Exemplar Generation

Assessing How Respondents Give Meaning to Rating Scales

Thomas M. Ostrom and Katherine M. Gannon

Many surveys use multiple interval rating scales to gather information. These rating scales provide a preset response format for the respondents while at the same time allowing for flexibility in gradients of intensity or certainty. From the researcher's perspective, choosing to use a closed response format may be based on considerations of minimizing difficulties associated with coding and analyzing the respondent's answers. This raises an important question, however. How does the respondent give meaning to rating scales?

Multiple interval rating scales are also used extensively in contexts other than surveys. For example, laboratory-oriented attitude researchers use a variety of multiple interval rating scales, such as one-item scales, semantic differential scales, and Likert scales to assess attitudes and attitude change. The one-item rating scale may use scale labels anchored by "very opposed to gun control" and "very supportive of gun control." The semantic differential scale (Osgood, Suci, and Tannenbaum, 1957) measures attitudes through a series of bipolar adjective scales with end labels such as "good" and "bad" or "fair" and "unfair." The Likert scale (Likert, 1932) contains a set of belief statements that are responded to on a scale ranging from "strongly agree" to "strongly disagree."

How do respondents make sense of survey questions presented in a multiple interval rating scale response format? In this chapter
The categorical model of thinking scales is based on the idea that the language of response options affects information processing and the construction of response options. If we present an example, let us consider the question: "Which of the following statements is true? A. categorical, B. complex, C. ambiguous.

Let's consider the two scales: one that is based on categorical thinking and one that is based on complex thinking. The categorical scale is likely to lead to a "yes" or "no" response, while the complex scale is likely to lead to a more nuanced response.

In the case of the categorical scale, the response options are "yes," "no," or "unsure." In the case of the complex scale, the response options are "true," "false," or "unsure." The complex scale is likely to lead to a more nuanced response, while the categorical scale is likely to lead to a "yes" or "no" response.

If we were to present an example, let us consider the following question: "Which of the following statements is true? A. categorical, B. complex, C. ambiguous."

The categorical scale is likely to lead to a "yes" or "no" response, while the complex scale is likely to lead to a more nuanced response.

In the case of the categorical scale, the response options are "yes," "no," or "unsure." In the case of the complex scale, the response options are "true," "false," or "unsure." The complex scale is likely to lead to a more nuanced response, while the categorical scale is likely to lead to a "yes" or "no" response.

If we were to present an example, let us consider the following question: "Which of the following statements is true? A. categorical, B. complex, C. ambiguous."

The categorical scale is likely to lead to a "yes" or "no" response, while the complex scale is likely to lead to a more nuanced response.

In the case of the categorical scale, the response options are "yes," "no," or "unsure." In the case of the complex scale, the response options are "true," "false," or "unsure." The complex scale is likely to lead to a more nuanced response, while the categorical scale is likely to lead to a "yes" or "no" response.
Suppose a respondent has seen another person return a very small amount of change to a cashier and is then asked to rate the person’s honesty on a rating scale anchored at either end by “not at all honest” and “very honest.” In answering the question, the highly prototypical exemplar “returned a lost wallet containing money,” which was given earlier, may be most accessible upon activation of the honest category. The person then locates the knowledge about the target—he returns change—on the internal gradient of the category honest. In this case, it is not the best fit instance, “returning a wallet with money still in it,” but it is highly representative of the category. The respondent integrates this into an overt response and selects perhaps “moderately honest.”

The categorical model differs from other models about judgment reports on rating scales. Other models may address the meaning of the endpoints on the scale but not the way in which respondents select the appropriate point on the rating scale gradient. For example, according to the perspective theory, the rating scale anchors determine the range of alternatives considered (Ostrom and Upshaw, 1968). Raters are believed to anchor the two end categories with beliefs, feelings, and actions that correspond to those labels. This theory bears some resemblance to the basic premise of the categorical model. But a key difference is that the perspective theory provides no dynamic for understanding how meaning is given to the intermediate intervals on the rating scale gradient. It offers a simple interpolation heuristic with no psychological justification.

