Surveyor 3.0: An Open Source Application for Sentiment Analysis

Jonathan H. Morgan[1] and Geoffrey P. Morgan[2]

Duke University, Durham, NC, USA[1]

Carnegie Mellon University, Pittsburgh, PA, USA[2]

Abstract: We introduce Surveyor 3.0, an open-source tool for the collection of sentiment data. We detail the relevant theoretical background of sentiment collection and show how this new tool has learned from the lessons of the past. We provide details of the tool's features, and provide some detail on how those features are achieved. Surveyor 3.0 has been used actively at multiple institutions and in multiple countries.

Keywords

Sentiment Analysis, Cross-Cultural Studies, Affect Control Theory

INTRODUCTION

Text analytics holds great promise for further developing language-based theories of social interaction. This promise, however, depends on developing more tractable data driven tools for data collection. We introduce such a tool, Surveyor 3.0, and discuss it with reference to the challenges associated with collecting cultural sentiments, particularly in cross-cultural contexts.

To be clear, when we refer to language-based theories of social interaction, we mean theories that share a common premise, namely that people rely on durable, widely held, cognitive, and *affective* associations to structure social interactions and make inferences (Freese and Burk 1994; Hogg, Terry, and White 1995; Smith-Lovin & Heise 1988). In sociology, symbolic interactionism in the form of identity theories such as structural symbolic interactionism (Stryker 1980; 2004), identity control theory (Burke 1980), and affect control theory (Heise 1979) provide complementary, but different accounts of the processes associated with this premise. Identity theorists have examined such phenomena as social differentiation (Serpe and Stryker 1987), identity maintenance (Robinson and Smith-Lovin 1992), self-esteem (Cast and Burke 2002), and emotion management (Francis 1997), among others. These theorists have used diverse methods to study social interaction and its implications, including semi-structured interviews, surveys, simulated vignette studies, and experiments.

The emergence of "big data" - in the form of tweets, blogs, posts on social networking sites, or e-mail corpuses – necessitates the development of both new theories and testing strategies for identity

theories. In the present research, we focus on affect control theory because we believe it offers several advantages. First, affect control theory's operationalization of theoretical concepts is immediately suited to computational methods. The theory's primary inputs, *cultural sentiments*, are represented as three dimensional vectors, while its mechanisms are defined by a set of regression equations (Heise 2007). Second, affect control theorists' sustained interest in construct validity has led to the construction of sentiment libraries suitable for longitudinal and comparative analysis. These libraries consist of samples from North America, Western Europe, East and Southeast Asia, and the Middle East. Although researchers typically use these libraries for simulation experiments and validation studies, these repositories are also well suited for data mining techniques. Researchers have started to use these libraries to inform machine learning techniques (Alhothali and Hoey 2015; Joseph, Wei and Carley 2016). These techniques offer researchers new tools to understand social psychological processes such as re-identification, role negotiation, and the effect of local and organizational sub-cultures in naturalistic settings.

However, until recently, the relatively small size of sentiment libraries has posed a challenge to more extensive affect control theory-driven text analysis, particularly those libraries that use non-Roman alphabets. Several efforts have been made to remedy this problem. One approach is to use third-party platforms such as Qualtrics™ to facilitate online data collection so to reach wider and more diverse audiences, as well as providing yet another tool to collect cultural sentiments. We have found that Qualtrics™-based studies can provide comparable ratings to those collected in a lab setting. We have had great success collecting ratings in the United States with this method. For more information on how to implement a semantic differential scale using Qualtrics™, see: http://brentcurdy.net/qualtrics-tutorials/scales/. Another approach is monitored machine learning which uses student coders to match words to their closest equivalent in an existing library to create a set of seed terms from which to infer ratings of similar but unknown terms (Alhothali and Hoey 2015). Machine learning approaches are capable of dramatically expanding a cultural sentiment library using texts.

There are some limitations to these approaches. We discuss Qualtrics™-based studies first. The first limitation of a Qualtrics™ -based study is that the researcher must have reliable access to Qualtrics™. The second is that replicating the semantic differential scales designed by Heise (2001) for the collection of cultural sentiments exactly is so programmatically intensive as to be functionally prohibitive, specifically replicating the spacing of the scale segments and implementing the randomization of scale end-points. In addition, there are some psychometric measures that are difficult to collect using Qualtrics™ such as the presentation order of a given concept's dimensions. Finally, even if condition randomization, concept randomization, and dimension randomization are all implemented in Qualtrics™

as specified by Heise (2010), it is extremely difficult and/or impossible to create a survey that allows participants to change their ratings. Therefore, it may not be appropriate to compare ratings drawn from a Qualtrics™ study to those drawn from a non-Qualtrics™ study.

Machine learning approaches sacrifice per participant quality checks for speed. Supervised machine learning techniques depend on both having high-quality testers (a pool of cultural experts) and reliable seed terms. The concern over high-quality testers is one shared by all affect control theory sentiment survey designs. All designs assume that the researcher is sampling from a competent pool of cultural experts who share cultural meanings. The second concern, reliable seed terms, can be more problematic. Machine learning techniques rely on seed terms that are affectively comparable to other terms in their class. Assessing whether two terms are both semantically and affectively comparable is an open area of research (Curdy, forthcoming), especially in new cultural contexts. Consequently, machine learning approaches are better suited for expanding upon existing cultural libraries rather than developing a new cultural or sub-cultural library.

We here introduce another, complementary means of data collection inspired by David Heise's original *Surveyor* application (Heise 2001). *Surveyor 3.0* includes a data collection tool as well data manipulation utilities to make data collection faster and statistical analyses easier to perform and has successfully been used for multiple large-scale data collection efforts at Duke University and the University of Georgia, as well as in Cairo, Egypt; Kuwait City, Kuwait; and Rabat, Morocco. We begin this report with a brief review of the major challenges associated with collecting cultural sentiments. We then discuss the practical concerns of running a study using *Surveyor 3.0* in light of these methodological challenges.

Social Psychological Approaches to Culture

Social psychological approaches to the study of culture have been deeply influenced by the pioneering work of Osgood (1953), Osgood, Suci, and Tannenbaum (1957), and Osgood, May, and Miron (1975). Collectively, this work provided the measurement tools necessary to differentiate concepts from one another along an affective continuum (whether identities, behaviors, settings, or modifiers such as emotions and traits), generate stable quantitative profiles of concepts, and from these profiles generate predictions based on the affective meanings of the identities and behaviors defining a given situation (Smith-Lovin and Heise, 1988; Heise 2010, p. 27). Critical to the development of affect control theory's measurement model, Osgood, Suci, and Tannenbaum (1957) identified in their work, *The Measurement of Meaning*, three universal affective dimensions critical to impression formation: *evaluation*, *potency*, and *activity*. Each dimension consists of a core set of contrasts captured using a bi-polar scale called a *semantic differential*. For example, *evaluation* refers to perceptions of goodness versus badness or

cleanliness versus dirtiness, *potency* to perceptions of powerfulness versus weakness or bigness versus smallness, and *activity* to perceptions of liveliness versus quietness or youth versus agedness.

Reviews of *The Measurement of Meaning* offered three major critiques (Heise 2010). First, Osgood, Suci, and Tannenbaum (1957) claim that evaluation, potency, and activity measure semantic differences. Weinrich (1980, p. 26) challenged the semantic differential as a measure of semantic meanings, proposing that the differential in fact measures affective associations attached to meanings. By 1962, Osgood and colleagues had reconciled themselves to this position (Heise 2010 p. 26). Nevertheless, researchers have continued to refer to scales of this kind as semantic differentials and use them for a variety of applications, including the measurement of attitudes and emotions (Bradley and Lang 1994). The second major critique was that past researchers had identified similar three-dimensional systems for measuring affective meanings, specifically Wilhem Wundt's system. Osgood again conceded this point (Osgood 1962, p. 19-20). The third and more serious critique involved the universality of the dimensions. Critics argued that the factor analyses featured in the cross-cultural studies validating the universality of evaluation, potency, and activity (EPA) were flawed. Specifically, the factors identified in these studies might be an artifact of translating English concepts into non-English languages.

Establishing the universality of the EPA dimensions involved an extensive 13-year pan-cultural research program. This project examined 21 language-culture communities in order to derive 1,050 bi-polar scales indigenously. Comparing samples of first 100 and then 200 male high school students for each community, Osgood, May, and Miron (1975) found invariably that the first three factors in each sample corresponded to evaluation, potency, and activity respectively. By showing the universality of the evaluation, potency, and activity dimensions, this work also showed that semantic differentials can serve as a generalized method for measuring sentiments, leading to the creation of sentiment libraries for longitudinal and cross-cultural comparison (Heise 2010, p. 37).

BACKGROUND: THE ART AND SCIENCE OF COLLECTING AFFECTIVE MEANINGS

Since its invention, the semantic differential has undergone several changes. We discuss these with reference to the semantic differential implemented in Surveyor 3.0, shown in Figure 1. We then discuss the methodological and analytical challenges that have resulted from the move from paper and pencil formats to modern distributed studies.

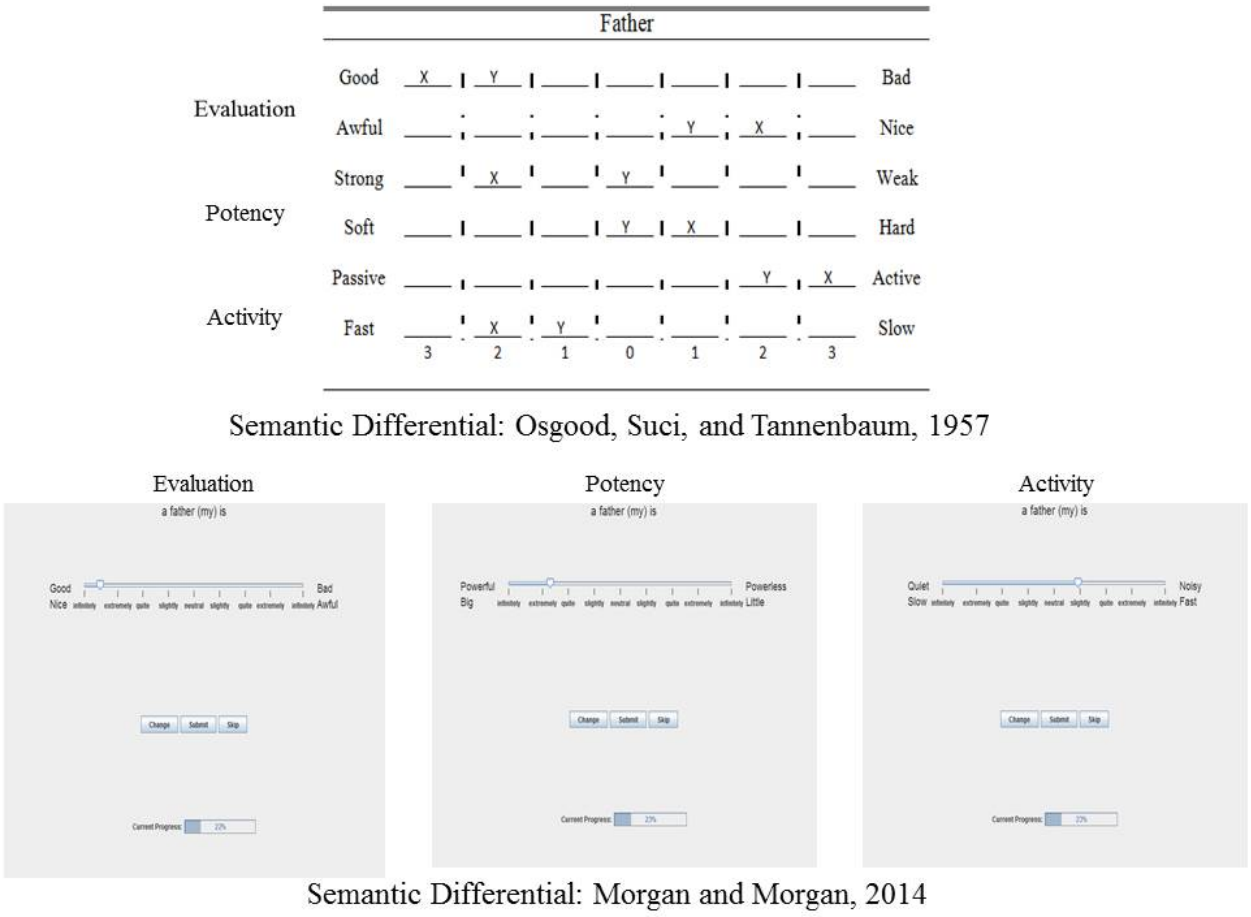

Figure 1. Semantic Differential, Past and Present

Figure 1 compares a reproduction of the semantic differential featured in *The Measurement of Meaning* to the modern format used by sentiment researchers found in Surveyor 3.0 (Osgood, Suci, and Tannenbaum 1957). We have added the dimension labels (evaluation, potency, and activity) for clarity. There are three major differences in the formats, each of which is a response to methodological challenges first identified by researchers in the 1960's (Kahneman 1963). The differences include the following: 1) the number of anchor concepts featured on each bi-polar scale, the scale's range, and its segmentation. There have also been procedural innovations regarding the presentation of the semantic differential. Within the affect control theory (ACT) community, these responses and procedures have been adopted as best practice. We will discuss these changes briefly, but acknowledge that there remains debate over scale formats and procedures, especially regarding the number of intervals, spacing of axis intervals, and the randomization of the anchor concepts (Cox 1980; Saris and Gallhofer 2014).

