respondents who finished the questionnaires) requested any definitions. If this occurs in general, it suggests that many misconceptions may go uncorrected if the definitions that were not requested differ from respondents' ordinary interpretation of the corresponding words. We address this below by examining the effect of usefulness of definitions on answers: if respondents do not obtain definitions that are unusual or counterintuitive, they are likely to misinterpret the question. The low overall rate of requests that we observed may reflect respondents' unwillingness to stray from the critical path, i.e., to do more than the minimum necessary to complete the task, and this could reduce the quality of their answers.

Examining data from those respondents who requested at least one definition, it is apparent that the number of requests is quite sensitive to the amount of effort (number of clicks) involved (see Figure 2). Overall, the number of requests decreased as the number of clicks required by the interface increased, $F(2, 452) = 9.71, p < .001$. When only one click was required, respondents obtained more than 2.5 out of 4 possible definitions but when two or more clicks were required, i.e., with the other two interfaces, respondents obtained closer to 1.5 out of 4 possible definitions (contrast of One-click to other interfaces $F[1, 452] = 64.20, p < .001$); there was no difference in number of requests for the Two-click and Click-and-scroll interfaces ($F[1, 452] = 0.37$, n.s.). Those respondents who had to click twice to get a definition abandoned the request after the first click 36% of the time (383 initial clicks but only 246 second clicks) providing additional evidence that effort (2 clicks versus 1) matters.

Response time patterns further support our contention that effort increases as the amount of clicking required to obtain a definition increases. The server recorded the time each page was submitted (i.e., the time that respondents pressed the "Next Screen" button) so it was possible to determine a global time for answering the four items in the grid by computing the difference between this time and the previous submission. Median response time for those who requested definitions was longer than for those who did not (90 vs 34 seconds, Wilcoxon-Mann-Whitney U, $<.001$). This difference most likely reflects the

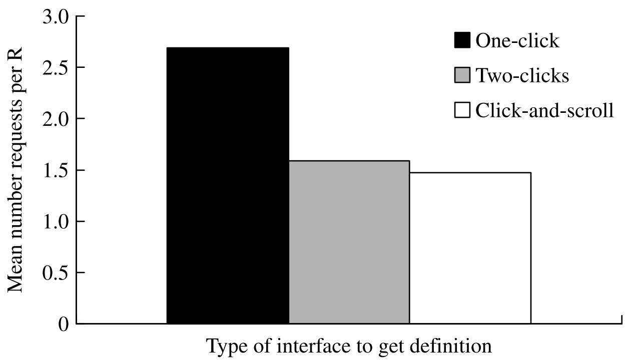

Fig. 2. Mean number of definitions requested (by respondents who requested at least one definition) for three user interfaces. Number of requests reliably larger with One-click than with other interfaces; number of requests with Two-click interface not reliably different than with Click-and-scroll interface

time to read definitions over and above time due to clicking.[8] Looking at just those respondents who requested definitions, as the number of required clicks increased so did the response times. Because the server did not time out as long as the respondent did not break the connection there was a small set of outliers whose times were more than four minutes (33 of the 2679 respondents). We assume that at least some of these times were so long because respondents were engaging in tasks other than completing the questionnaire while the timer continued to increment; thus we excluded them from our analyses. In addition, the data were transformed into log times to reduce the effect of the remaining slow, (i.e., positively skewed), outliers[9]. Figure 3 displays mean response times (transformed back to seconds) for those respondents who requested definitions with each of three interfaces. The overall effect is marginally significant ($F[2, 349] = 3.38$, $p = .094$) and response times for the One-click interface are reliably faster than response times for the other user interfaces ($F[1, 347] = 2.115$, $p = .035$). While the click and scroll interface may sometimes involve just two clicks, on average it involves more actions, either more clicking in the scroll bar or dragging the slider. All of this takes time.

Respondents seemed to recognize the potential value of a definition more often for technical than ordinary concepts: 89% of definitions requested concerned technical concepts. However, requesting a definition for a technical concept was no more likely to lead to follow-up requests than was a request for the definition of an ordinary concept. Respondents who were presented technical concepts and who requested at least one definition requested definitions for 2.32 concepts on average. Respondents who were presented ordinary concepts and requested at least one definition requested definitions for 1.54 concepts on average, $F(1, 452) = 2.51$, n.s.). This equivalence could be because definitions for technical and ordinary concepts were useful equally often in this experiment.

For ordinary concepts – the ones people are more likely to assume they understand without a definition – getting a useful definition led to more total requests for other definitions than did getting a non-useful definition. The pattern is reversed for technical concepts – where people were more likely to believe they might need a definition in the first place, interaction of concept × usefulness $F(1, 452) = 3.79$, $p = .052$. Figure 4 displays the mean number of total requests for both kinds of definitions and both kinds of concepts. It could be that for a complex technical concept, "useful" information, i.e., information that is surprising or counterintuitive, is more than people can assimilate. For example, thinking about cholesterol may be sufficiently difficult so that learning about "good cholesterol" is more than respondents want to think about.

This pattern is actually moderated by the number of clicks needed to obtain a definition, interaction of interface × concept × usefulness, $F(2, 42) = 3.49$, $p = .033$ (see, Figure 5). The patterns of requests are quite different for ordinary (Figure 5a) and technical (Figure 5b) terms. For ordinary terms, when only one click was required, useful definitions were

[8] We cannot rule out the possibility that respondents who do not request definitions are simply faster at answering web questionnaires than those who do. We address this directly in Experiment 2.