Another related social judgment theory is Hovland and Sherif’s assimilation contrast theory (1952), which states that a person’s attitude serves as an anchor and the stimuli close to it are assimilated whereas those distant from it are contrasted. This theory, however, does not address how one determines where own’s attitude is located on the rating scale in the first place. Other models, such as Tourangeau and Rasinski (1989) and Strack and Martin (1987), address the reporting of a judgment but have little to say about the cognitive processes involved in mapping the judgment to the particular scale point.

The Exemplar Generation Methodology

In this section of the chapter, we will discuss a new methodology that has been used to investigate categorical meaning of graded rating scales.

The exemplar generation methodology uncovers the cognitive meaning given to rating scale responses by having respondents provide instances for each of the scale labels (see Exhibit 12.1 for sample instructions). That is, rather than having a respondent rate an instance in terms of a scale, the task is reversed and the respondent is asked to generate instances that would exemplify each of the rating scale labels.

This first step provides the researcher with the content associated with each of the scale labels. Next, the respondents are asked to rate each of their own exemplars in terms of how easy or difficult they were to generate (see Exhibit 12.2 for sample instructions). The analysis of these ease ratings can be used to assess the categorical properties of the ratings scale. Specifically, the best-fit instances in a category are those which are highly accessible and come to mind easily. The ease ratings, then, can be used to index the internal gradient of the categories used to give meaning to the scale.

Thus, a set of respondent-generated exemplars provides a rich array of actions, feelings, and beliefs evoked by the rating scale labels. Exhibit 12.3 illustrates exemplars generated for the traits kind and stupid, based on a probability gradient. Even more important to this methodology are the ease ratings. These provide information about which intervals on the rating scale most closely correspond to the respondent’s activated cognitive categories.

A related form of the exemplar generation methodology has occasionally been used by researchers investigating categorical structure (Barsalou, 1985; Battig and Montague, 1969; Rosch, 1975). Subjects were asked simply to list members of categories or attributes of categories that came to mind. This research, however, did not address the cognitive conditions in which an external gradient is provided for respondents, which is the case when using a rating
Exhibit 12.1. Example Instructions for Exemplar Generation Task.

Conventionally graded scale

Below is a rating scale for the personality trait of "honesty." Your task is to give descriptions of behaviors characteristic of each of the labels on the rating scale. For example, if the personality trait you were writing for was "kindness," "helped a neighbor mow the lawn" might be an example of a behavior you might have written for one of the "kind" scale labels. This behavior would indicate that the person performing the behavior would appear to be kind to the degree indicated by the specific label.

Not at all 1 2 3 4 5 6
Not honest Slightly honest Somewhat honest Moderately honest Fairly honest Very honest Completely honest

On the lines provided, write a behavior that corresponds to each label on the rating scale that describes how honest a person performing the behavior would seem.

Note: Scale labels are listed in either scrambled or graded order.

Probability graded scale

During this experiment, we would like you to generate examples of behaviors that fit at different levels on the probability scale. As you know, probabilities range on a continuum from 0 to 1.0 where 0 is the lowest probability and 1.0 is the highest probability. During this study we would like you to generate an example of a behavior that you see as fitting each of the eleven probability levels on the scale given below.

0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0
Lowest probability Highest probability

For example, if you were generating behaviors for the trait "kind," you might write "helped a neighbor mow the lawn" as a behavior that fits at the probability level of .8. This means that a person who would do this behavior has a .8 probability of being of being kind. But you might write "laughed when they saw an elderly person fall" as a behavior that fits the probability level of .2.

0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0
Lowest probability Highest probability

Your task is to provide examples of behaviors that an honest person might do at several different levels of probability. Write them on the lines below.

Note: Probability labels are listed in either scrambled or graded order.

Exhibit 12.2. Example Scale and Instructions for Rating Ease of Exemplar Generation.

Ease rating instructions

We would like you to rate how easy or difficult it was for you to come up with examples of behaviors for each of the rating scale labels. Feel free to go back to the first booklet if it will help you make your ratings. For each of the following labels, circle the number that corresponds to how easy or difficult it was for you to generate a behavior.