Multiple Anchor Points

Early versions of the semantic differential were bi-polar scales consisting of a single set of contrasting concepts (e.g., a continuum ranging from sweet to sour). The change to bi-polar scales characterized by more than one anchor concept was a response to concept-scale interactions. Concept-scale interactions frequently occur when a concept has a denotative meaning that is associated with a given anchor concept. Concept-scale interactions can either provide an exaggerated rating or a false average. Sentiment researchers use multiple scales or anchoring concepts when ascertaining a concept's meaning because different metaphors associated with each dimension are more or less appropriate. For example, participants can become confused when asked to rate the goodness or badness of a concept, like an apple, if this dichotomy is not supplemented by another dichotomy such as sweet or sour, which itself may be culturally dependent. Nevertheless, there are risks associated with this strategy. Each dichotomy introduces its own denotative content which can result in a concept-scale interaction.

Heise (2010) recalls Osgood, May, and Miron's lemon example to illustrate such an interaction (1975, p. 33). Imagine participants rating a lemon for the purpose of determining the widely held affective meaning associated with lemons. Participants rating a lemon in early studies would encounter multiple bi-polar scales meant to capture the meanings associated with each dimension (evaluation, potency, and activity). For example, participants would encounter dichotomies such as good/bad, awful/nice, and sweet/sour when rating a concept's evaluation, as shown in Figure 1. Presenting the scales separately led some participants to interpret each dichotomy separately and not as part of an overall metaphorical grouping pulling on a common latent concept. Consequently, some participants would rate lemons as very sour, quite good, and neutral on the awful/nice scale. If considered as separate denotative objects, the logic of these ratings is quite clear. Lemons are very sour but quite good in a martini or lemon meringue pie. As for a lemon's awfulness or niceness, does one have conversations with a lemon? The resulting rating, however, is a false average resulting from an extreme sour rating offset by a rating of quite good, and this average provides no insight into the evaluative dimension of a lemon's affective or connotative meaning because it is an artifact of the denotative association between sour and lemons.

Provided we accept the validity of the evaluation, potency, and activity dimensions or any other latent concept we are trying to measure, using multiple concepts to anchor a bi-polar scale affords two significant advantages. First, it provides the participant multiple metaphors with which to interpret the concept's meaning, while simultaneously highlighting for the participant the underlying construct they are to apply when rating the concept (evaluation, potency, or activity in this case). Second, it is efficient. The alternative to using multiple anchor concepts is to average many more scales to compensate for the potential distortion introduced by any set of anchoring concepts. Such an approach would drastically

reduce the number of stimuli a participant could reasonably rate in an experimental session. In addition, there would be even more point estimates that could be confounded by outliers.

Range and Segmentation

The range of the semantic differential used in sentiment analysis has also changed. The classic format is a seven-point scale, with 0 indicating neutral and three equally spaced points indicating more extreme positions. This format is still used by marketing researchers and some psychologists (Saris and Gallhofer 2014). For sentiment analysis, however, Heise (2010b) notes that the seven-point scale frequently introduces ceiling and floor effects. For example, participants using a seven-point scale would rate *friends* and *God* as both "extremely" good (Heise 1978, p. 64). Heise's remedy is to change the scale to a continuum segmented into nine intervals, with infinitely marking the most extreme positions. Sentiment researchers have also largely abandoned numeric labels because the difference between positions, for example between 1 and 2, is not always intuitive. On the other hand, participants encountering adjectival labels generally have some sense of the difference implied by a change in position because adjectives are routinely used to describe magnitude. For example, most participants have some sense of the difference between something being *slightly* important versus *quite* important.

The introduction of nine intervals led to other changes to the semantic differential scale. The classic format also features equally spaced intervals as shown in Figure 1. Heise (1978) finds that the equal interval assumption is problematic for sentiment analysis. Using a least squares solution for successive intervals scaling developed by Diederich, Messick, and Tucker (1957), Smith-Lovin (1987, pp. 42-44) provides further support for this finding, showing that the size of the intervals are unequal and vary with respect to the type of stimulus (nouns, verbs, adjectival modifiers, and settings), the gender of the participant, and the dimension being rated.[1] As a result of these analyses, the affect control theory community has adopted a scale that ranges from -4.3 to +4.3, with the first three positions being equally spaced and the final position, *infinitely*, being 1.3 times the distance of the others.

To summarize, the format of the semantic differential used by sentiment researchers has changed significantly since its inception. There remains debate regarding the number and specific anchoring concepts that are most appropriate, the exact range of the scale, and its segmentation. We believe that

[1] Diederich, Messick, and Tucker (1957) assume the ratings of stimuli are normally distributed. They set the upper category boundary of g^{th} category as a function of scale value of stimulus i plus the product of the conditional probability of the category and the dispersion of the stimulus. Using weights to handle missing cases (where the weights are scaled as a function of the mean scale value of the category boundaries and the standard deviation), the algorithm iteratively approximates the boundary values until the normality assumption is maximally fulfilled for all the stimuli.

research tools that incorporate the semantic differential as a stimulus type need to be flexible on each of these aspects, as these choices are largely informed by the research question and phenomena under study. Nevertheless, the semantic differential is a mature measurement instrument whose essential features have been informed by over fifty years of research. Effectively deploying the semantic differential in both local and cross-cultural studies, however, remains a challenge. In the following section, we discuss how Surveyor 3.0 collects affective meanings and where it fits in this methodological tradition.

COLLECTING AND ANALYZING AFFECTIVE MEANINGS IN SURVEYOR

Surveyor 3.0's design envisions the experimental process as consisting of three distinct tasks: designing and piloting a study, running a study, and analyzing data. The requisite level of knowledge required to perform these tasks varies. Our goal in developing and implementing Surveyor is to reduce the level of application-specific knowledge necessary to perform these tasks while preserving data quality. We will now discuss Surveyor 3.0's tools and features with regards to each of these tasks.

Running a Study in Surveyor 3.0

When we refer to Surveyor 3.0, we generally mean a suite of tools and files contained in a common folder (shown in Figure 2) rather than a single application. We will discuss these files and tools with respect to three tasks: running a study, designing a study, and analyzing data. We begin with how to run a study in Survey 3.0 so that the reader has a better understanding of how the tool works. By running a study, we refer to both setting-up Surveyor for an experimental session and backing-up the data created during that session. We will discuss each task in turn.

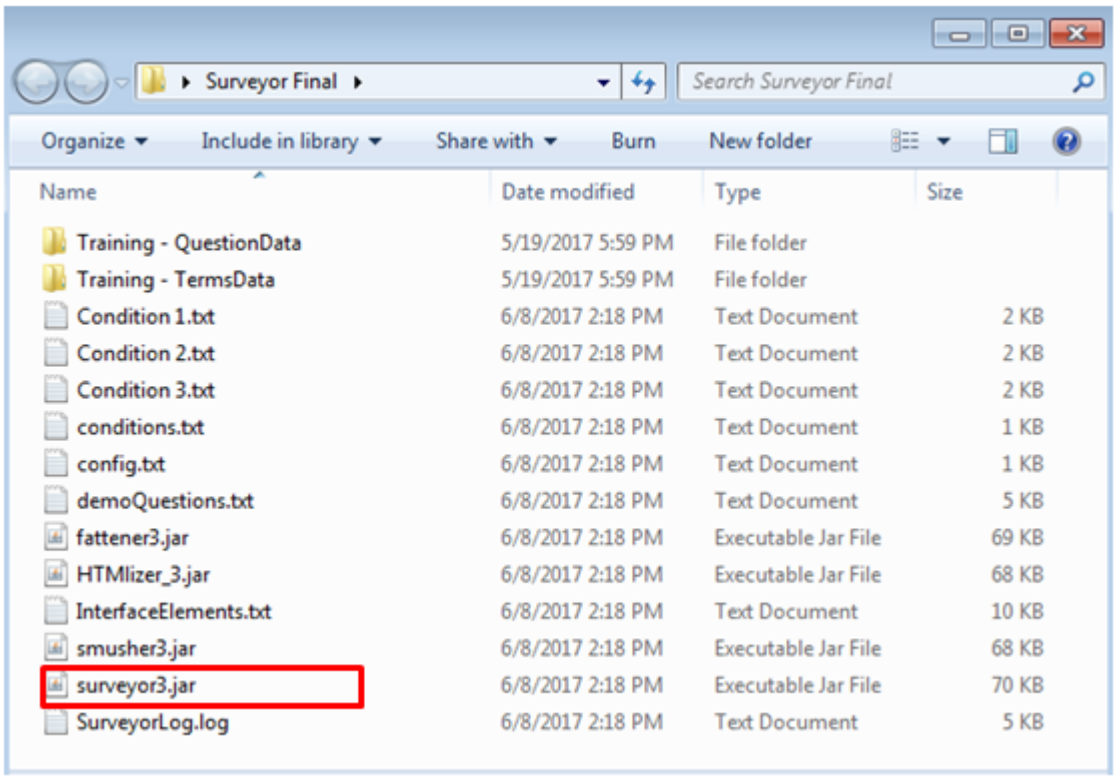

Figure 2. Surveyor 3.0 Application Folder

Setting-Up

To set Surveyor 3.0 up for an experimental session, first double-click the file *surveyor3.jar*, indicated by the border in Figure 2. After double clicking *suveryor3.jar*, Surveyor 3.0's loading screen will appear followed by the participant ID dialog box, shown in Figure 3. Surveyor always loads the participant ID dialog box. This presentation order is a safety feature to ensure that there is always an ID for any data collected by the application. Surveyor prohibits entering the same participant ID within a given 24-hour period; this is to prevent inadvertently overwriting data before the data has been backed up.

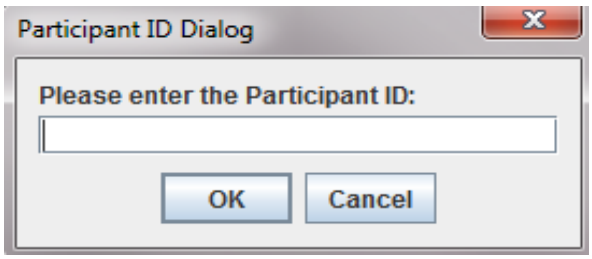

Figure 3. Participant ID Dialog Box

The second step of setting up a study is to select the study condition. The condition screen, shown in Figure 4, appears immediately after the Participant ID dialog box. At minimum, there must be one condition because conditions in Surveyor specify the set of concepts the participant will rate.

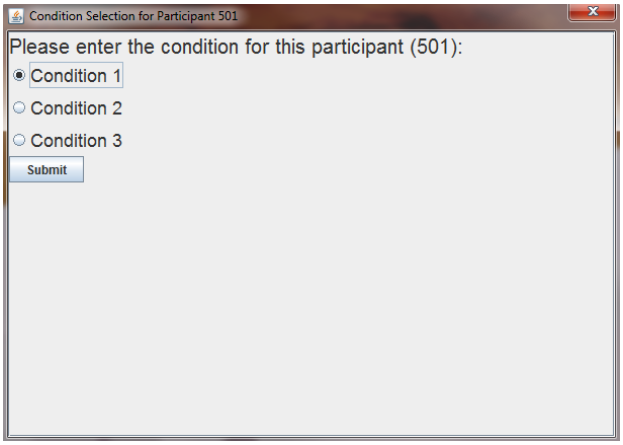

Figure 4. Condition Selection Dialog Box

Session Execution

After the researcher completes the participant ID dialog box and condition selection dialog box, the survey will begin. Surveyor 3.0 follows a sequence: 1) it presents the survey questions that the researcher has specified to come before the concept rating phase, 2) it presents a tutorial that instructs the respondent on how to rate concepts, 3) it presents the set of concepts to be rated, and 4) it presents survey question specified by the researcher to occur after the concept rating phase. By survey questions, we mean any questions other than concept ratings. Survey questions typically include demographic questions and questions assessing cultural familiarity but are ***not*** limited to these kinds of questions. Currently, survey questions can only be specified using a multiple choice format. If no survey questions have been specified to occur before the concept rating phase, Surveyor will immediately begin with the study tutorial.

Figure 5 shows a survey question that appears before the tutorial and concept rating phase. Surveyor 3.0 allows the researcher to specify whether a survey question comes before or after the concept rating phase because question order is important in some studies. For example, studies focusing on gender or racial identities may require questions about race or gender to appear at the end of the survey so as to avoid influencing the participant's ratings. On the other hand, there can also be instances where collecting information before the concept rating phase may be useful such when the researcher wishes to identify patterns of missing data associated with survey locations. In the example, we ask an administrative question – school enrollment – primarily for record keeping purposes. Other researchers, however, may find a question about school enrollment useful for distinguishing between potential subcultures; see Smith-Lovin and Douglass's (1992) study examining two religious subcultures as an example.