[9] Response time distributions tend to be positively skewed because there is a lower bound of zero but no upper bound. It is possible, therefore, that very long but rare response latencies can distort the mean. To reduce the effect of such outliers on the mean, behavioral scientists commonly transform raw response times to make the distribution less skewed. Log transformations are one such approach.

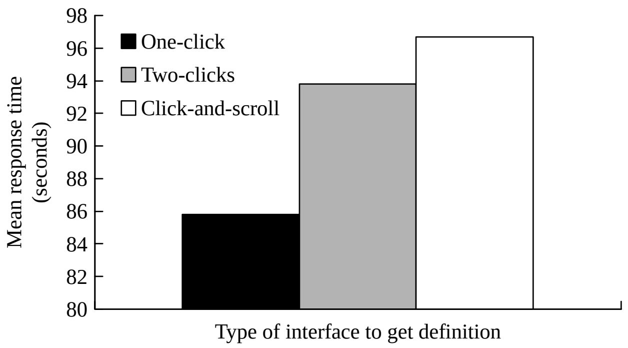

Fig. 3. Mean response time (seconds) for respondents who obtained definitions with the three user interfaces. Responses are reliably faster with the One-click interface than with the others

requested 3.67 out of 4 times on average by respondents who requested any definitions and only 1.39 times out of 4 on average for not useful definitions by respondents who requested any definitions. However, when more than one click was required, respondents rarely requested more than one definition whether useful or not, contrast of usefulness for One-click versus all other combinations of user interface and usefulness, $F(1, 452) = 8.53$, $p < .001$). Apparently, if more than one click is required, it seems that little will convince respondents to request a definition. For technical terms, definitions that were not useful were requested more often than definitions that were useful ($F[1, 388] = 12.73$, $p < .001$), again, presumably because respondents were overwhelmed by useful (counterintuitive) definitions of unfamiliar terms. In addition, the overall number of requests decreased as more clicks were required, irrespective of usefulness ($F[1, 388] = 17.86$, $p < .001$). The general lesson is that if respondents have a reason to question their understanding of words in a question as is the case for technical terms, the content (usefulness) of definitions matters and the ease of obtaining definitions matters in a sensible way. If respondents do not have a reason to question their interpretation, as is more likely for ordinary terms, they are unlikely to request definitions unless it is very easy to do so (one click). Content (usefulness) has an effect in only these cases.

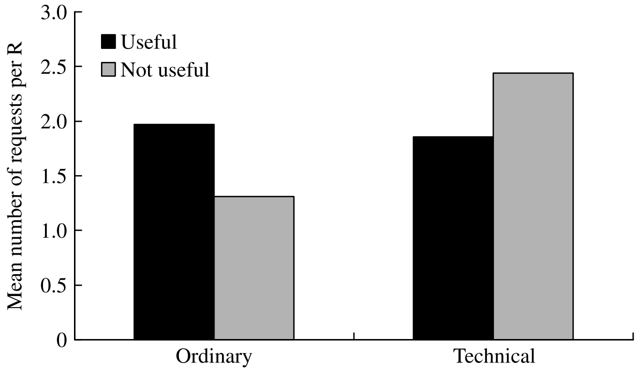

Fig. 4. Mean number of useful and non-useful definitions requested (by respondents who requested at least one definition) for ordinary and technical concepts

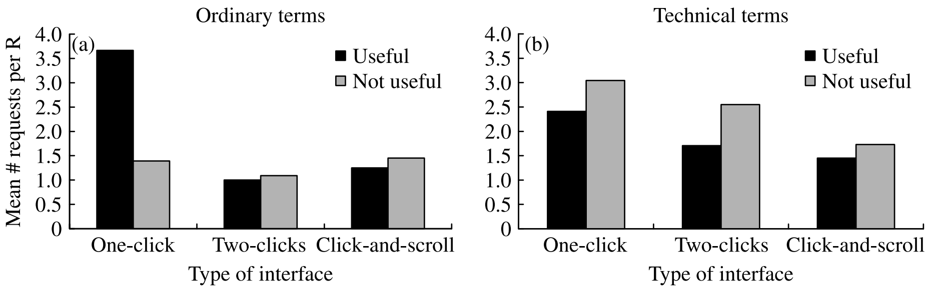

Fig. 5. Mean number of definitions requested per respondent (who requested at least one definition) for (a) ordinary and (b) technical terms that were useful or non-useful, requested with One-click, Two-click or Click-and-scroll user interfaces. In panel (a), respondents requested more definitions of ordinary terms when the definitions were useful if the request involved only a single click; if more clicks were required, requests were rare and usefulness did not affect their frequency (usefulness has a reliably larger effect for One-click than other interfaces). In panel (b), respondents requested definitions for technical terms reliably more often when the definitions were not useful than when they were; overall number of requests decreased reliably as the number of clicks required by the three user interfaces increased

Researchers are motivated to make definitions available because they assume that respondents who obtain them will be more likely to understand questions as intended and thus to respond more accurately. Obtaining hyperlinked definitions certainly led to more accurate responses to factual questions in the Conrad, Schober, and Coiner (forthcoming) laboratory study for ambiguous scenarios. In the current study it is hard to measure (or even define) accuracy because answering the questions involves a subjective judgment about consumption quantities (e.g., "more than I should"). What we *can* determine is whether respondents who obtain definitions answer differently from those who do not. If this is the case, it would suggest that definitions are aligning respondents' understanding with the meanings that the question authors intended which in some cases leads to different answers.