Scale label: Fairly Honest

1 2 3 4 5 6 7
Very difficult Very easy

Note: Probability labels are listed in either scrambled or graded order.

scale format. Nor was this research concerned with the way in which respondents coordinate the category exemplars with such an external gradient. The exemplar generation methodology allows for an examination of production of instances when constrained by the rating scale format.

The key to the present approach is the assumption that instances generated for the rating scale endpoints correspond to the best-fit (or ideal) instances and so may be more prototypical. These instances are more accessible as indicated by the increased ease of generation in comparison with those instances generated for the intermediate levels. This reflects the graded categorical structure. Figure 12.1 shows a typical outcome of the exemplar generation methodology.

The figure presents the results of one study of probability-gradi
dent ratings scales in which subjects generated exemplars for the personality traits kind and intelligent (Carnot, 1990). Consistent with the categorical model, exemplars generated for the endpoint labels, 0.0 lowest probability and 1.0 highest probability, were the easiest to generate. This indicates that the respondents accessed cognitive categories to provide meaning to those endpoint labels.
Exhibit 12.3. Example Exemplars Generated for Probability Scales.

\[ Trait = \text{Stupid} \]

<table>
<thead>
<tr>
<th>Probability</th>
<th>Exemplar</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>They studied three days for an exam and got plenty of sleep the night before the test.</td>
</tr>
<tr>
<td>.1</td>
<td>The person left one answer blank on a multiple choice test.</td>
</tr>
<tr>
<td>.2</td>
<td>They forgot to go to an important meeting.</td>
</tr>
<tr>
<td>.3</td>
<td>The person didn’t go to a free class offering self defense when there have been five attacks in their neighborhood.</td>
</tr>
<tr>
<td>.4</td>
<td>The person didn’t lock their doors and windows when seven burglaries had happened in the neighborhood within the last month.</td>
</tr>
<tr>
<td>.5</td>
<td>The person got drunk and picked a fight with a person three times his size.</td>
</tr>
<tr>
<td>.6</td>
<td>The person didn’t study for a huge final on purpose and went out and got really drunk the night before the test.</td>
</tr>
<tr>
<td>.7</td>
<td>The person left all the answers blank on a multiple choice test.</td>
</tr>
<tr>
<td>.8</td>
<td>The person left her animal in a car with no windows cracked on a 98 degree day.</td>
</tr>
<tr>
<td>.9</td>
<td>The person refused to go to the hospital after getting into a severe car accident and they were bleeding terribly.</td>
</tr>
<tr>
<td>1.0</td>
<td>The person scored an 8 on the ACT math level.</td>
</tr>
</tbody>
</table>

\[ Trait = \text{Kind} \]

<table>
<thead>
<tr>
<th>Probability</th>
<th>Exemplar</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Cold-blooded murder.</td>
</tr>
<tr>
<td>.1</td>
<td>Seeing a little child on the street and beating him up.</td>
</tr>
<tr>
<td>.2</td>
<td>Taking money from a blind person.</td>
</tr>
<tr>
<td>.3</td>
<td>Betraying a friend to get ahead.</td>
</tr>
<tr>
<td>.4</td>
<td>Hitting a person for no reason.</td>
</tr>
<tr>
<td>.5</td>
<td>Telling someone exactly what you think, even if rude.</td>
</tr>
<tr>
<td>.6</td>
<td>Putting off what you have to do to help someone.</td>
</tr>
<tr>
<td>.7</td>
<td>Helping an old lady across the street.</td>
</tr>
<tr>
<td>.8</td>
<td>Stopping on the highway to help a stranded car.</td>
</tr>
<tr>
<td>.9</td>
<td>Giving your last dollar to someone who needs it.</td>
</tr>
<tr>
<td>1.0</td>
<td>Doing anything to help a stranger, that is, giving money, a ride, the time to help.</td>
</tr>
</tbody>
</table>