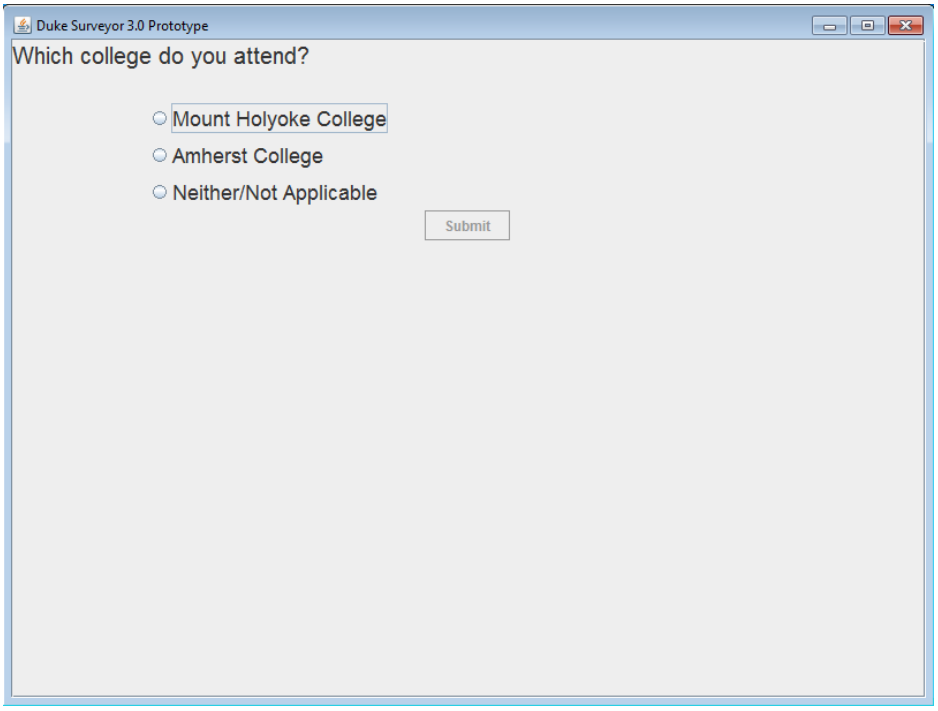

Figure 5. Example Survey Question

After responding to the last survey question specified to appear before the concept rating phase, Surveyor 3.0 alerts the participant that there will be a tutorial. A brief tutorial is necessary because semantic differential questions are often unfamiliar to participants. The tutorial also provides experimenters a final opportunity to respond to participant questions before participants start rating concepts.

The tutorial walks through a pre-programed set of examples. These examples come from Heise's (2001) Surveyor programs, helping to ensure consistency between data collected with previous versions of Surveyor and Surveyor 3.0. The experimentalist can, however, change the text of the tutorial to render it in different languages or in non-Roman alphabets.

Surveyor 3.0 first informs the participant that they will rate concepts based on how they feel about them and that concepts will appear at the top of the screen. Next, the tutorial introduces an example, *helping someone*, and reinforces that the experimental task is to rate how she feels about concepts such as helping someone. In the following screen, the tutorial introduces a slider and instructs the participant to move the cursor along the slider. The tutorial next adds the slider positions of neutral, slightly, quite, extremely, and infinitely good, with neutral being at the center. The tutorial also informs the participant that she will be rating concepts using three different rating scales. The next three screens introduce the end points and scales: evaluation, potency, and activity respectively. On the activity screen, the participant must rate the concept and select submit for the first time to proceed with the tutorial.

After familiarizing the participant with semantic differential scales generally, the tutorial walks the participant through how to rate a concept as neutral, how to change a rating, and how to skip a concept. The participant encounters a new example, *a stranger*. At this point, Surveyor 3.0 has the participant practice rating a concept as neutral by moving the slider away from the middle of the scale and back again. This part of the tutorial is necessary because Surveyor requires participants to move the slider before being able to submit their rating of the concept. This feature of the program is to encourage participants to use the scale. By practicing how to rate concepts as neutral, the tutorial both demonstrates that it is possible and helps encourage the participant to view rating a concept as neutral as a purposeful act rather than a default assumption. Forcing the participant to move the slider to proceed through the survey might result in rater fatigue, possibly exhibited by ratings that are always to the right (or left) of the screen. Surveyor 3.0's AxisDirection data quality measures allow the researcher to detect such ratings.

The tutorial next has the participants practice how to change a rating of a concept the participant has already rated. To change a rating, the participant selects the *change* button. A dialog box appears listing the concepts the participant has previously rated. The participant selects the concept, selects the submit button, and is then re-directed to that concept. Surveyor 3.0 will show the participant her previously selected ratings of the concept's evaluation, potency, and activity. If the participant is satisfied with her rating, the participant simply selects submit. To change a rating, the participant moves the slider to a new position and then clicks submit for the new rating. Surveyor 3.0 logs both the respondent's original and changed ratings.

The part of the tutorial describing how to change a rating can be confusing for participants. The cause of the confusion is that the change dialog box has no concepts to rate. So, there is nothing to do. This is because Surveyor 3.0 currently does not log ratings of concepts shown in the tutorial. Consequently, Surveyor 3.0 does not detect that the participant has rated any concepts, and so does not show any concepts to change. We are currently addressing this source of confusion by having Surveyor 3.0 record participant ratings generated during the tutorial to a temporary file. This change will appear in a future release. In the meantime, be aware that this part of the tutorial can cause confusion.

Finally, the tutorial has the participant practice how to skip a concept, shown in Figure 6. To skip a concept, the participant selects the skip button. Surveyor 3.0 presents a skip submenu that requests a reason for why the concept was skipped. Options include: 1) I don't know what the concept means; 2) the concept makes me feel uncomfortable, or 3) Some other reason. In the event the respondent skips one of the concept dimensions, Surveyor 3.0 will not present any further dimensions for that concept. For example, if the respondent skips rating the activity of a polymath, she will not be asked to rate the evaluation or potency of a polymath. The rational for asking the participant to indicate a cause for

skipping the concept is to both isolate potentially problematic concepts, and to discourage heedless skipping. Researchers can change the number, language/alphabet, and wording of the skip submenu prompts. Having chosen a reason to skip, the participant encounters the tutorial's final screen which notifies her that she is about to complete the tutorial and begin the next phase of the study.

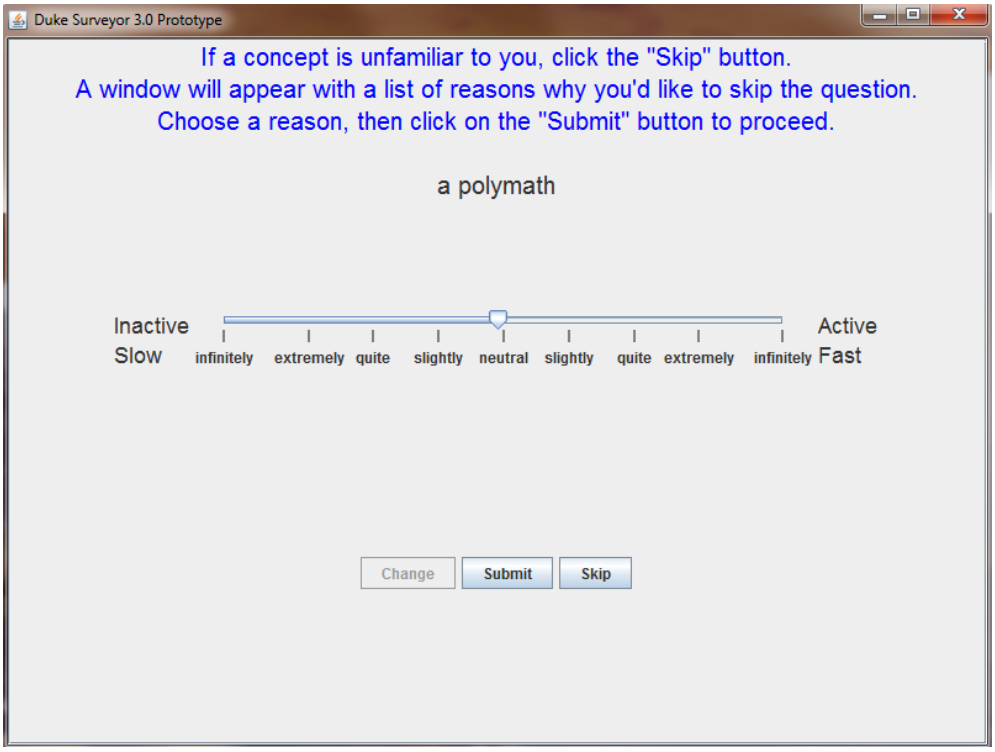

Figure 6. Surveyor Tutorial, Practicing How to Skip

The majority of the respondent's time is spent rating concepts. For each concept, the respondents will see a separate screen for each dimension of that concept. Both the presentation order of the concepts and the order of the concept dimensions (i.e., evaluation, potency, and activity) are randomized to prevent the early onset of rating fatigue; however, the experimenter can change these settings. Each rating instance consists of a semantic differential, demarcated by the labels neutral, slightly, quite, extremely, and infinitely. Surveyor 3.0 records responses as its respective value between -4.3 to 4.3. Surveyor 3.0 also displays a progress bar to the respondent. This is especially important for longer studies as it helps respondents cope with rater fatigue. As noted earlier, participants have three options: rating the concept, changing a previously rated concept, or skipping the concept as shown in Figure 7.

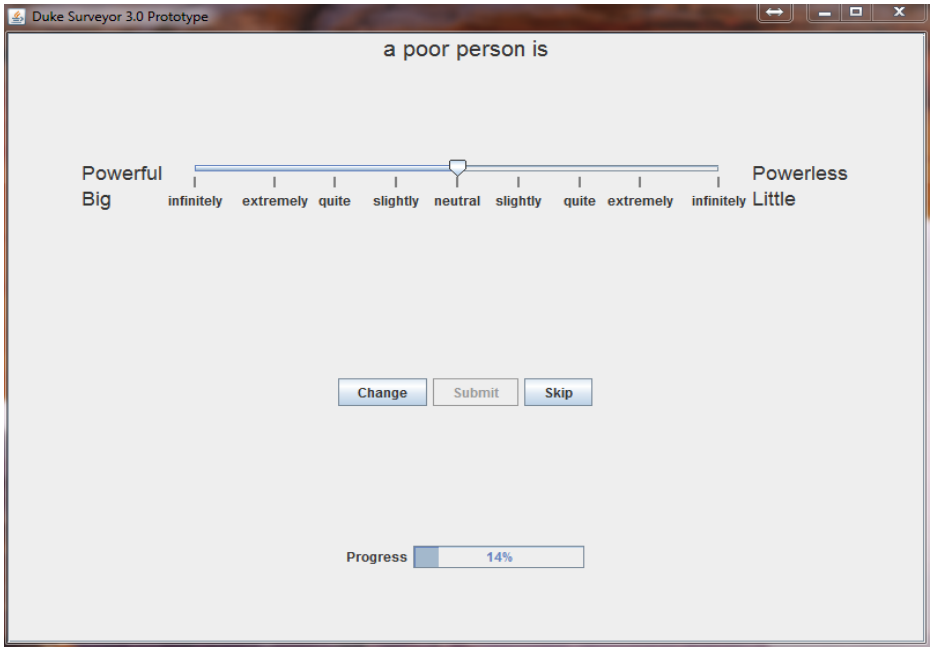

Figure 7. The Semantic Differential for the Identity, Poor Person, along the Power Dimension

The concept rating phase ends with an exit screen that indicates that the respondent has completed the concept rating phase, shown in Figure 8. Surveyor 3.0 also provides the respondent one last chance to change any of her responses before concluding. This screen also concludes the session unless the researcher has programmed questions to follow the concept rating phase. If questions follow, Surveyor 3.0 will terminate after the last question. Although response data is stored as it is recorded so that sudden termination does not result in loss of data, we advise including text in the last question that indicates to the respondent that she has completed the survey so that the respondent isn't startled by the sudden termination.

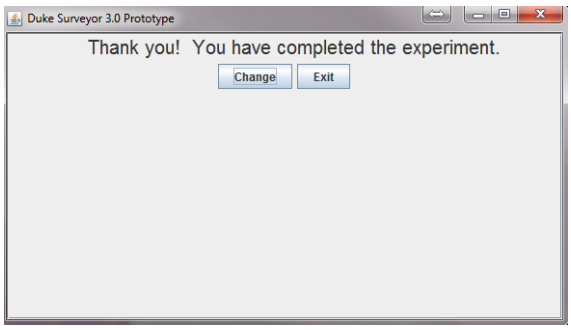

Figure 8. Concluding Screen of the Survey's Concept Rating Phase

Backing-Up Data

The final phase of running a study in Surveyor 3.0 involves backing-up the data. Surveyor writes two types of output files, a questions file and terms file to two folders (a questions folder and terms folder), indicated by the border in Figure 9. The file names for both file types include the participant ID, a

time stamp, and a file type indicator (e.g., _questionEntries for question files or _termEntries for terms files).