To explore this possibility we examine response distributions for questions whose definitions were requested relatively often. Six technical concepts elicited at least 50 requests for clarification. The five consumption categories were coded numerically, with 1 representing "less than I should" and 5 representing "more than I should." Consumption scores for three out of the six concepts differed reliably for respondents who obtained definitions from those who did not and, for one additional concept, the scores were marginally different (see Table 1). In all cases, respondents who obtained these definitions registered reliably lower consumption scores (i.e., less than they should) than when they did not obtain the definition.

Differences in consumption scores for respondents who did and did not request definitions suggest that respondents read the definitions and took them into account when deciding how to answer. The definitions for Antioxidants and Polyunsaturated Fatty Acid, which were associated with reliably lower scores, include information about the health benefits of the corresponding food or nutrition. It is sensible therefore that those respondents who had read them would be more likely to report consuming less of these substances than they should. However, the definition for Hydrogenated Fat was truly uninformative with respect to the health consequences of its consumption, yet also led to lower scores. It could be that after reading other definitions which mentioned the benefits of the corresponding substance, respondents assumed that this substance was also

Table 1. Average consumption scores (1 = Much less than I should; 5 = Much more than I should) for respondents who obtained and did not obtain definitions (number of observations in parentheses) in Experiment 1

Concept	Definition	
	Did not obtain	Obtained
Hydrogenated fat*	3.36 (524)	3.08 (105)
Antioxidants*	2.51 (547)	2.06 (82)
Polyunsaturated fatty acid*	3.17 (555)	2.40 (96)
Saturated fatty acid**	3.44 (533)	3.23 (93)
Complex carbohydrates	3.11 (484)	3.08 (146)
Folic acid	2.59 (579)	2.46 (69)

*$p < .01$; **$p = .06$.

beneficial and so adjusted their scores downward. Or it could be that respondents who requested definitions happened to interpret the response scales differently than those who did not obtain definitions – or were different in some other way. Although there is no theoretical reason why this should be the case, we cannot rule it out on the basis of the current experimental design.

2.4.5. Summary of Results

In the current study, web survey respondents requested definitions infrequently even though the request involved only a small number of mouse clicks. When they did request definitions, respondents were substantially more likely to do so for technical than ordinary concepts. If the request involved a single click, respondents who requested one definition obtained more follow-up definitions than when two or more clicks were required. Overall response times were faster when definitions were obtained with one click than with multiple clicks, consistent with the idea that clicking is effortful. If the content of the definition was useful, i.e., contained surprising information, respondents requested definitions for more ordinary concepts than if the definitions were less useful; usefulness had the opposite effect for technical concepts. However, the effect of usefulness only led to a larger number of requests when the request involved one click, indicating that effort is the main determinant of using the definition feature. Finally, for several items, respondents who obtained definitions answered differently than respondents who did not, suggesting that they read and considered the content of the definitions in deciding how to answer.

3. Very Low Effort Methods of Obtaining Definitions

Although respondents can ensure that they understand questions as intended by obtaining definitions, the benefits of this may be outweighed by the effort involved. Even a single mouse click seems to involve more effort than respondents are willing to expend except under special conditions (e.g., technical concepts, useful definitions for ordinary concepts). We have suggested that there is more to clicking than the simple motor action involved (e.g., anticipating and then evaluating the consequences of clicking, not to mention reading the definitions once they have been displayed), but the act of clicking itself is at least one component of the effort. If this is so, then it should be possible to increase requests for definitions by making them available through simpler respondent actions.

One interface technique that would enable respondents to request a definition without clicking the mouse is the "rollover." In our version of the technique, the respondent positions the mouse pointer over a term in the grid and this is interpreted as a request for a definition. The system then displays the definition as "hovering text" in front of the questionnaire, much as help is often provided in conventional desktop programs. Because this action – positioning the pointer over the term – is a prerequisite for clicking, the rollover clearly involves fewer actions by the respondent than the One-click interface in Experiment 1. Nonetheless it does still require the respondent to move the mouse and even this could feel effortful given that it is off the critical path.

In Experiment 2, we test whether rollovers are used more frequently than the One-click or Two-click interface to request definitions. The Rollover interface was designed so that as soon as the mouse pointer was positioned over one of the terms in the grid, the definition appeared as hovering text (see Figure 6).

In addition we tried to manipulate how respondents viewed obtaining definitions – on the critical path or not – by varying the instructions. One set of instructions encouraged respondents to obtain definitions in order to be sure they had correctly interpreted the question; we refer to these as "encouraging" instructions: "Ordinary words can be used in unexpected ways in surveys so, to be sure you understand them, please request more information." Another set of instructions simply indicated that definitions were available just as in the first experiment; we refer to these as neutral instructions. Both sets of instructions described how to request a definition appropriate to the rollover or click interfaces.