Evidence that the end categories of a rating scale have special cognitive status emerges from two other methodologies for studying categorial structure. The endpoints' labels were found to be the most accessible instances as assessed by these two other methodologies. One methodology used in categorization research is to measure the time it takes to respond to whether an instance is a category member. Response time is found to be faster for instances rated as more typical than for those rated as less typical (Glass, Holyoak, and O'Dell, 1974; Rips, Shoben, and Smith, 1973; Rosch, 1973). Another methodology has measured typicality by having subjects rate the extent to which an instance fits their image of a category. The mean of the subjects' responses for each instance has been used as an index of typicality (Rosch, 1975).
The U-shaped pattern of ease ratings shown in Figure 12.1 is robust despite changes in some of the procedural details. The pattern occurs in both cases in each of the following instances: (1) Exemplars are generated per page or all on the same page. (2) The order of generation is random or in scale order. (3) Scale intervals are designated verbally or numerically. (4) The rating scale has seven or eleven intervals. (5) The ease-of-generation scale has seven or nine intervals. For example, in studies conducted by Carnot (1990) the order in which the exemplars were generated was randomly determined and exemplars were generated one per page. The rating scale consisted of eleven labels with only the endpoint labels verbally labeled. The ease ratings were on a scale with nine intervals. In studies conducted by Gannon (1991), the scale labels appeared in scale order with all labels for the scale on the same page. The rating scales had seven intervals with all intervals labeled verbally. The ease ratings were on a scale with seven intervals.

Relevance to Rating Scales

The exemplar generation methodology allows us to examine a key issue in the categorical basis of rating scales. This relates to how many cognitive categories are activated in giving meaning to the scale.

Activation of Multiple Categories

The ease ratings in Figure 12.1 illustrate a number of facets of this issue. The probability rating scale only explicitly activates one cognitive category. As the example in Exhibit 12.1 shows, only one trait was designated by the instructions, the honesty trait. Asking a respondent to generate an exemplar that would be present at a probability of 1.0 for an honest person explicitly activates the category honesty. It is an open question if more than that one category becomes activated by the labels of the probability scale.

Figure 12.1 raises the possibility that two other cognitive categories could be activated by the rating scale. The ease ratings show a U-shaped pattern in which it is not only easy to generate exemplars at the explicit end of the scale (where the probability is 1.0) but also easier to generate exemplars at the implicit end of the scale (where the probability is 0.0). Presumably, the cognitive category dishonesty is implicitly activated to give meaning to the zero probability end of the scale.

It is also possible that a third cognitive category is activated by the scale. Figure 12.1 shows a small elevation in the center of the probability scale (where the probability is 0.5). The category here may be of uncertainty, ambiguity, or irrelevance as defined by the designated trait.

When multiple categories are activated, the strength of activation may vary from category to category. The ease ratings in Figure 12.1 show an asymmetry to the U-shaped pattern. The ease ratings are higher at the end of the scale where the trait is explicitly designated (probability 1.0) than at the end where the trait is only implicit in the scale labels (probability 0.0). The difference in ease ratings between the explicit and implicit ends of the probability rating scale is reliable. It holds up even when counterbalancing for trait valence (that is, when for some respondents the designated trait is honest while for others it is dishonest).

The difference between the two ends suggests that a single activation process might be at work for probability rating scales. Only one trait is explicitly mentioned in the instructions and so the corresponding cognitive category becomes activated. The opposite end of the scale is given meaning by searching for a category that stands in contrast to the designated category. Some rating scales explicitly activate cognitive categories for both ends of the scale. An example is the semantic differential scale where terms like kind and unkind are used to anchor the scale. This dual activation process should lead to equally strong ease ratings given to exemplars generated for both ends of the scale.
Typology of Rating Scale Formats

A typology of rating scale formats was developed based on a survey of formats used in the social psychological literature (Gannon, 1991). Attention was paid to the presence or absence of explicit labels that would activate categories. The survey excluded scales of physical values such as those assessing weight or height. (See Table 12.1.)