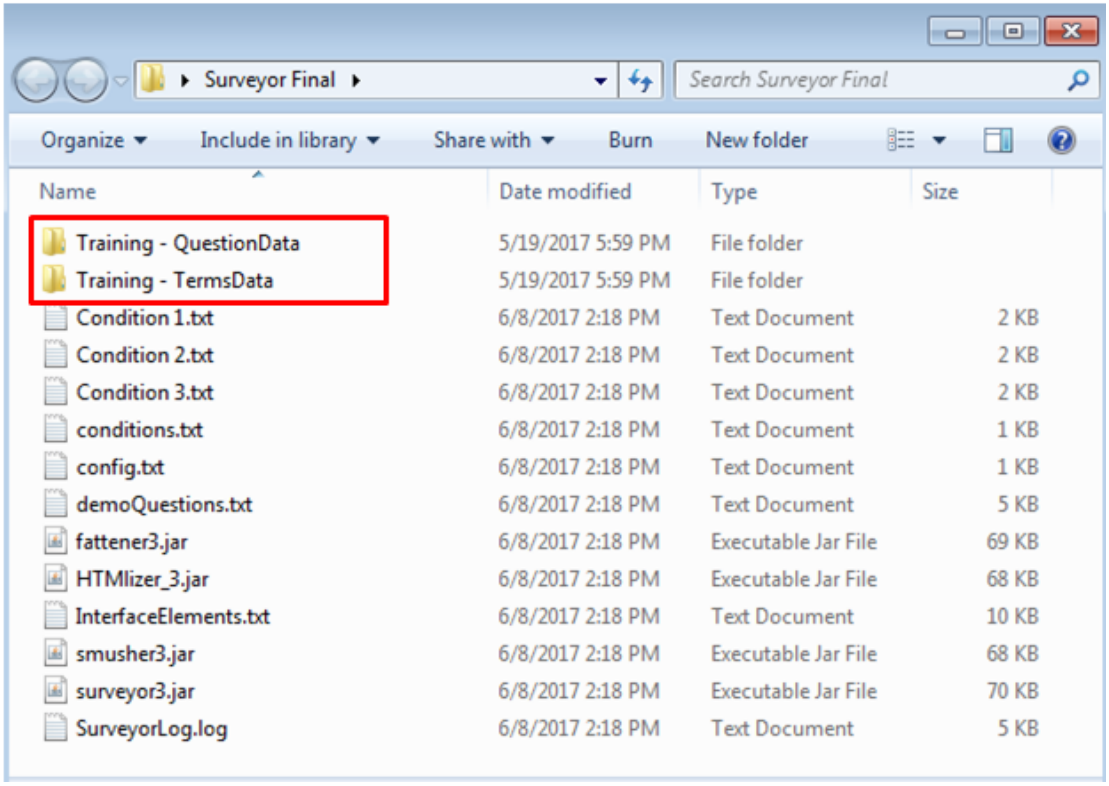

Figure 9. Question and Term Folders

Each respondent's file inside the *QuestionData* folder contains her responses to the survey questions such as demographic, cultural familiarity and administrative questions; the respondent's terms file inside the *TermsData* folder contains her concept ratings. These files can be linked through their participant ID number (which researcher entered in the Participant ID dialog box) that appears in the file name. The questions and terms folders contain all files of that type generated by Surveyor 3.0 unless the investigator has deleted files, or has changed the files' output directory by changing the experiment name in the configuration file.[23] The QuestionData and TermsData folders are found in Surveyor's main application folder. At the end of a survey session, we recommend copying the files inside the QuestionData and TermsData files to an external hard drive or other storage device. We recommend not

[2]Because Surveyor requires very few data permissions (the ability to read and write files in its local directories), it is relatively easy to setup Surveyor on administered systems as well as on personal devices or networks.
[3] In the event the output file designation has been changed, the investigator will see two sets of questions and terms folders in Surveyor's main application folder, two folders corresponding to the old experiment name and two to the new name.

changing the name of the QuestionData and TermsData folders on your backup drive so as to facilitate aggregating the files later using Surveyor 3.0's *smusher* utility.[4]

Designing and Piloting a Study

Study design in Surveyor 3.0 consists of four steps: 1) identifying and configuring meta-features, 2) specifying the demographic questions and responses presented to the participant, 3) specifying the terms presented, and 4) specifying the interface elements where necessary (e.g., when designing cross-cultural studies). All these steps can occur in Excel and entail working in pre-formatted text files. Surveyor's application folder contains examples of all these files. Surveyor comes with an example study which provides templates for the files associated with each of these steps. The advantages of this approach include: 1) *Surveyor 3.0* supports studies using both Roman and non-Roman alphabets; 2) investigators can now easily customize the interface to meet unique study needs; and 3) rapid implementation and testing of the survey.

Configuring Surveyor 3.0 requires thinking about the survey's study design. An important preliminary question is whether the survey will consist of multiple conditions, and if so, how the conditions will differ from each other. Typically, when conducting sentiment analysis of a single culture, the survey instrument consists of multiple conditions, the sole difference between them being the list of concepts to be rated. Other survey features, such as the user interface, demographic questions, and randomization settings, are held constant. Cross-cultural surveys, however, may consist of conditions that differ with respect to the user interface, depending on the experimental setting (e.g., the language used for the survey). There may also be reasons to test different randomization settings during pilot studies. Consequently, for each condition, the investigator specifies the following: the files that will be used to generate the interface; the demographic questions; the concept terms that will be rated; and the randomization settings in what is known as the conditions file, shown in Figure 10.

[4] We suggest matching the names of the folder because in the event of a data loss the investigator will be able to generate new aggregate logs from the backup files simply by copying the two folders, placing them into Surveyor's main application folder, and running Surveyor data aggregation utility.

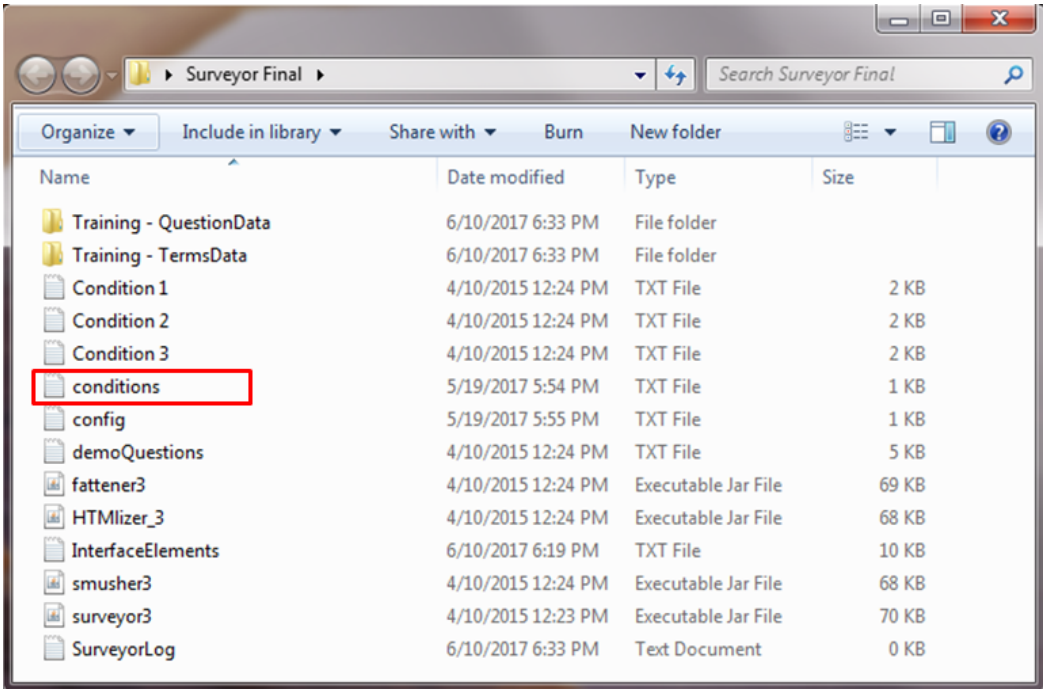

Figure 10. Suveryor 3.0's Conditions File

Although configuring Surveyor 3.0 may sound tedious, it is a relatively straightforward process consisting of entering row values (one row per condition) for eight columns in Excel or some other text editor. ***Note***: for all pre-formatted text files, the column headers must remain the same while row values can change. Table 1 is an example condition file from Surveyor's training module. Columns 1 and 2 indicate the condition and experiment name, respectively. Columns 3, 4, and 5 specify the files Surveyor will call to generate the condition. Columns 6, 7, and 8 specify the randomization settings. [5]

Table 1. The Conditions File from Surveyor's Training Module

Condition	Experiment	Interface	Questions	Terms	Randomize Term Order	Randomize Axes Order	Randomize Polarity
Condition 1	Training	InterfaceElements.txt	demoQuestions.txt	Condition 1.txt	True	True	True
Condition 2	Training	InterfaceElements.txt	demoQuestions.txt	Condition 2.txt	True	True	True

Having determined the number of conditions and whether they will differ from each other, the next step in configuring Surveyor 3.0 is to create the question and term lists that the application will use to generate the survey questions and concepts that the respondent will encounter. We discuss generating survey questions first.

[5] The settings, Randomize Axes Order and Randomize Polarity, refer to the presentation of the semantic differential. Randomize Axes Order refers to the presentation of the evaluation, potency, and activity dimensions for each concept, while Randomize Polarity refers to the left-right orientation of the scale with respect to positive and negative values.

Survey Questions

Creating and specifying survey questions in Surveyor 3.0 is a straightforward process involving entering row values, with rows in this case corresponding to questions in demoQuestion.txt, shown in Figure 11. Researchers can change the name of the file from *demoQuestion.txt* if she changes the values of the Questions column in the Conditions file. In addition, you can also have multiple survey question files associated with different conditions if these files are specified in the conditions file.

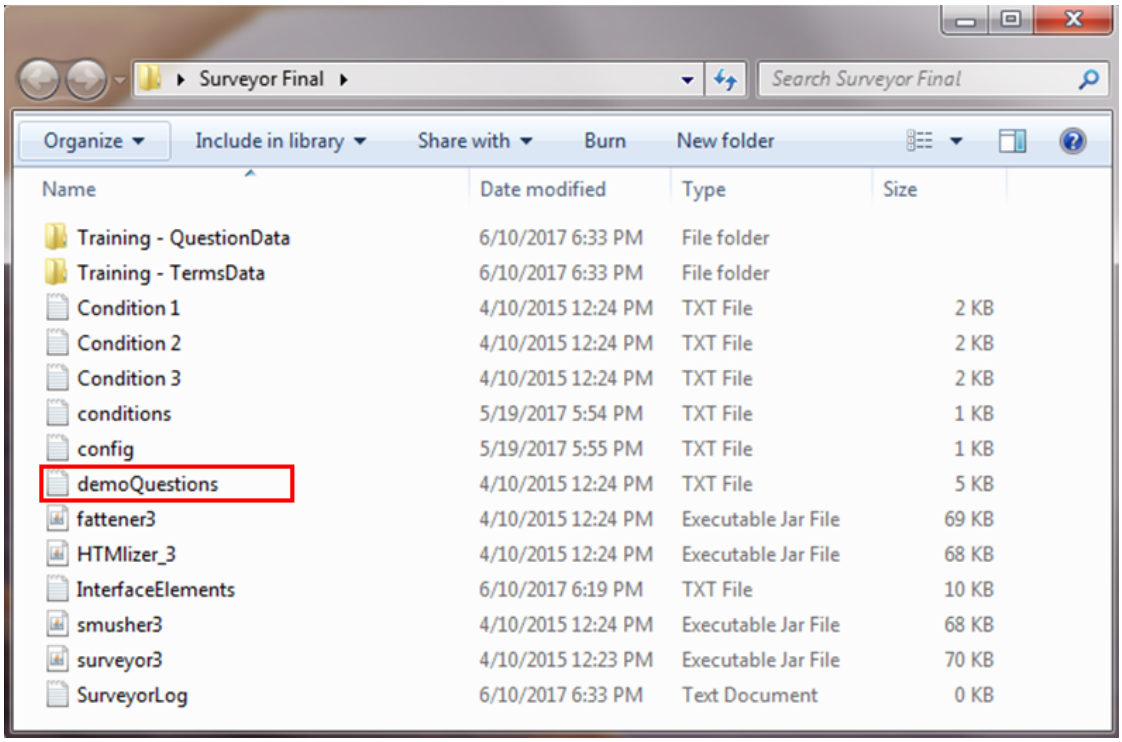

Figure 11. Survey Questions File

Researchers are able to easily name, customize, and sequence the demographic questions. If left unspecified, Surveyor 3.0 is programmed to provide a default set of questions and values, taken from the original Surveyor. We discuss these features in more detail, below (see Table 2).