Finally, we also vary the usefulness of the definitions as in the first experiment (see Appendix B). All of the concepts were (relatively) ordinary terms in the current study (cholesterol and calcium being perhaps slightly more technical than the others but certainly in

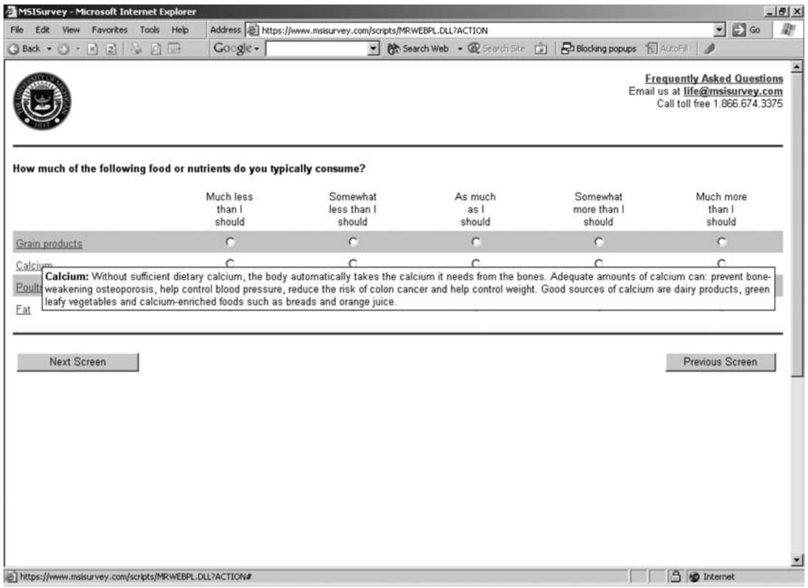

Fig. 6. Definition displayed in response to rollover, Experiment 2

everyday use). This was done to make the test of the experimental factors more stringent given that in Experiment 1 respondents requested definitions even less frequently for ordinary than for technical concepts. We selected eight concepts and developed a useful and not useful definition for each. The two versions of each definition were matched in word length.

The eight concepts appeared in two grids of four rows for questions and five columns for response options, modeled on the grid used in Experiment 1. The order of the concepts in each grid, i.e., which row each concept appeared in, was randomized for each respondent.

3.1. Experimental Design

Respondents were randomly assigned to one of three interface conditions (Rollover, One-click and Two-clicks), one of two instruction conditions (Encouraging or Neutral), and one of two definition types (Useful or Not Useful Definitions). This created 12 experimental groups. In addition, a group of respondents roughly equal in size to each of the 12 experimental groups was randomly assigned to a No Definitions control condition.

3.1.1. Respondents

Half of the respondents were recruited from one of the opt-in panels used in Experiment 1 and half were recruited from AOL's "opinion place," a service that routes AOL members to web surveys through banner advertisements. The opt-in panel members were invited by e-mail to answer a questionnaire administered on the web concerning a variety of "lifestyle" topics, much like the procedure in Experiment 1. The AOL participants clicked on a banner advertisement which makes it impossible to know how many potential respondents chose not to participate. As a result we cannot determine the response rate for the current experiment. 3,050 respondents completed the definition questions. It is possible that some of the respondents in Experiment 2 had also been respondents in Experiment 1. However, the large size of the panel makes this unlikely. In addition, the survey in which the second experiment was embedded was administered about 18 months after the survey in which the first experiment was embedded, further reducing the likelihood of overlap.

3.1.2. Results

Requests for clarification were relatively infrequent but more frequent than in Experiment 1. Overall, 22.4% of respondents who could request clarification did so in contrast to the 13.8% figure in the first experiment. This difference is largely due to the Rollover interface: four times as many respondents obtained definitions by means of rollovers (36.5%) as did so by clicking (8.9% and 6.5% with the One-click and Two-click[10] interface respectively), $X^2(2, N = 2,851) = 369$, $p < .001$. Those who requested at least one definition with the Rollover interface requested definitions for more concepts on average (2.29) than did their counterparts using the One-click (1.84) and Two-click (1.44) interfaces, $F(2, 2848) = 180.17$, $p < .001$.

While it is clear that more respondents request more definitions with rollovers than by clicking, it could be that many of the rollover requests are less deliberate than requests

[10] Note that 26 of the 89 respondents who clicked for a definition using the Two-click interface abandoned requests for definitions after clicking once. The 6.5% figure does not include these cases.

registered by clicking. For example, respondents could move the mouse over the "hot" region in order to respond by clicking a radio button and this would be interpreted by the system as a request for a definition even though one was not intended; alternatively respondents who trace the text they are reading with the mouse pointer might unintentionally invoke a rollover definition. To click the mouse, in contrast, the respondent must intentionally press the button. If the difference in number of requests is due to differences of intention, respondents should be less likely to read definitions displayed via the Rollover interface than would their counterparts who clicked. Less frequent reading of definitions would reduce the effect of definitions on responses in the Rollover condition, relative to the One-click and Two-click interfaces. However, this does not seem to be the case. Consumption scores differed reliably for six out of the eight concepts when respondents obtained useful versus non-useful definitions with rollovers (see Table 2). This suggests respondents read the definitions they obtained with rollovers and took them into account when deciding how to answer.

In fact, the effect of the definitions' content is greater for rollovers than clicks. In the One-click and Two-click interfaces, the effect is only visible for one concept, dietary supplements. So requesting definitions with rollovers seems deliberate. However, even if some requests are accidental, respondents in general seem to be incorporating the content of definitions into their answers.