Four types of rating scale formats were seen to be prominent in the literature: bipolar, unipolar, agree/disagree, and probability scale formats. Bipolar scale formats are those that explicitly label two categories (for example, very dishonest, moderately dishonest, slightly dishonest, neither dishonest nor honest, slightly honest, moderately honest, very honest). Unipolar scale formats are those that explicitly label only one category (not at all honest, slightly honest, somewhat honest, moderately honest, fairly honest, very much honest, completely honest).

An additional characteristic of the rating scale formats is the nature of the terminology used to convey the gradient. Agree/disagree scale formats explicitly label one category and convey the gradient in terms of agreement versus disagreement (strongly disagree, moderately disagree, slightly disagree, neither disagree nor agree, slightly agree, moderately agree, strongly agree). Probability scale formats also explicitly label one category and convey the gradient in terms of probability or likelihood (completely improbable, very improbable, moderately improbable, uncertain, moderately probable, very probable, completely probable).

The exemplar generation methodology was used to examine the categorical activation of these four scale formats (Gannon, 1991). The results indicated that there are essentially two patterns of activation (see Figure 12.2). The bipolar scale format and the agree/disagree format produced the dual activation pattern of ease ratings. That is, the two endpoints were nearly equivalent in terms of ease of generation and the intermediate intervals farthest from the endpoints were the most difficult to generate. The pattern of ease ratings was essentially symmetrical. The midpoint was found to be easier to generate exemplars for than the surrounding labels, which suggests that this scale activated a third category (that is, irrelevant to both end categories).

The unipolar and the probability scale formats produced ease ratings that corresponded to the single activation pattern. The ease ratings were asymmetrical. However, the implicit endpoint label (that is, not at all honest or not at all probable) was a relatively easy label for which to generate exemplars. This pattern suggests that the contrast category of the explicit category label may have been partially activated.

To investigate the idea that the contrast category had been partially activated, another study was conducted (Gannon, 1991). In this study, the actual exemplars generated for unipolar scales by
Figure 12.2. Ease of Exemplar Generation for Seven-Interval Ratings Scales as a Function of Rating Scale Format.

Figure 12.3 shows the pattern of results. Exemplars that were originally generated for the implicit end of the rating scale were less extreme than those generated for the explicit end. In some cases (see scale labels three, four, and five in the figure), the generated exemplars were opposite in valence. For example, an exemplar generated for the unipolar label “somewhat honest” may have been rated under the bipolar label “somewhat dishonest.” These results provide support for the notion that the unipolar scale format partially activates the contrast category and fully activates the explicit category.

A follow-up study investigating this pattern of activation is worth noting. The label used at the implicit end of the unipolar scale for the preceding study was “were not at all ______.” It is plausible that this verbal label strongly suggested the use of a contrast category. That is, the complete absence of one category can be most readily understood as the presence of the semantically opposite category. This assumption can be supported by Holyoak and Glass (1975) and Ebbesen and Allen (1980) who found that highly related instances that are clearly contradictory are rejected more quickly in instance verification tasks than other nonmember instances.

Consider the use of an alternative endpoint label “very slightly ______.” If only subtly, this label suggests the presence of the explicit category. If the “not at all ______” label activates the contrast category as suggested, the ease ratings should be higher when this label is used than when the alternative label of “very slightly ______” is used.

A study using the exemplar generation methodology was conducted in which the endpoint label was experimentally varied (Gannon and Ostrom, 1992). Half the subjects generated exemplars for a rating scale that had the implicit endpoint label of “not at all ______” and half generated exemplars for a rating scale that had the implicit endpoint label of “very slightly ______.”

The results are shown in Figure 12.4. The ease ratings for the
end labels were significantly higher when the "not at all ____" label was used compared with when the "very slightly ____" label was used. These results are consistent with the categorical interpretation of the exemplar generation.

This study has an important implication for rating scale methodology. There are occasions when researchers wish to obtain separate empirical measures for both a category and its opposite. For example, it would be useful in surveying the public's perceptions about politicians to know both how trustworthy and how untrustworthy they are perceived to be. In such a case, the label "not at all ____" should be avoided and the label "very slightly ____" preferred.