Table 2. Survey Questions File from Surveyor's Training Module (demoQuestions.txt)

QuestionID	SubID	Question Text	Right Answer	Ask After Terms				
Gender		What is your gender?		1	Female	Male		
Age		What is your age?		1	18	19	20	21
Math	0	2 + 2	4	0	1	4		
Math	1	2^3	8	0	5	8		

Table 2 lists example survey questions used in past studies. The first column, QuestionID, lists the question's name. This can be a character string or a numeric value. The second column, SubID, forces questions to appear in a specific order, even if questions are set to be randomized by the

"randomQuestionOrder" option (see Appendix A). For example, the question forced to be shown first will be Math0 and the question forced to be shown second will be Math1. The combination of a common QuestionID (in this case Math) and values in SubID column, thus, allows the investigator to specify question blocks (groups of questions that appear together in a particular order). In the outputted data files, block questions are labeled with their QuestionID followed by their SubID number, Math0 and Math1 in this case. The third column, Question Text, specifies the question prompt, exactly as it will be displayed to the participant. The fourth column, Right Answer, transforms a simple question into an "attention check" by specifying a correct answer that the participant must enter to continue the survey. If Right Answer is left blank, then the participant can enter any response to continue the survey. In this case, we have specified that the answers 4 and 8, for Math0 and Math1, respectively, are the only answers that will allow the participant to continue the survey. Because Surveyor 3.0 only supports using a multiple choice format for survey questions, the participant will have to enter a value, precluding the possibility of blank or no responses. The fifth column, Ask After Terms, specifies whether Surveyor presents the demographic question before or after the survey's concept rating segment. This feature is useful in instances where there is concern about ordering effects (e.g., influencing the ratings of racial categories by first asking respondents to declare their own racial identity). In our example, the two attention checks appear first, but the questions asking the participant to indicate their gender and age appear at the end of the survey. Surveyor's default setting is to have questions appear before the concept rating phase, with any non-zero value entered in the column indicating that the question will appear at the end of the survey. The order of the survey questions appearing before and after the concept rating phase will correspond to order they appear in the survey questions file, unless the experimenter has chosen to randomize the question. Finally, the sixth column and every column to right of the sixth column specify the response options Surveyor 3.0 will present to the respondent.

Concept Questions

Survey conditions typically differ solely with respect to the concepts they present the participant. For a study with conditions that differ in this way (*i.e.,* each condition presents a different list of concepts to a respondent), there will be at least two "Condition" spreadsheets in the Surveyor 3.0 application folder, shown in Figure 12.

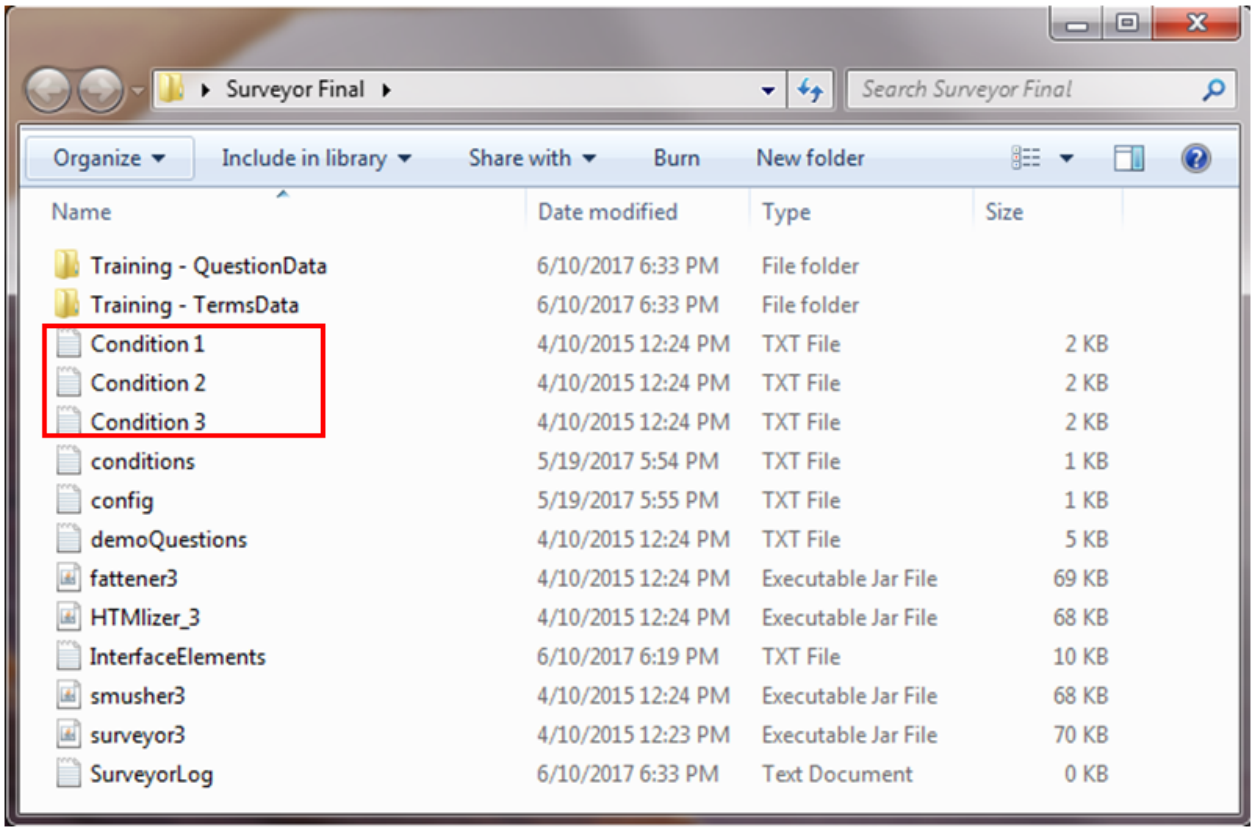

Figure 12. Condition Files (Files Specifying the Concepts that Will Be Rated in each Condition)

In each spreadsheet, the concepts are listed as rows, with the columns specifying the concepts' name for analytical purposes (Term_ID), and the actual wording of the concept as presented to the respondent (Phrase). Table 3 shows an example from the first condition of Surveyor 3.0's training module. The first column, Term_ID, is an identifier that allows Surveyor to differentiate between concepts. Note that in conditions that are differentiated by language, the Term_ID can be the *same* in two different conditions, while the Phrase can be different in two different conditions. This is because the respondent only ever sees the text in the Phrase column. For example, a study with an *English* Condition 1 might have the Term_ID "i_father" and Phrase "a father is", while its *German* Condition 2 might have Term_ID "i_father", but Phrase "Ein Vater ist".

Table 3. The First Condition's Concept File from Surveyor's Training Module (TrainingTerms_Condition1.txt)

Term_ID	Phrase
i_auctioneer	an auctioneer is
i_Austrian	an Austrian is

Configuring Surveyor 3.0

The last step in configuring Surveyor 3.0 is to specify the interface elements the participant will encounter in the InterfaceElements file, shown in Figure 13. For studies occurring in English speaking contexts, Surveyor's default interface file will most likely be sufficient.

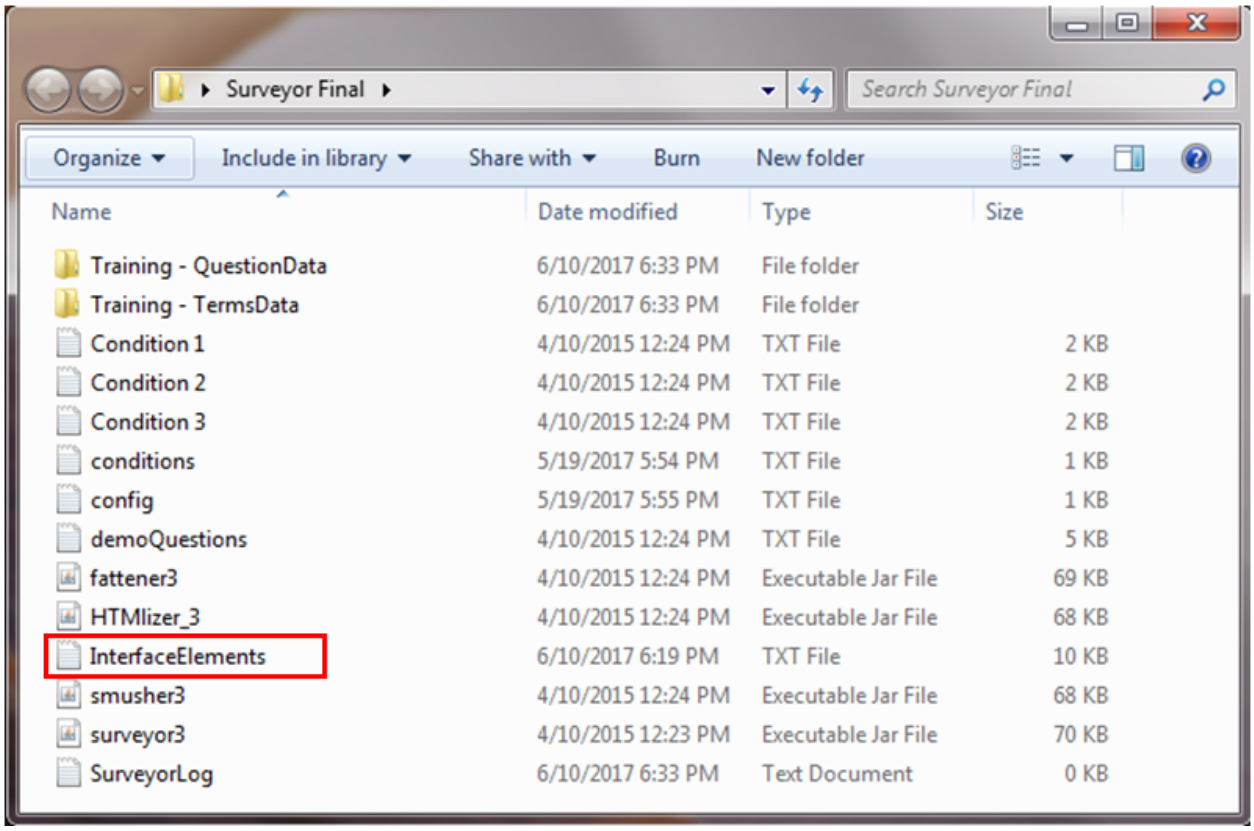

Figure 13. Surveyor 3.0's Interface Elments File

However, investigators can conduct studies using non-Roman alphabets by configuring the interface file. Surveyor's interface file consists of two columns: an elements column and a value column. The elements column lists the interface components the researcher wishes to specify; the value column lists the text that will be presented to the participant. All unspecified elements appear in the log along with their default values; we include an example log in Appendix A. Checking the log allows the researcher to identify any configurable features or elements for which Surveyor is using defaults. The experimenter can change an interface element from its default setting by including it in the interface file and specifying a new value. Configuring this file follows a similar process as configuring the survey question files and

concept files.[6] The investigator changes row values in the value column to change the text presented to the participant. For example, to change the text of the submit button from English to Arabic, we would change the entry in value column of the *SubmitButtonText* element from "submit" to "حفظ." We include the interface file used for studies conducted in the Middle East as an example of an interface file configured for a non-English speaking / non-Roman alphabet study context in Appendix B.

Piloting Surveyor 3.0

Having configured the study, we suggest a two-part piloting process: first, reviewing the concepts that will be rated during the concept rating phase; and second, piloting the entire survey using testers. To make reviewing the concepts easier, *Surveyor 3.0* offers a testing mode called Experimenter Mode.[7] When in Experimenter Mode, Surveyor 3.0 skips both the survey questions and the tutorial. In this mode, the investigator can move quickly from one concept to the next without rating them. These changes to the usual survey format are useful for both quickly checking the presentation of each stimulus item. To check the survey questions, it can be helpful to first set the questions to appear at the beginning of the survey before going through them. This step allows the investigator to check the survey questions without going through the entire survey.

The investigator can exit the survey using the escape password. Exiting Surveyor 3.0 is password protected unless specified otherwise in order to avoid inadvertent terminations of the application; we have found this feature to be helpful in lab settings.

Finally, if a condition fails to load, checking Surveyor 3.0's log file can help pin down the source of the error, whether there was a problem with the questions file, the terms file, or the interface file. The log file is generated upon termination of the application and notes what program features Surveyor 3.0 successfully and unsuccessfully loaded, see Appendix B for an example.

The second phase of piloting involves running a pilot sample of participants through the survey. Surveyor provides several quality statistics to assess the data and check for systemic errors. We discuss these quality measures and quality checks in the analysis section. Finally, we suggest keeping the log files created during the pilot sessions to ensure repeatability; Surveyor 3.0's log list all the experimental settings and can be used to re-implement the study if necessary.

---

[6] We suggest using a text editor such as Notepad++ instead of Excel when configuring the interface file because Excel encapsulates text elements it cannot identify with quotation marks. Alternatively, you can use Surveyor 3.0's HTMLizer to eliminate quotation marks from Excel edited files.

[7] Setting Surveyor to be in Experimenter Mode requires changing the element, Experimenter_Mode, from the default value of FALSE to TRUE in the configuration file.

Analyzing Study Data

There is an extensive literature spanning a variety of theoretical perspectives discussing the analysis of sentiment data. We provide an abbreviated discussion that focuses on data construction and some of Surveyor 3.0's data quality measures. For a more comprehensive discussion of sentiment analysis, we recommend Heise's (2010) book, *Surveying Cultures: Discovering Shared Conceptions and Sentiments*.