Consistent with the idea that less effort is involved with rollovers than clicking, response times were shorter when respondents had obtained definitions with the Rollover than One-click and Two-Click interfaces. Mean response time for a grid of questions when a definition was requested was 79.9 seconds for the Rollover interface, 107.4 seconds for the One-click interface and 99.9 seconds for the Two-click interface (overall $F(2, 512) = 15.41$, $p < .001$; contrast of Rollover to other two interfaces $F(1, 513) = 29.7$, $p < .001$). This advantage for the Rollover interface is observed despite the greater number of requests and concomitant reading time with this interface than with the click interfaces.

Despite the clear reduction in effort introduced by the Rollover interface, many respondents did not request any definitions. Our attempt to frame the task with

Table 2. Average consumption scores (1 = Much less than I should; 5 = Much more than I should) for respondents who obtained non-useful and useful definitions (number of observations in parentheses) with Rollover interface in Experiment 2

Concept	Definition	
	Non-useful	Useful
Fat**	3.63 (35)	2.81 (36)
Dietary supplements**	2.19 (109)	3.47 (73)
Grain products**	2.87 (71)	2.4 (100)
Poultry	2.85 (47)	2.67 (46)
Vegetables	2.36 (22)	2.11 (27)
Dairy products*	2.69 (39)	2.04 (45)
Cholesterol*	3.17 (36)	2.70 (51)
Calcium**	2.23 (22)	3.35 (57)

$*p < .05$; $**p < .001$.

encouraging instructions as one in which accurate understanding was an essential component had no effect. Only 18.1% of respondents who were presented encouraging instructions requested any definitions, not reliably different from the 16.8% who requested definitions after being presented neutral instructions ($\chi^2[1] = 0.780$, n.s.) and those who requested definitions did not request more of them as a function of instructions ($t[504] = 0.684$, n.s.). Instructions did not affect the use of definitions differently for the different interfaces, interaction of instructions and interface for number of clicks, ($F[2, 2845] = 0.323$, $p = .724$). This suggests that most web respondents are simply not inclined to use clarification features, presumably because they do not consider them to be essential for adequately completing the task as they conceive of it. This non-use could also be prevalent for respondents filling out paper questionnaires but, unlike web questionnaires, there is no effective way to measure definition use for paper.

It seems likely that some respondents are simply more inclined to inquire about the meaning of survey concepts than are others, irrespective of the instructions, the interface, or the content of definitions. Perhaps it is a matter of how curious or reflective or motivated to be accurate different respondents are. The more curious, reflective or motivated a respondent, the longer it should take the respondent to answer because the respondent is more prone to ponder the question meaning relative to the respondent less curious, reflective or motivated counterpart who simply answers without belaboring the process.

We can test this idea by comparing the latencies for respondents who do not request definitions but could do so if they chose to, i.e., they were assigned to one of the three interface conditions, to the latencies for respondents assigned to the No Definition control group. The former group of respondents should be relatively fast because they respond without questioning their interpretation, i.e., they are less curious, etc. The latter (control group) should include respondents who would have requested definitions if they could have as well as respondents who would not have. As a result, this group should be comprised of both slower and faster respondents. On average, the control group should respond more slowly than should those who did not request definitions despite their availability. This is exactly what we observed. Respondents who did not request definitions but could have, answered the items in 60.3 seconds on average whereas the average response time for the No Definition control group was 65.5 seconds, $F(1, 2508) = 4.90$, $p < .05$.

3.1.3. Replication

One question about the results is to what extent they generalize. The respondents recruited from the SSI panels are likely to differ on many characteristics from members of the general population such as age, education and web use. Although those from the AOL Opinion Place may be more typical of the general population, they were not selected randomly. In principle, both groups might use definitions differently than would a more representative sample of respondents, even if there is no theoretical reason to expect this to be the case. As a check on this possibility we conducted a small replication using the Knowledge Networks (KN) panel. These panel members are recruited through Random Digit Dialing techniques so they are more likely to resemble the general population than our other sample members. While KN panel members need not have internet access to be in the panel (KN provides web TV to those without access), we recruited a sample with access in order to keep the user interface was comparable to that used in the main experiment. We randomly assigned 941 KN panel members to the One-click, Two-click and Rollover user interfaces. In addition, 78 respondents were not able to obtain definitions, providing a control condition. KN respondents were exposed to the same concepts, definitions and grid format used in the main experiment.

The main finding is that rollover requests for definitions were more frequent than requests with either click interface, just as was the case with SSI and AOL respondents. Respondents with the rollover interface requested definitions for 52.5% of all concepts in contrast to 12.3% and 16.5% for respondents using the One-click and Two-click interfaces, respectively ($\chi^2[2] = 152.7,\ p < .0001$). In addition, respondents using the rollover interface answered more quickly than their counterparts using click-interfaces, similar to the pattern observed with SSI and AOL respondents. Average response times were 37.35 seconds for Roll-over respondents and 53.48 and 45.25 seconds for One-click and Two-click respondents, respectively, $F(2, 204) = 5.96,\ p < .001$). This pattern extends the original findings to different and presumably more representative respondents.

3.1.4. Summary of Results

Experiment 2 demonstrated that respondents obtain definitions relatively often if it is easy enough to do so. Rollovers led to more than a four-fold increase in requests for definitions over the One-click and Two-click interfaces. For respondents who requested any definitions, rollovers led to requests for definitions of more concepts than did either of the click interfaces. Respondents' answers were more affected by the content (usefulness) of definitions obtained with rollovers than by clicking, suggesting that they took the definitions into account when deciding how to answer even though accidental rollovers were possible. Respondents who obtained definitions with rollovers responded more quickly overall than did those who obtained definitions by clicking, even though respondents in the rollover group had more definitions to think about than did their counterparts in the click interface groups. This emphasizes the greater ease of using of rollovers than click interfaces for clarification.