Assessing the Exemplar that Comes to Mind

We argued earlier that rating scales influence the way in which responses are given. Asking how honest someone is is different from asking how dishonest someone is. The categorical model posits that the endpoint labels activate the cognitive categories associated with them. These labels are associated with highly accessible exemplars or units of knowledge. What first comes to mind when encountering a rating scale is thus likely to be these highly accessible exemplars or units of knowledge.

Responding to a series of survey questions may well motivate participants to respond quickly. This is especially true in the case of
rating scales where respondents need only select one interval from a limited number and are not obligated to provide a rationale for selection. In such situations, respondents are cognitively busy and may be unable to use all the information at their disposal (Gilbert, Pelham, and Krull, 1988). Information that is most accessible may be more likely to be used. Furthermore, even if not cognitively busy respondents may be cognitive misers conserving cognitive energy unless motivated otherwise and thus they may simplify survey answering (Fiske and Taylor, 1991; Krosmick, 1991). One way in which they simplify the task is to make use of the most accessible information.

According to Schwarz and Bless (1992), survey responding involves retrieving a standard of comparison as well as a representation of the stimulus. It is plausible that the categorical representation that is highly accessible upon seeing the rating scale may be used as a standard of comparison in making a judgment.

The categorical model approach may be applied to rating scales to help understand context effects in survey responding. Specifically, understanding whether assimilation or contrast of a judgment to the context will occur involves understanding which categorical representation is activated by the context, according to Schwarz and Bless (1992). The best-fit or ideal exemplars that are highly accessible when the category is activated by a particular rating scale may differ depending on the context. Research by Roth and Shoben (1983) demonstrates that the best-fit exemplar listed for the category bird can be influenced by the context. In a farm context, the chicken is often listed as the prototypical bird. In the context of Australia, the ostrich is often listed. Thus, the most prototypical exemplar of honesty in the context of politics (for example, admitting responsibility for a poor decision) may be different from the prototypical exemplar of honesty in the context of religion (for example, confessing to a mortal sin).

Furthermore, the category activated by the rating scale may influence the construction of the representation of the stimulus during the responding task. Strack and Martin (1987) argued that individuals do not retrieve all information relevant to a judgment and that their responses are based heavily on what is most accessible. Thus, the category activated by the rating scale may direct or at least influence the search for information about the target because it is what is most accessible. For example, when encountering a unipolar scale of dishonesty, one may search for anything dishonest the target has done rather than search for an honest act.

To address more directly the question of which exemplar comes to mind first when encountering a rating scale, several studies have been conducted using a version of the exemplar generation methodology (Gannon, 1993). This version allowed for an examination of the exemplar produced first by the respondent. Previous research on categories suggested that best-fit exemplars come to mind first when respondents are asked to generate categorical instances (Battig and Montague, 1969; Rosch, 1973; Barsalou 1985). Based on the categorical model, one would expect then that the first exemplars generated would correspond to the endpoint scale labels.

In the present study, rating scale labels were typed on individual index cards instead of being presented in booklet form. The cards were placed facedown on a table in front of the subject. The order of the cards was counterbalanced across subjects. Respondents received instructions similar to those used in the other exemplar generation studies described in this chapter. The key difference was that in this study subjects were also instructed to generate the exemplars in any order they wished. When finished with a label, they were to place it facedown in a pile. The dependent variable in this study was the order in which respondents generated exemplars, which was easily measured by noting the order in which the cards had been placed face down in the pile. The mean order of exemplar generation for each of the scale labels is shown in Figure 12.5.

To facilitate comparisons with other exemplar generation results reported in this chapter, we have reversed the rank order. Thus, if a label was generated first we gave it a rank of seven.
The by-now familiar U-shape curve is evident in the graph. The endpoint labels tended to be selected for exemplar generation earlier in the sequence than the intermediate scale labels. On the average, little difference appeared between the implicit and explicit ends of the unipolar scale.