Data Construction

The first data analysis step is data construction. Surveyor 3.0 provides two utilities related to data construction: the *smusher* and the *fattener*. The smusher is a data aggregation utility that generates a long-form data set (a *smushed set*), where rows correspond to concepts while columns correspond to EPA ratings, survey question responses, and quality statistics (27 columns in addition to question responses).[8] Because the rows correspond to concepts, this file format is often referred to as a concept-level data set. The fattener transforms a *smushed* data set from a long to a wide format (a *fat set*), where rows correspond to individuals and columns correspond to ratings, survey question responses, and quality statistics for each concept. Because rows correspond to participants in this file format, this format is often referred to as a participant-level data set. Participant-level data sets are useful for exploring within-group variation (e.g., comparing scale usage patterns for different age cohorts). Finally, Surveyor's fattener utility allows the investigator to generate data subsets by selecting specific variables of interest. We next discuss how to use these utilities.

To aggregate ratings files (term files) and question files, first double-click the smusher utility in the Surveyor application folder, shown in Figure 14.

[8] In this format, the respondent ID is a column value that is repeated for every concept the respondent rated.

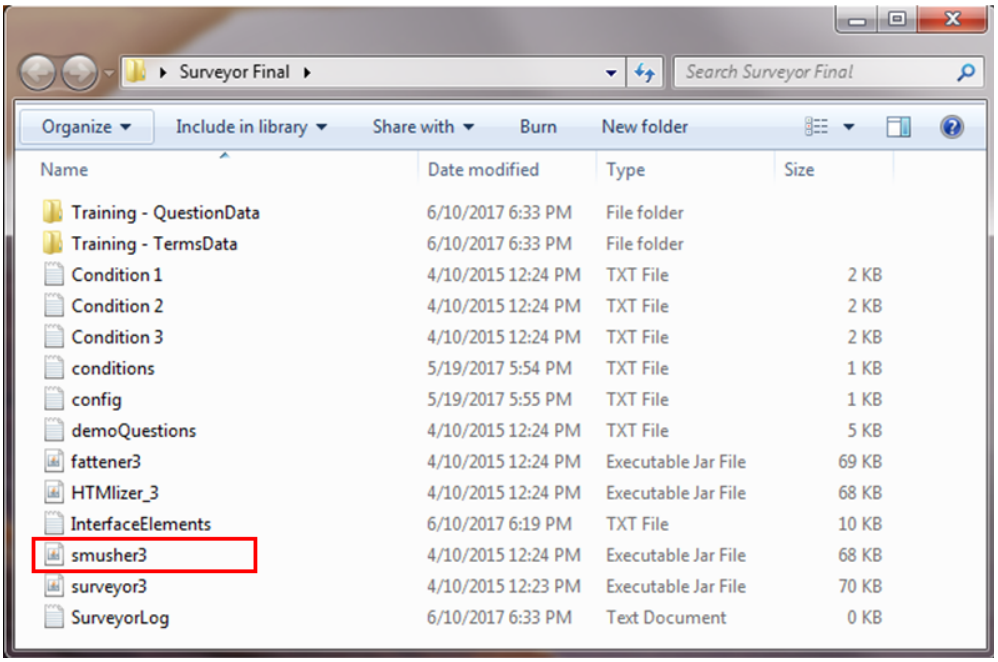

Figure 14. Suveyor 3.0's Smusher Utility

The smusher's utility interface will appear, shown in Figure 15. By default, the utility will try to identify term and question files in the Surveyor application folder for aggregation (this is one reason why we recommend not changing the names of these folders). To select another folder, select another directory in the *Look In* window. Once you have selected the directory listing the terms and question folders, select the QuestionData folder and then select the *Open* button. The smusher progress bar will then appear, report the percentage complete, and disappear once the smusher has completed aggregating the files.

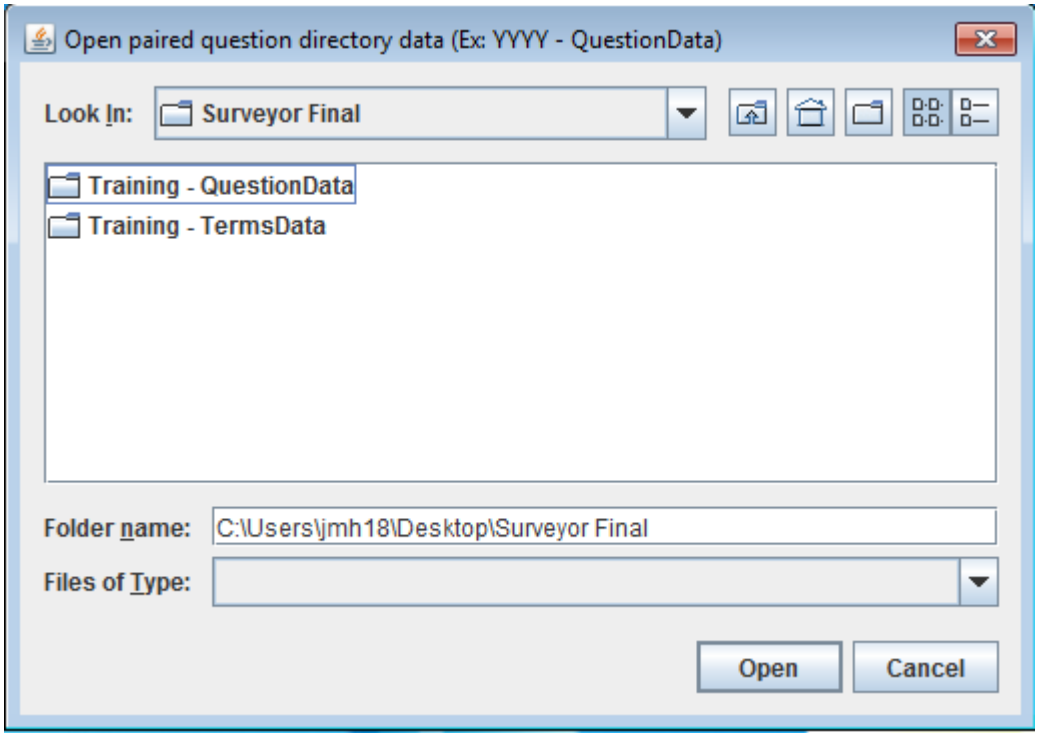

Figure 15. Suveyor 3.0 Smusher Utility Interface

After aggregating the files, the smusher will generate a new concept-level data set entitled "SurveyorData-yoursurvey-Standard.txt" (where "yoursurvey" is whatever you named your folder), shown in Figure 16. In Figure 16, the survey is entitled "Training", as seen in the Training-QuestionData and Training-TermsData folders. After smushing the folders, the resulting file's name is SurveyorData-Training-Standard.txt.

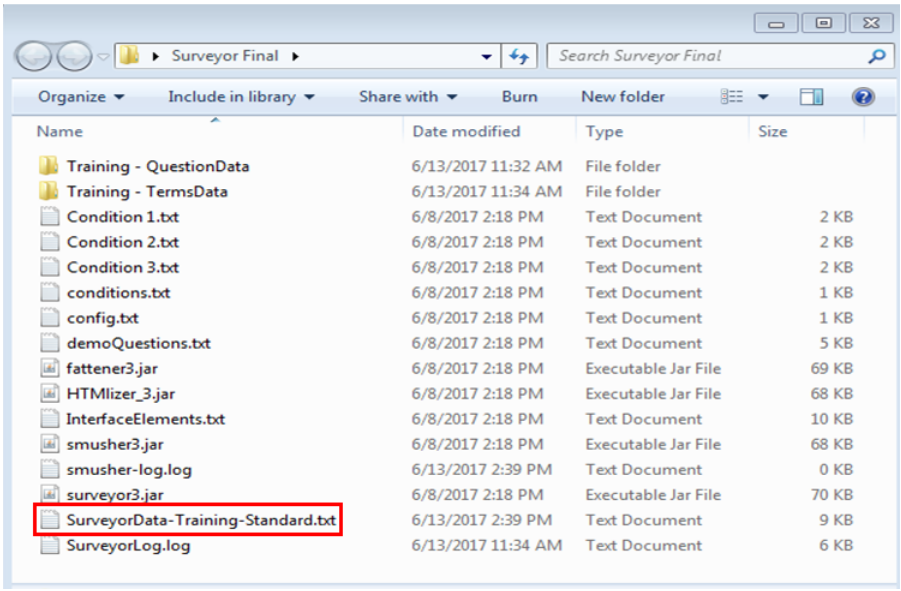

Figure 16. Surveyor 3.0 Concept-Level File.

To create a participant-level data set for the purpose of quality checks or other analyses, double click the fattener utility in the Surveyor 3.0 application folder, shown in Figure 17.

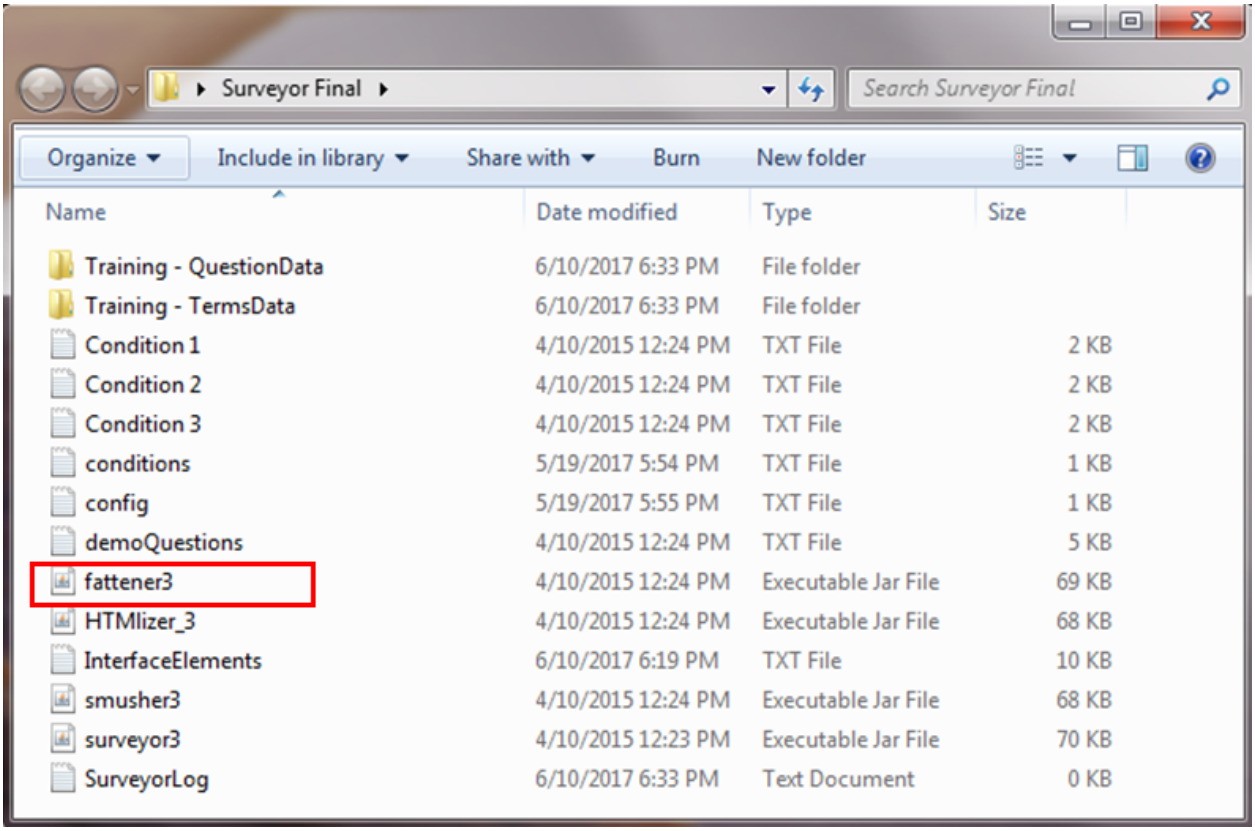

Figure 17. Surveyor 3.0's Fattener Utility

The fattener utility interface will appear, shown in Figure 18. Select the concept-level data set you wish to transform into a participant-level data set, in our case the file SurveyorData-Training-Standard.txt. Once you have selected the concept-level file, click on the Open button.