In contrast to ease of obtaining definitions, instructions designed to encourage the use of clarification had no effect on frequency of requests relative to neutral instructions. It seems that some respondents may simply be disinclined to obtain definitions because they reflect less on the meaning of questions than do other respondents. Such respondents should answer faster than those who give more thought to such issues, and this seemed to be the case: respondents who did not use the clarification feature when it was available answered more quickly than respondents who might have used it had it been available. Finally, a small replication with a probability sample reproduced the pattern of requests and response times observed in the main experiment, generalizing the results beyond the current testing situation.

4. General Discussion

Survey respondents can provide acceptable answers in good faith without ever verifying that they have understood the questions exactly as intended. However, at least some respondents who do not obtain definitions will misunderstand the question. While a click would seem to be as low an effort as one can imagine, respondents in the current study did not click for definitions very often. One explanation is that any motor action involves more effort than retrieving what one already knows. In particular, if a definition is available for an ordinary term, it is easier to apply one's everyday meaning than to click for a definition which must then be read and incorporated into the response process. Moreover, obtaining a definition could require one to rethink what one already believed to be established which could further deter the use of a clarification feature.

This point is underscored by a demonstration in Gray and Fu (2004). Experimental participants were asked to program an on-screen VCR to record several TV shows, entering start and stop times for each show. Participants had an opportunity to view start and stop times on the screen prior to beginning the programming task, leaving some trace of this information in their memories. Once they started to program each show, the start and stop times were obscured for some participants but could be brought to the foreground with a click. For others, the show information remained in full view so could be seen with only a minor eye movement. As it turned out participants were relatively unlikely to access the show information – either by clicking or moving their eyes – while programming the VCR, relying more often on what they remembered. It seems that people are reluctant to engage in the simplest motor actions – even eye movements – when they already know something about the topic, much like our respondents knew the conventional meanings of ordinary words. Nonetheless, participants in this study did access the show information more (though still not often) by moving their eyes than by clicking, suggesting that their behavior depended on both knowledge and ease of use.

This is in line with our observation that rollover requests are more frequent and have a larger effect on answers than click requests. It may be that beyond ease of use, rollovers are effective because the sudden appearance of the hovering text in respondents' field of view is hard to ignore and makes it possible for respondents to consult the definition without even moving their eyes. This raises the possibility of simply presenting definitions within the questions by default. We suspect that text that is "always on" would not lead more respondents to read definitions because while the eye movements are low effort the presentation of the text is not under respondents' control and this kind of control may be essential if respondents are to consult definitions.

These results almost certainly extend beyond on-line definitions and even beyond web surveys to web use in general. People seem to be impatient when they use the web, perhaps because of the vast amount of information that is available through very minor actions, e.g., pressing a mouse button or just moving a mouse pointer. For example, Hert and Marchionini (1997, Section 4.3.2) observed that a substantial proportion of visits to a U.S. Federal web site involved just one page (the proportion of one-page sessions ranged from 22%–52% depending on the content area of the site) suggesting that if the information users sought was not immediately available they quickly went elsewhere. Yet unsolicited information that is easily available, e.g., pop-ups, banners, etc., are routinely ignored or actively suppressed by users. Users seem to desire both immediacy and control.

In the end, many respondents (and web users generally) may simply choose not to use certain interactive features. Simply creating an interactive feature for web survey respondents does not guarantee they will use it. Landauer (1995) used the phrase "creeping featurism" to describe the phenomenon of including features in software because designers believe they will make the product more competitive but not because they are helpful to users. He describes a survey of one software company's user base that found that less than one third of the available features in the company's products were ever used. Presumably many of those used were used only once, as we observed for definitions that are not useful. Web survey designers should carefully consider the way respondents think about their task before providing features they believe will help.

Appendix A

The concepts for which definitions appeared are grouped according to two factors: technicality, i.e., whether the terms were ordinary or technical, and the usefulness of the definitions available for each term, i.e., whether the definition was useful or non-useful. The four groups of concepts and their experimental definitions appear below.

Measures of word frequency (instances per million in a variety of English texts) and rated concreteness (100–700 with higher numbers reflecting greater concreteness) were obtained for each concept using Coh-Metrix (Graesser, McNamara, Louwerse, and Cai 2004; Louwerse, McCarthy, McNamara, and Graesser 2004). The concepts were chosen so that those for which useful definitions were available and those for which non-useful definitions were available were equally frequent and concrete (i.e., not reliably different). The ordinary concepts were chosen to be more frequent (33.1 versus 8.2, $F[1, 15] = 7.303$, $p = 0.019$), and concrete (427.9 versus 119.9, $F[1, 15] = 6.199$, $p = 0.028$) than the technical concepts. The interaction of technicality and usefulness was not significant for either frequency or concreteness.

A.1. Ordinary Concepts/Non-Useful Definitions

Fat – A chemical compound containing one or more fatty acids. Fat is one of the three main constituents of food (the others are protein and carbohydrate). It is also the principal form in which energy is stored in the body.
Dietary supplements – Products (other than tobacco) intended to supplement the diet that bears or contains one or more of the following dietary ingredients: a vitamin, mineral, amino acid, herb or other botanical.
Beer – A fermented extract of barley malt, with or without other sources of starch, flavored with hops, and containing more than 0.5% alcohol by volume.
Beef – A full-grown steer, bull, ox, or cow, especially one intended for use as meat.