Under the time pressure of a survey, respondents may only spontaneously generate exemplars for one or two scale labels. Another way of looking at these data is to examine the proportion of subjects who generated exemplars for each of the particular scale labels first. As Figure 12.6 shows, the endpoint labels were selected first by a higher proportion of respondents. The implicit scale label was generated first more often and the explicit label second most often.

In the case rating studies, the explicit end is easier to generate an exemplar for than the implicit end. The pattern of results in this order of generation study may not purely serve as a measure of accessibility, in that the pattern may partially reflect a strategy of generation. Specifically, respondents’ strategy may have been to start with the contrast category because in some ways it is more poorly defined and ambiguous than the designated category. This allows them first to set the anchors (or boundaries) and then proceed to the easier task of generating exemplars for the designated category. The designated endpoint exemplar still may be the most accessible in this case.

A second explanation is that it is more than just a strategic decision regarding which label to write an exemplar for. It is possible that concepts are given meaning by their opposite and that the
opposite category acquires cognitive dominance. Whereas the non-designated category may be more difficult for exemplar generation than the designated category, it may ultimately become more memorable and thus more influential in guiding the search for the personal experiences and beliefs that the respondent draws upon in answering the question.

To sort out these alternative explanations, a surprise recall task was given after respondents had generated the exemplars and completed a ten-minute memory distraction task. The recall task asked respondents to write down as many of their exemplars as they could in the order in which they came to mind.

Examination of the exemplars recalled first was also anticipated to be informative in understanding survey responding. As mentioned previously, Strack and Martin (1987) have found that previous questions can affect the interpretation of later questions in a survey. If exemplars are used as standards of comparison as argued in the previous section, then an interesting question is which exemplar does an individual remember from a previous encounter with a rating scale?

In the example, respondents tended first to recall exemplars that they had generated for the explicit endpoint label (see Figure 12.7). This pattern differs from the order in which the exemplars were generated (that is, the implicit end label was generated first by a higher proportion of subjects).

The present recall findings mirror the ease rating data pattern of the previous studies. These data support the interpretation that exemplars for the non-designated end of the scale were generated first because of strategic, defining purposes. The data argue against the suggestion that the implicit cognitive category acquires a dominance in defining and representation of the scale.

The results of this study suggest that the exemplars generated for the explicit endpoint label tend to be the most accessible exemplars in a recall task. This is consistent with our interpretation of the ease ratings in the earlier studies, that is, that they reflect cognitive accessibility. Although subjects tended to generate exemplars for the implicit endpoint label first, they were most likely to generate the exemplars for the explicit endpoint label next and most likely to recall the explicit endpoint label first.

We have argued that the rating scale activates a category and along with it associated exemplars. These exemplars are organized in a graded structure from the most highly representative to the least. Once the highly accessible instance are generated, which scale labels are generated next? If the most highly accessible instance is generated first, perhaps the graded structure of the category dictates which exemplar is generated next.

A calculation was made of the conditional probabilities of generating an exemplar for a particular label given that one has just previously generated an exemplar for a different scale label. The
conditional probabilities displayed in Table 12.2 identify two sequential stages. The first stage of generation is to start with the extreme labels. This is consistent with the findings shown in Figures 12.5 and 12.6. The second stage involves a choice between two patterns of activation. Respondents moved either to the adjacent label at the original end of the scale or to the labels at the other extreme. For example, if the first scale label selected was "extremely honest" the respondent selected either the adjacent "very honest" or the opposite label "not at all honest."

Examining rows 0 and 6 in Table 12.2 shows that 35 percent and 23 percent of the respondents, respectively, selected a scale label adjacent to the extreme label initially selected. This gradient-oriented activation pattern appeared to continue in successive generation responses until all labels to the midpoint were completed. The table also shows strong evidence for an activation pattern that involves providing an exemplar base for the two extremes of the scale. Rows 0 and 6 show that 27 percent selected the opposite scale label second. For each of the two rows, this represented by a substantial margin the first or second most probable conditional response.

Implications for Item Construction

Using the exemplar generation methodology in the context of the categorical model can be an informative tool for the survey researcher. The methodology may be used in many facets of item construction.