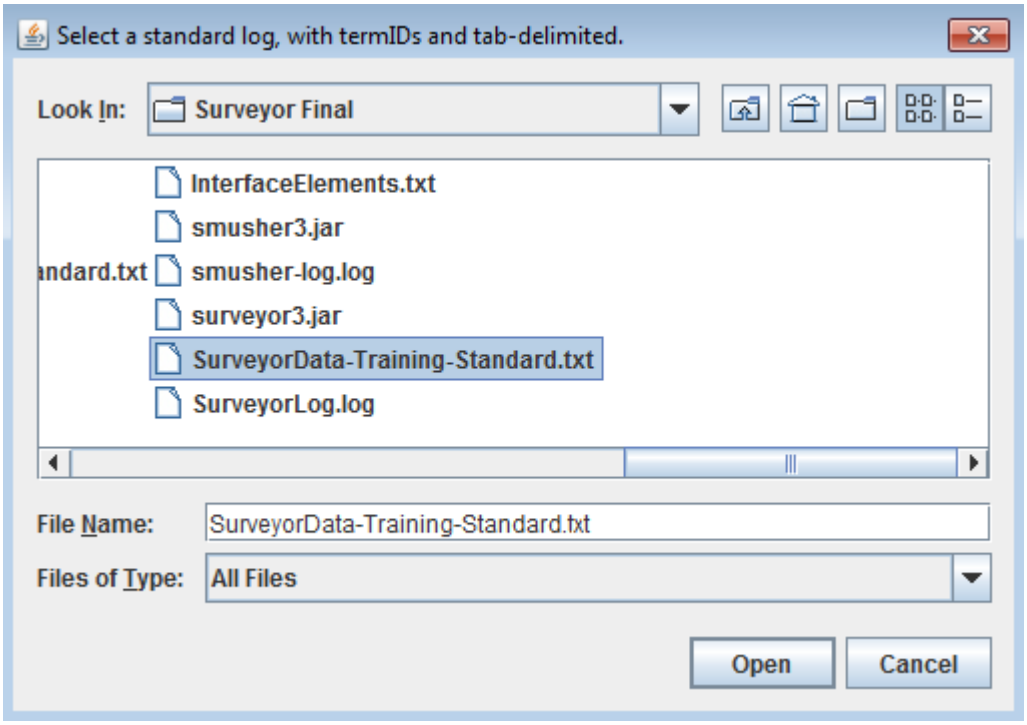

Figure 18. Surveyor 3.0's Fattener Utility

The fattener utility will next ask if you wish to submit a reserve term list, shown in Table 4. A reserve term list specifies a subset of concepts you wish to examine. Note, the reserve list expects a list of term IDs corresponding to the term IDs listed in the Concept files, see Table 3. If the researcher does not submit a reserve list, the fattener utility will create a column for each concept characteristic (e.g., time taken to rate the concept, slider values, and whether the concept was skipped or changed) for all concepts found in the selected file.

Table 4. Example Reserve List

My Reserve List
to_charm_someone
an_elder

Next, the fattener utility will ask you to identify which characteristics you wish to include. The fattener's default setting is to select all of Suveyor 3.0's data quality measures, resulting in a data set of *n* x 20 columns. These characteristics correspond to the data quality measures listed in Table 5. Once you have identified which characteristics you wish to include, select okay. The fattener utility will write a participant-level data set to the Surveyor 3.0's application folder, shown in Figure 19. Surveyor's participant-level data set is designed to facilitate data quality checks. **Note:** *The utilities presume that you will use the smusher first, to aggregate the files, and the fattener second, for data quality checks*. We now turn to Surveyor's data quality measures.

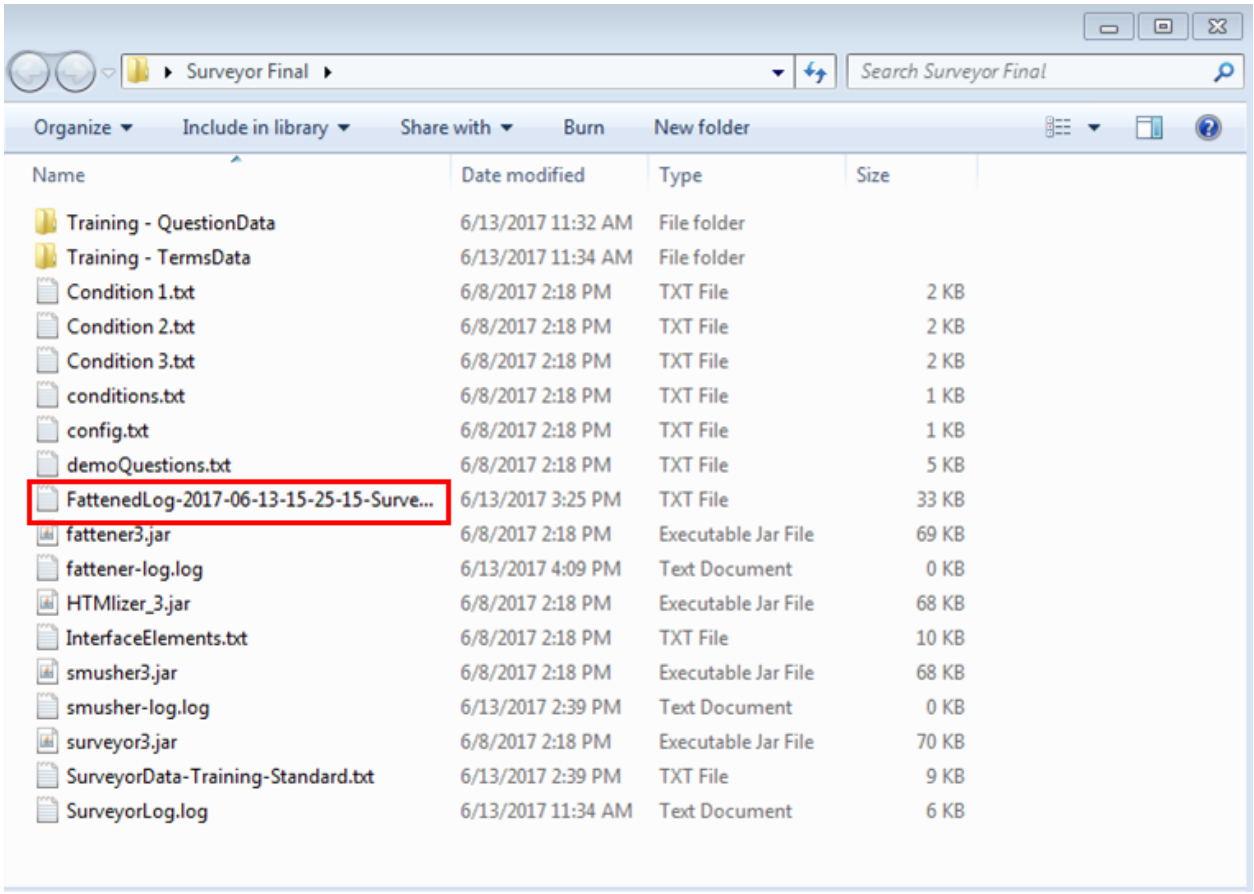

Figure 19. Example Participant-Level File Produced by Surveyor 3.0's Fattener Utility

Surveyor 3.0's Data Quality Measures

Surveyor provides sixteen measures to assess data quality. These measures when combined with Q-Factor analyses of the concept ratings provide a basic data quality picture (Rummel 1970; Heise 2010). We concentrate on Surveyor's data quality measure here; but for more information about applying Q-Factor analysis to the study of sentiments, we suggest reading Heise's discussion of sentiment analysis found in *Surveying Cultures* (2010, p. 167).

Table 4 lists Surveyor's quality and diagnostic measures. Surveyor's various measures help assess data quality in two ways. First, cultural variables (primary language, nativity, and the proportion of the respondent's life lived in the cultural context of interest) indicate the respondent's level of cultural familiarity. Second, rating statistics such as the average time spent on each concept, whether there are clear directionality biases (for example, the participant rated all the concepts by moving the slider to the right), and the skipping frequency provide measures of data quality. We recommend flagging respondents who on average spend 3 milliseconds or less per rating, whose ratings show a clear direction bias (90% or more of the ratings are in one direction), or who skip more than 35% of the ratings. We generally recommend excluding respondents who have failed either the cultural checks or who have been flagged

on two or more quality checks.[9] The diagnostic statistics such as presentation order and the *WasChanged* variable are useful for piloting, allowing the investigator to check for ordering effects or unclear concepts.

Table 5. Surveyor's Quality Measures, Measures Coming from Data Collected in Egypt (2014)

Quality Measures	Values
Primary Language	Egyptian Arabic
Egyptian Born	TRUE
Proportion Lived	For 91-100% of your life, you have lived in Egypt
E-SliderValue	46.51
P-SliderValue	69.77
A-SliderValue	23.26
E-Time (msec)	10594
P-Time (msec)	8938
A-Time (msec)	7625
E-AxisDirection	-1
P-AxisDirection	-1
A-AxisDirection	1
PresentationCategorical	EAP
E-PresentationOrder	1
P-PresentationOrder	3
A-PresentationOrder	2
WasChanged	FALSE
WasSkipped	FALSE
ReasonSkipped	-1

DISCUSSION AND CONCLUSION

In this report, we discussed the evolution of sentiment analysis from the pioneering work of Osgood and colleagues (1957; 1975) to modern studies using computer assisted collection techniques. We then introduced a new data collection platform, Suveyor 3.0, which was inspired by Heise's (2001) Surveyor platform. The development of this platform led to several technical and methodological innovations. Technical innovations include an explicit easily customizable configuration process that rapidly speeds up study implementation; data safety features such a human-readable log files, file overwrite controls, and password protected exit features; and data management utilities that facilitate data analysis. Methodological innovations include the elaboration of data quality procedures.

[9] These guidelines were developed by Dawn Robinson and Jesse Clark in conjunction with this effort.

The overall goal of this effort was to develop Surveyor 3.0, a flexible secure data collection tool capable of supporting distributed cross-cultural studies. The successful data collection efforts reported here resulted primarily from efforts of a dedicated transnational team of researchers who used this tool.

REFERENCES

Bradley, Margaret M., and Peter J. Lang. 1994. "Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential." *Journal of Behavior Therapy and Experimental Psychiatry* 25(1): 49-59.

Burke, Peter. J. 1980. "The Self: Measurement Implications from a Symbolic Interactinist Perspective." *Social Psychology Quarterly* 43(1): 18-29.

Cast, Alicia D., and Peter J. Burke. 2002. "A Theory of Self-Esteem." *Social Forces* 80(3): 1041-1068.

Cox, Eli. P. III. 1980. "The Optimal Number of Response Alternatives for a Scale: A Review." *Journal of Marketing Research* 17(4): 407-422.

Diederich, Gertrude W., Samuel J. Messick, and Ledyard R. Tucker. 1957. "A General Least Squares Solution for Successive Intervals." *Psychometrika* 22(2): 159-173.

Francis, Linda. 1997. "Ideology and Interpersonal Emotion Management: Redefining Identity in Two Support Groups." *Social Psychology Quarterly* 60(2): 153-171.

Freese, Lee, and Peter J. Burke. 1994. "Persons, Identities, and Social Interaction." Pp. 1-24 in *Advances in Group Processes,* vol. 11, edited by Barry Markovsky, Karen Heimer, and Jodi O'Brien. Greenwich, CT, JAI.

Heise, David. R. 1978. *Computer-Assisted Analysis of Social Action: Use of Program INTERACT and SURVEY.UNC75*. Chapel Hill, NC, Institute for Research in the Social Sciences (IRSS).

Heise, David R. 1979. Understanding Events: Affect and the Construction of Social Reality. Cambridge, UK, Cambridge University Press.

Heise, David. R. 2001. "Project Magellan: Collecting Cross-Cultural Affective Meanings via the Internet." *Electronic Journal of Sociology* 5(3).

Heise, David R. 2007. *Expressive Order: Confirming Sentiments in Social Action*. Bloomington, IN, Springer.

Heise, David R. 2010. *Surveying Cultures: Discovering Shared Conceptions and Sentiments*. Hoboken, NJ, Wiley.

Hogg, Michael A., Deborah J. Terry, and Katherine M. White. 1995. "A Tale of Two Theories: A Critical Comparison of Identity Theory with Social Identity Theory." *Social Psychology Quarterly* 58(4): 255-269.

Kahneman, Daniel. 1963. "The Semantic Differential and the Structure of Inferences among Attributes." *The American Journal of Psychology* 76(4): 554-567.

Osgood, Charles E. 1953. *Method and Theory in Experimental Psychology*. New York, NY, Oxford University Press.

Osgood, Charles E. 1962. "Studies on the Generality of Affective Meaning Systems." *American Psychologist* 17(1): 10-28.

Osgood, Charles. E., William H. May, and Murray S. Miron. 1975. *Cross-Cultural Universals of Affective Meaning*. Urbana, IL, University of Illinois Press.

Osgood, Charles E., George C. Suci, and Percy H. Tannenbaum. 1957. *The Measurement of Meaning*. Urbana, IL, University of Illinois Press.

Robinson, Dawn T., and Lynn Smith-Lovin. 1992. "Selective Interaction as a Strategy for Identity Maintenance: An Affect Control Model." *Social Psychology Quarterly* 55(1): 12-28.

Rummel, Rudolf J. 1970. *Applied Factor Analysis*. Evanston, IL: Northwestern University Press.

Saris, Willem E., and Irmtraud N. Gallhofer. 2014. *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. 2nd ed. Hoboken, NJ, John Wiley & Sons, Inc.

Serpe, Richard T. and Sheldon Stryker. 1987. "The Construction of Self and Reconstruction of Social Relationships." Pp. 41-66 in *Advances in Group Processes*, edited by Edward Lawler and Barry Markovsky. Greenwich, CT: JAI.

Smith-Lovin, Lynn. 1987. "Impressions from Events." *Journal of Mathematical Sociology* 13(1-2): 35-70.