A.2. Ordinary Concepts/Useful Definitions

Dairy products – In addition to milk, cream, ice cream, butter, cheese, and yogurt include frozen yogurt, creamy soups like cream of mushroom, and cheesy foods like pizza and lasagna.

Poultry –Although often low in fat and cholesterol, when poultry is fried or eaten with its skin it can be high in fat and cholesterol. Include chickens, turkeys, ducks, geese and other game birds. Also, include eggs.
Vegetables – In general, vegetables include the edible stems, leaves, and roots of a plant. Potatoes, including French fries, mashed potatoes, and potato chips are vegetables.
Grain products – Bread and foods made with bread such as French toast and stuffing, baked goods (such as muffins and cake), cereals, popcorn, rice, and pasta.

A.3. Technical Concepts/Non-Useful Definitions

Complex carbohydrates – A large group of sugars, starches, celluloses, and gums that contain carbon, hydrogen, and oxygen in similar proportions. They are a good source of minerals, vitamins, and fiber.
Hydrogenated fat – A fat whose molecular structure has been modified by the addition of hydrogen. Vegetable shortening and margarine are hydrogenated fats.
Saturated fatty acid – In saturated fatty acid, the carbon atoms are bonded with single bonds; they share one set of electrons. Saturated fatty acids are mostly found in animal products.
Antioxidants – Antioxidants are either nutrients or enzymes (and sometimes both) which mop up damaging free radicals in our bodies. Free radicals (resulting from among other things, smoking, alcohol consumption and pollution) readily damage the tissues of our body and have been linked with the onset of many diseases including heart disease and cancer. Selenium is indirectly an antioxidant as it is required for the production of the major antioxidant enzyme glutathione peroxidase.

A.4. Technical Concepts/Useful Definitions

Cholesterol – Good cholesterol (HDL) carries bad cholesterol (LDL) away from the arteries. Levels of HDL are elevated by consuming vitamin C and niacin, exercising and not smoking. Levels LDL are lowered by limiting consumption of foods such as beef and rich cheeses.
Calcium – Calcium builds bones and teeth, and provides many other benefits. Dairy products and green leafy vegetables are good sources, but foods such as breads and orange juice are calcium-enriched. For example, one English muffin has more than 1/3 as much calcium as a cup of skim milk.
Folic acid – Folic acid promotes cell growth and may prevent birth defects. It is found in avocado, bananas, orange juice, asparagus, leafy vegetables, dried beans and peas and is added to commercial grain products.
Polyunsaturated fatty acid – Polyunsaturated fatty are an important part of a healthy diet. They are found in plant and sea foods as well as safflower oil, canola oil, and corn oil.

Appendix B

Concept	Definition	
	Not useful	Useful
	Grid 1	
Fat	Fat is a chemical compound containing one or more fatty acids. It is one of the three main constituents of food (the two others are protein and carbohydrate). In some recipes, "fat" refers to solid vegetable shortening.	Fat is an essential part of the diet. It is the most concentrated source of energy we get, it supplies us with essential fatty acids, it promotes absorption of fat-soluble vitamins and it helps maintain healthy skin.
Dietary supplement	Products intended to supplement the diet, containing one or more of the following: a vitamin, mineral, amino acid, herb or other botanicals. Dietary supplements are found in many forms such as tablets, capsules, softgels, gelcaps, liquids, or powders. Available in health food stores, grocery stores, pharmacies, and by mail.	A multivitamin supplement taken daily is recommended to help insure adequate levels of necessary vitamins and micronutrients. In addition, dietary supplements help protect cells against aging (Vitamins C and E), improve sexual performance (Zinc and Vitamins C and E) and reduce stress (Omega-3 fatty acids, Valerian), among other benefits.
Grain products	Cereals, bread (from wheat or rice flour), pasta and rice. Typically high in complex carbohydrates.	Bread and foods made with bread, including muffins, French toast, stuffing, popcorn, and presweetened cereals.
Poultry	Domestic fowl raised for either meat or eggs. Include chickens, turkeys, ducks, geese, guinea fowl, black swan, peafowl, pigeon, pheasant, quail, peacock and other game birds. The chief poultry bird is the chicken and the dominant chicken today is a cross between the fast-growing female White Plymouth Rock chicken, and the deep-breasted male Cornish chicken.	Chicken, the most popular poultry bird, is less healthy than one might think. Pesticides and fungicides in chicken feed are passed to humans and could damage our central nervous system and cardiovascular system. Scientists have linked contaminated poultry to numerous cases of salmonella each year. Eating chicken is generally no healthier than eating lean beef.