First, the model provides a conceptual basis for label selection and the methodology can be used to assess the selection. For example, the methodology can be used to detect the presence of the implicit, nondesignated category. Rather than relying on his or her intuition about what constitutes a contrast category, the researcher can make an empirical investigation of it. For example, Solomon (1978) argued that scales assessing situational and dispositional attributions need to be measured on two unipolar scales. He argued that the two concepts do not vary inversely. Indeed, in one of our studies discussed briefly at the beginning of the chapter, we found that exemplars generated for the implicit endpoint label for an attribution scale using situational and dispositional labels ("due very much to John," "due very much to the situation") were not placed in the opposite category when rated on a bipolar scale, but rather in the midpoint category ("due neither to John nor to the situation"). This is in contrast to other content areas in which the exemplars generated for the implicit endpoint label were placed in the other bipolar category.

Second, the exemplar generation methodology proved useful for rating scale format evaluation. For example, the results of the studies using this methodology have shown that unipolar scale formats provide more category relevant discrimination in responding than the bipolar scale formats. That is, for the unipolar scale formats, two thirds of the scale corresponds to the designated category and one third to the contrast category (see Figure 12.3), whereas for the bipolar scale formats half of the scale corresponds to each of the categories activated by the endpoint labels. The researcher can ask how much categorical discrimination is desirable. Thus, the model additionally provides a conceptual basis for determining the number of intervals to be used.
Third, the methodology can be used as a conceptual basis for survey item critique and revision. For example, by examining the content of the exemplars the respondent is generating and the ease with which they are generated, ambiguities in language may be detected. It may also be possible to detect unintended interpretations of a survey question or a rating scale (for example, respondents interpret “home health care” to mean everyday health care given by family members, such as providing aspirins, instead of health care workers coming into the home to provide health care services, such as changing surgical dressings).

Fourth, the methodology may be used to track the meaning of item responses over the years. For example, this methodology could help assess whether a change in ratings over the years reflects an actual change in beliefs or perceptions or is the result of a change in the content definition of the category. For example, a new meaning may have become generally accepted for the concept “successful” or “conservative.” If ratings have changed, one may also look to see if the exemplar basis of the rating scale itself has been changed. Our model suggests that a change in the language used for the endpoint labels can lead to the activation of different cognitive categories.

Finally, the methodology can be used to investigate basic research questions. Most of the studies reported in this chapter address the conceptual question of how respondents make sense of the survey questions in which a multiple interval rating scale response format is used.

We have discussed how this methodology and variations of it have been used to examine a categorical model of rating scales responses. The model provided an explanation for several basic research questions: What information is most assessable upon encountering a rating scale? Do rating scales activate cognitive categories? Can a rating scale activate more than one category? How can we determine which categories are activated? Is there a typology of rating scales based on patterns of categorical activation?

The use of reaction time to study mental processes has had a long and sometimes controversial history in experimental psychology. More than a century ago F. C. Donders (1868), a Dutch physiologist, explored the speed of mental processes on the assumption that thinking consists of a chain of successive processes that build on each other in a cumulative manner. Donders’s interest and that of many experimental psychologists after him was to gain insight into the thought processes that underlie mental tasks by measuring the duration of their constituent stages.

Early research using reaction time was handicapped by methodological and conceptual weaknesses that eventually brought the approach into disrepute (Johnson, 1955; Woodworth, 1938). The problems raised by early critics of the approach, however, were substantially resolved with the advent of modern methods of experimentation and data analyses (Sternberg, 1969). Also, the cognitive renaissance that began in the 1960s replaced a long-standing behavioristic aversion to mentalism with a keen interest in the processes that are responsible for perception, thinking, and memory (Neisser, 1967). Reaction time provided an important metric for mental processes and was embraced by a new generation of skilled researchers (for example, Collins and Quillian, 1972; Sternberg, 1969). Following the publication of a large number of influential
Norbert Schwarz and
Seymour Sudman, Editors

Answering Questions

Methodology for Determining
Cognitive and Communicative
Processes in Survey Research

1976

Jossey-Bass Publishers
San Francisco