Smith-Lovin, Lynn., and David R. Heise, Ed. 1988. *Analyzing Social Interaction: Advances in Affect Control Theory*. New York, NY, Gordon and Beach Science Publishers.

Stryker, Sheldon. [1980] 2002. *Symbolic Interactionism: A Social Structural Version*. New York, NY, Blackburn Press.

Stryker, Sheldon. 2004. “Integrating Emotion into Identity Theory.” Pp. 1-23 in *Advances in Group Processes*, vol. 21, edited by J. H. Turner. Emerald Group Publishing.

Weinreich, Uriel. 1980. "Travels through Semantic Space." Pp.14-36 in *On Semantics,* edited by W. Labov and B. S. Weinreich. Philadelphia, PA, University of Pennsylvania Press.

APPENDIX A

Surveyor 3.0 Log File

For details, see Morgan G.P. & Morgan J.H. (2014)

Loading master config file (config.txt) ...

No 'randomizeQuestionOrder' config specified, using default: false

No 'TERM_FONT_STYLE' config specified, using default: SanSerif

No 'QUESTION_FONT_STYLE' config specified, using default: SanSerif

No 'RESPONSE_FONT_STYLE' config specified, using default: SanSerif

No 'DESCRIPTOR_FONT_STYLE' config specified, using default: SanSerif

No 'TUTORIAL_FONT_STYLE' config specified, using default: SanSerif

No 'TUTORIAL_FONT_SIZE' config specified, using default: 14

No 'WRITE_FILES' config specified, using default: true

No 'SHOW_PROGRESS_BAR' config specified, using default: true

No 'TUTORIAL_FONT_COLOR' config specified, using default: java.awt.Color[r=0,g=0,b=255]

No 'MIN_SLIDER_VAL' config specified, using default: -43

No 'MAX_SLIDER_VAL' config specified, using default: 43

No 'NEG_HIGH_TICK_VAL' config specified, using default: -43

No 'NEG_MEDHIGH_TICK_VAL' config specified, using default: -30

No 'NEG_MEDLOW_TICK_VAL' config specified, using default: -20

No 'NEG_LOW_TICK_VAL' config specified, using default: -10

No 'NEUTRAL_TICK_VAL' config specified, using default: 0

No 'POS_LOW_TICK_VAL' config specified, using default: 10

No 'POS_MEDLOW_TICK_VAL' config specified, using default: 20

No 'POS_MEDHIGH_TICK_VAL' config specified, using default: 30

No 'POS_HIGH_TICK_VAL' config specified, using default: 43

No 'ALLOW_USER_TO_EXIT' config specified, using default: false

Loading conditions.

Loaded Condition: Condition 1,Training,Condition 1.txt,demoQuestions.txt,InterfaceElements.txt

Loaded Condition: Condition 2,Training,Condition 2.txt,demoQuestions.txt,InterfaceElements.txt

Loaded Condition: Condition 3,Training,Condition 3.txt,demoQuestions.txt,InterfaceElements.txt

Submit Button for Condition Dialog has been entered.

Loading interface elements from 'InterfaceElements.txt'

Finished reading interface text elements.

Element 'FileOverWrite_Title' not defined. Using 'User and condition already logged today!' as the default.

Element 'Tutorial_NextButtonText' not defined. Using 'Next' as the default.

Element 'CancelButtonText' not defined. Using 'Cancel' as the default.

Element 'CHOICE_QUESTION_TEXT' not defined. Using 'Select a term to edit:' as the default.

Element 'ExitButtonText' not defined. Using 'Exit' as the default.

Element 'ExitPrompt' not defined. Using 'To exit, please type the experimenter exit password:' as the default.

Element 'Tutorial_IntroPopupTitle' not defined. Using 'Warning: Tutorial Ahead!' as the default.

Element 'FileOverWrite_Message' not defined. Using 'A file for this user name and condition was already submitted today.

The application will exit. Please see your proctor to report this message.' as the default.

Element 'SkipQuestionDialogTitle' not defined. Using 'Skipping Term?' as the default.

Element 'OnlyRightAnswerPopUp' not defined. Using 'Your answer to this question was not correct. The application will now close.' as the default.

Element 'ChangeTermTitle' not defined. Using 'Change Term Entry' as the default.

Element 'Tutorial_ExampleTerm3' not defined. Using 'a polymath' as the default.

Element 'Tutorial_ExampleTerm2' not defined. Using 'a stranger' as the default.

Element 'Tutorial_ExampleTerm1' not defined. Using 'helping someone' as the default.

Loading terms from 'Condition 2.txt'.

Term File processing complete. Loaded 24 terms correctly.

Loading questions from 'demoQuestions.txt'.

Question File processing complete. Loaded 11 correctly.

Constructing UI elements...

Constructing Demographic Question Panels.

Demographic Question Panels constructed.

Constructing Term Panels.

Term Panels constructed.

Shuffling terms.

Generating panel queues

Showing Interface.

User identified as 2

Checking that this user hasn't already participated today.

Creating data writers.

Begin working through questions and terms.

Submit Button for Question 'Location' has been entered.

Slider value changed. Current value is: -4.65

Slider value changed. Current value is: 11.63

Slider value changed. Current value is: -2.33

Slider value changed. Current value is: 48.84

Slider value changed. Current value is: 51.16

Slider value changed. Current value is: 51.16

Slider value changed. Current value is: -2.33

Slider value changed. Current value is: 30.23

Slider value changed. Current value is: 30.23

Slider value changed. Current value is: 0

Slider value changed. Current value is: 18.6

Slider value changed. Current value is: 37.21

Showing panel for Term ' i_queer_person-P'.

Slider value changed. Current value is: -20.93

Showing panel for Term ' i_queer_person-E'.

Slider value changed. Current value is: 11.63

Showing panel for Term ' i_queer_person-A'.

Slider value changed. Current value is: 32.56

Showing panel for Term ' i_Latino/a-A'.

Slider value changed. Current value is: 16.28

Showing panel for Term ' i_Latino/a-E'.

Slider value changed. Current value is: 37.21

Showing panel for Term ' i_Latino/a-P'.

Slider value changed. Current value is: -16.28

Showing panel for Term 'b_respecting-E'.

Slider value changed. Current value is: 62.79

Showing panel for Term 'b_respecting-P'.

Slider value changed. Current value is: -25.58

Slider value changed. Current value is: -11.63

Showing panel for Term 'b_respecting-A'.

Slider value changed. Current value is: -34.88

Showing panel for Term ' i_black_person-E'.

Slider value changed. Current value is: 62.79

Showing panel for Term ' i_black_person-A'.

Slider value changed. Current value is: 16.28

Showing panel for Term ' i_black_person-P'.

Slider value changed. Current value is: 9.3

Slider value changed. Current value is: -6.98

Showing panel for Term ' i_gay_person-A'.

Application exited with experimenter input.

Finalizing Terms Log.

System closing.

APPENDIX B

Table 6: The Interface File from Studies Occurring in Egypt, Kuwait, and Morocco (arabicInterfaceElements.txt)

Element	Value
TitleBar	Duke Surveyor 3.0
SubmitButtonText	<html> <font size="6">حفظ</html>
ChangeButtonText	<html> <font size="6">تغيّير</html>
SkipButtonText	<html> <font size="6">تخطيّ</html>
EPA_E_High1	جيد
EPA_E_High2	جميل
EPA_E_Low1	سيّئ
EPA_E_Low2	قبيح
EPA_P_High1	قوي
EPA_P_High2	ضخم
EPA_P_Low1	ضعيف
EPA_P_Low2	صغير
EPA_A_High1	صاخب
EPA_A_High2	سريع
EPA_A_Low1	هادئ
EPA_A_Low2	بطيء
EPA_Magnitude_Zero	محايد
EPA_Magnitude_Low	قليلا
EPA_Magnitude_MedLow	الي حد ما
EPA_Magnitude_MedHigh	للغاية
EPA_Magnitude_High	بلا حدود
Tutorial_IntroPopupTitle	تعليمات
CancelButtonText	<html> <font size="6">إلغاء</html>
Tutorial_NextButtonText	<html> <font size="6">حسنا</html>
CHOICE_QUESTION_TEXT	<html> <font size="6">لتغييرها مفهوم اختيار </html>
Skip_Question_Prompt	<html> <font size="5">المفهوم هذا تخطي أنت لماذا</html>
Skip_Question_Ignorance	<html> <font size="5">المفهوم هذا تعنيه ما أعرف لا</html>
Skip_Question_Offended	<html> <font size="5">متضايق يجعلني المفهوم هذا</html>
Skip_Question_Other	<html> <font size="5">تخطيها أن فقط أريد و آخر سبب</html>
OnlyRightAnswerPopUp	<html> <font size="6">البرنامج يغلق سوف هذا و صحيح غير السؤال هذا على جوابك كان الآن</html>
Exit_Text	<html> <font size="6">الدراسة أكملت قد كنت و شاركتكم على شكرا
Tutorial_IntroPopup	<html> <font size="8">استخدام كيفية شرح نقوم سوف الدراسة بدء قبل الواجهة </html>
Tutorial_BeforeTermAppears	<html> <font size="8">مفاهيم تجاه مشاعرك بتقيم ستقوم المشروع هذا في يتم يفـك لترى "حسنا" الزر على بالنقر قم المفهوم تشرح عبارة أو كلمة عليك سيعرض مختلفة المفاهيم أحد عرض </html>
Tutorial_BeforeSliderAppears	<html> <font size="8">بهيش مقياس خلال من معين مفهوم عن مشاعرك بتقيم ستقوم لترى "حسنا" زر على بالنقر قم لأخرى جهة من تحريكه يمكن ساحب المسطرة لهذه المسطرة المسطرة </html>
Tutorial_SliderHasAppeared	<html> <font size="8">على الفأرة بوضع قم ،التقيم بممارسة قم ألان في يه عندما الفأرة زر بإفلات قم المسطرة على الساحب جر و الفأرة زر على بالكبس قم الساحب حسنا" الزر على بالنقر قم ثم تريده الذي المكان </html>

Tutorial_BeforeEndTermsAppear	<html> <font size="8">يمكن التي الضروف أقصى يمثلان المسطرة طرفي ترى لكي "ناحس" زر على بالنقر قم محايد المتقف وبالتالي الطرفين من أياً يمثل لا المنتصف تخيلها المسطرة بنهايتي المقصود </html>
Tutorial_EndTermsEvaluation	<html> <font size="8">"سيئ" مقابل في "جيد" تمثلان هنا المسطرة نهايتي التي طرافالأ من أخر زوجاً لترى "حسنا" على بالنقر قم المقياس على "السيئ"و "الجيد" عكس سيتم أحياناً مفهوم كل مع ستستعملها </html>
Tutorial_EndTermsPotency	<html> <font size="8">أو "قوي" مفهوماً كان إذا ما بتقيم لك تسمح المفاهيم هذه لترى "حسنا" لىء بالنقر قم المثالي التفكير على وليس الأول انطباعك على تقيمك ببناء قم "ضعيف" مفهوم كل مع ستستعملها التي الأطراف من ثالثاً زوجاً </html>
Tutorial_EndTermsActivity	<html> <font size="8">قم ما مفهوم حيوية تجاه بمشاعرك تتعلق الأطراف هذه حينما " حفظ " زر على بالنقر قم خامل اله أم حيوي الحال بطبيعة السلوك هذا هل ما شخص مساعدة بتقيم تريده الذي المكان على المؤشر تضع </html>
Tutorial_RateTermNeutral	<html> <font size="8">م لتقي ما مفهوم بتقيم لتقوم المؤشر تحريك من لابد طريق عن نألا ذلك بممارسة قم بإعادته قم ثم المنتصف عن بعيداً المؤشر بتحريك قم "محايد" إنه على شيء محايد إنه على "الغريب" تقيم </html>
Tutorial_NotNeutralPopup	<html> <font size="8">قيّم للمتابعة محايد أنه على المفهوم </html>
Tutorial_DiscussChangeButton	<html> <font size="8">ندع "تغيير" زر على بالنقر قم ،معين تقييم لتغير فذتنا ستظهر و "تغيير" زر على بالنقر قم ألان بتقييمها قمت التي بالمفاهيم قائمة ستظهر ذلك جديدة للإلغاء الجديدة فذتالنا اغلاق "حسنا" الزر على بالنقر قم ثم </html>
Tutorial_DiscussSkipButton	<html> <font size="8">الزر على بالنقر قم ثم المفهوم تعترف لا كنت إذا أخرى ذةنافذ ستظهر الموسوعي معنى تعرف كنت لو حتى "تخطي" الزر على بالنقر قم الآن "تخطي" "حفظ" الزر على بالنقر وقم سببا اختيار تخطي أن ترغب كنت لماذا الأسباب من قائمة مع </html>
Tutorial_CloseMessage	<html> <font size="8">ىعل بالنقر قم المفاهيم تقيم في تبدأ أن ألان يمكنك جاهزاً تكون عندما الزر </html>
Tutorial_EndTutorialButtonText	<html> <font size="8">حسنا</html>
ProgressBarLabel	<html> <font size="6">التقدم</html>
Tutorial_ExampleTerm1	مساعدة شخص ما
Tutorial_ExampleTerm2	الغريب
Tutorial_ExampleTerm3	الموسوعي