Appendix B. Continued

Concept	*Definition*	
	Not useful	*Useful*
	Grid 2	
Vegetables	In general, vegetables include the edible stems, leaves, and roots of a plant. Include lettuce and other salad greens, carrots, cucumbers, celery and tomatoes.	In general, vegetables include the edible stems, leaves, and roots of a plant. Potatoes, including French fries, mashed potatoes, and potato chips are vegetables.
Dairy products	Any product manufactured from milk or any derivative or product of milk. Usually refers to milk from cows (bovine) but also refers to milk from goats, sheep, reindeer, and water buffalo. Include all forms of milk (e.g., skim, reduced fat, whole, powdered, evaporated and condensed) and cream (including half and half). Include cheese and yogurt and whey.	The consumption of cow's milk may not be as healthy as has traditionally been assumed. For example, international studies now show a strong correlation between the use of dairy products and the incidence of diabetes; milk proteins are among the most common causes of food allergies; and milk sugar has been implicated in ovarian cancer and cataracts.
Cholesterol	Cholesterol is a soft, waxy substance found among the lipids (fats) in the bloodstream and in all the body's cells. Too high a level of cholesterol is a major risk for coronary heart disease. It's also a risk factor for stroke.	Good cholesterol (HDL) carries bad cholesterol (LDL) away from the arteries. One can increase levels of HDL by consuming vitamin C and niacin, exercising and not smoking. One can lower LDL by consuming fewer foods such as beef and rich cheeses.
Calcium	Calcium is a soft silver-white element that is an alkaline earth metal constituting about three percent of the earth's crust. Thus it is the fifth most abundant element in the earth's crust occurring especially as calcium carbonate. Limestone is primarily calcium carbonate. Calcium is present in many foods and is used as a deoxidizer in steel.	Without sufficient dietary calcium, the body automatically takes the calcium it needs from the bones. Adequate amounts of calcium can prevent bone-weakening osteoporosis, help control blood pressure, reduce the risk of colon cancer and help control weight. Good sources of calcium are dairy products, green leafy vegetables and calcium-enriched foods such as breads and orange juice.

5. References

Belson, W.A. (1981). The Design and Understanding of Survey Questions. Aldershot, UK: Gower.

Card, S.K., Moran, T.P., and Newell, A. (1983). The Psychology of Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum, Inc. Publishers.

Clark, H.H. and Schober, M.F. (1992). Asking Questions and Influencing Answers. In Tanur, J.M. (ed.), Questions about Questions: Inquiries into the Cognitive Bases of Surveys. New York: Russell Sage Foundation, 15–48.

Conrad, F.G. and Schober, M.F. (2000). Clarifying Question Meaning in a Household Telephone Survey. Public Opinion Quarterly, 64, 1–28.

Conrad, F.G., Schober, M.F., and Coiner, T. (in press). Bringing Features of Human Dialogue to Web Surveys. Applied Cognitive Psychology.

Fowler, F.J. and Mangione, T.W. (1990). Standardized Survey Interviewing: Minimizing Interviewer-related Error. Newbury Park, CA: Sage Publications, Inc.

Gentner, D. and Collins, A. (1981). Studies of Inference from Lack of Knowledge. Memory and Cognition, 9, 434–443.

Graesser, A.C., Cai, Z., Louwerse, M., and Daniel, F. (2006). Question Understanding Aid (QUAID): A Web Facility that Tests Question Comprehensibility. Public Opinion Quarterly, 70, 3–22.

Graesser, A., McNamara, D.S., Louwerse, M., and Cai, Z. (2004). Coh-Metrix: Analysis of Text on Cohesion and Language. Behavioral Research Methods, Instruments, and Computers, 36, 193–202.

Gray, W.D. and Fu, W. (2004). Soft Constraints in Interactive Behavior: The Case of Ignoring Perfect Knowledge In-the-world for Imperfect Knowledge In-the-head. Cognitive Science, 28, 359–382.

Gray, W.D., John, B.E., and Atwood, M.E. (1993). Project Ernestine: Validating a GOMS Analysis for Predicting and Explaining Real-world Performance. Human-Computer Interaction, 8(3), 237–309.

Hert, C.A. and Marchionini, G. (1997). Seeking Statistical Information in Federal Websites: Users, Tasks, Strategies, and Design Recommendations. http://www.ils.unc.edu/~march/blsreport/mainbls.html

Landauer, T.K. (1995). The Trouble with Computers: Usefulness, Usability and Productivity. Cambridge, MA: MIT Press.

Louwerse, M.M., McCarthy, P.M., McNamara, D.S., and Graesser, A.C. (2004). Variation in Language and Cohesion Across Written and Spoken Registers. In K. Forbus, D. Gentner, and T. Regier (eds), Proceedings of the 26th Annual Meeting of the Cognitive Science Society, 843–848. Mahwah, NJ: Erlbaum.

Reeves, B. and Nass, C. (1996). The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places. Stanford, California: CSLI Publications.

Schober, M.F. and Conrad, F.G. (1997). Does Conversational Interviewing Reduce Survey Measurement Error? Public Opinion Quarterly, 61, 576–602.

Schober, M., Conrad, F., Ehlen, P., and Fricker, S. (2003). How Web Surveys Differ from Other Kinds of User Interfaces. Proceedings of the American Statistical Association,

Section on Survey Research Methods. Alexandria, VA: American Statistical Association.
Schober, M.F., Conrad, F.G., and Fricker, S.S. (2004). Misunderstanding Standardized Language. Applied Cognitive Psychology, 18, 169–188.
Tourangeau, R., Couper, M.P., and Steiger, D.M. (2003). Humanizing Self-administered Surveys: Experiments on Social Presence in Web and IVR Surveys. Computers in Human Behavior, 19, 1–24.
U.S. Department of Commerce. (1994). Current Population Survey Interviewing Manual (CPS-250). Washington, DC: U.S. Bureau of the Census.

Received April 2004
Revised September 2005