PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES

Looking Back at Anger: Reference Periods Change the Interpretation of Emotion Frequency Questions

Piotr Winkielman University of Michigan Bärbel Knäuper Freie Universität Berlin

Norbert Schwarz University of Michigan

When asked to report on behaviors and experiences, participants draw on the specified reference period to infer the question meaning: Short reference periods suggest that the question pertains to frequent experiences; long reference periods suggest that it pertains to rare ones. Because frequent experiences are typically less intense than rare ones, this meaning shift results in reports of different experiences. Three experiments support this analysis in the domain of emotion reports. Participants asked how frequently they get angry (a) assumed that the question refers to less intense and more frequent episodes when presented with a short (1-week) rather than a long (1-year) reference period, (b) reported more extreme episodes in the latter case, and (c) provided differential frequency reports. These differences reflect conversational inference processes and cannot be fully accounted for by memory search biases.

Research participants are often asked to report on their thoughts, behaviors, emotions, and experiences during a specified reference period. Depending on the purpose of the study, the length of the reference period may range from short time spans, such as 1 hr, to more extended time spans, such as a day, a week, a month, or even a year. In choosing a reference period, researchers often take the assumed frequency of the target phenomenon into account (see Sudman & Bradburn, 1983, for a discussion). When the behavior or experience is likely to be relatively infrequent, researchers prefer an extended reference period, or else too many participants may report a nonoccurrence, thus limiting the possible analyses. Conversely, when the behavior or experience is likely to be relatively frequent, researchers prefer shorter reference periods, based on the concern that participants may have forgotten more distant instances of a frequent behavior or experience. In fact, frequent experiences tend to blend into generic, knowledge-like representations of the type of behavior or experience, which lack the time and space markers that allow for the episodic recall of distant events (see Schwarz, 1990; Sudman, Bradburn, & Schwarz, 1996, for reviews).

What is often overlooked, however, is that the chosen reference period may influence which behaviors or experiences participants consider to be the referent of the question. Suppose, for example, that participants are asked how often they were angry during a specified reference period. To provide a meaningful answer to the question, participants, as cooperative communicators, have to determine the intended meaning of the term angry (Grice, 1991, Sperber & Wilson, 1995). Does the researcher refer to minor irritations, such as waiting in a long line, or to major annoyances, such as fighting with one's spouse? To infer the intended meaning of ambiguous questions, individuals are likely to draw on the context of the utterance-both in daily life and in research situations. In research situations, this context includes apparently formal features of the research instrument, such as the length of the reference period or the nature of the response alternatives (for reviews, see Clark & Schober, 1992; Hilton, 1995; Schwarz, 1994, 1996). Suppose, for example, that the question reads, "How often were you angry yesterday?" In this case, participants may infer that the researcher is interested in relatively frequent anger experiences—or else focusing on a single day would make little sense. Moreover, the wording of the question presupposes that the participant has experienced some anger yesterday, and may, in fact, have experienced more than one anger episode, or else the question would be better worded "Were you angry yesterday?" (see Knäuper, 1998).

Piotr Winkielman, Department of Psychology, University of Michigan; Bärbel Knäuper, Department of Psychology, Freie Universität Berlin, Berlin, Germany; Norbert Schwarz, Department of Psychology, University of Michigan.

The reported research was supported by a Max Planck Society Fellowship awarded to Bärbel Knäuper. We thank Gwen Dewar, Kent Harber, Tim Ketelaar, Uli Schimmack, Dan Sperber, Diederick Stapel, and Bob Wyer for their helpful comments. We are also grateful to Roberta Mancuso, Magdalena Ponurska, and our undergraduate assistants for their help with data collection and coding.

Correspondence concerning this article should be addressed to Piotr Winkielman, who is now at the Department of Psychology, University of Denver, 2155 South Race Street, Denver, Colorado 80208. Electronic mail may be sent to pwinkiel@du.edu.

Journal of Personality and Social Psychology, 1998, Vol. 75, No. 3, 719-728 Copyright 1998 by the American Psychological Association, Inc. 0022-3514/98/\$3.00

Conversely, suppose that participants are asked, "How often were you angry *last year*?" In this case, they may infer that the researcher is interested in relatively rare anger experiences. After all, the researcher can hardly expect them to remember all instances of a frequent experience for the lengthy reference period of 1 year. Moreover, why would one present such an extended recall task unless one is interested in rare events, which cannot be captured by the easier task covering a short reference period? Hence, participants may infer from a long reference period that the researcher is interested in relatively rare instances but may infer from a short reference period that the researcher is interested in relatively frequent instances.

It is important to note, however, that frequent experiences are likely to be less intense than infrequent ones: Whereas major annoyances usually do not occur on a daily basis, minor annoyances are a routine part of daily life. Because we are unlikely to remember the "small stuff" for a long time, this link further increases the likelihood that participants will infer that the above anger question can't plausibly refer to minor annoyances when it pertains to a reference period of 1 year. A participant may say to himself or herself, "The researcher would not expect me to remember all the small experiences that happen over the whole year, so she must be asking about the serious ones." This reasoning suggests that the same question may be interpreted as referring to substantively different experiences, depending on the length of the reference period. If so, participants asked about different reference periods would deliberately report on different experiences, undermining the comparability of the obtained data. This possibility has important implications for the assessment and interpretation of self-reports. We address these implications below, using discrepancies between concurrent and retrospective reports of emotion intensity as an example.

Consistent with our conceptual analysis, Schwarz, Strack, Müller, and Chassein (1988) observed that formal characteristics of a question can profoundly affect participants' interpretation. Specifically, these researchers asked participants to report how often they felt "really annoyed" on one of two different scales: a low-frequency scale that ranged from less than once a year to more than once every 3 months, or a high-frequency scale that ranged from less than twice a week to several times a day. Following the frequency report, participants described a typical example of an annoying situation. Participants who were exposed to the low-frequency scale provided more extreme examples than participants who were exposed to the high-frequency scale. Presumably, participants in the low-frequency scale condition inferred that the researchers were interested in relatively rare, and hence more extreme, events, whereas the participants in the high-frequency scale condition inferred that the researchers were interested in relatively frequent, and hence less extreme, events. Although these findings are compatible with the present conversational analysis, they are also compatible with an alternative account offered by Thomas and Diener (1990) to explain discrepancies between concurrent and retrospective reports of emotion intensity.

Concurrent and Retrospective Reports of Emotion Intensity: Memory Bias or Question Interpretation?

In two studies concerned with memory accuracy in the recall of emotion intensity, Thomas and Diener (1990) asked participants to provide concurrent or retrospective ratings of various positive and negative emotions. The concurrent ratings were obtained by asking participants to make their ratings four times a day (Study 1) or at the end of the day (Study 2), thus introducing reference periods of a few hours (Study 1) or 1 day (Study 2). The retrospective ratings were obtained after the completion of a 3-week period (Study 1) or a 6-week period (Study 2). As may be expected on the basis of the above analysis, the retrospective (long reference period) ratings of the intensity of positive and negative affect were significantly higher than the concurrent (short reference period) ratings. Similar findings were reported by Parkinson, Briner, Reynolds, and Totterdell (1995).

Findings of this type are consistent with our analysis of the role of reference periods in question interpretation, but they are also consistent with a memory bias account proposed by Thomas and Diener (1990). These authors suggested that "very emotional times in people's lives are more salient to them than the more neutral occasions" (p. 295). Accordingly, past intense events are more likely to come to mind than less intense ones when participants are asked to provide a retrospective report, resulting in the observed discrepancy between concurrent and retrospective reports of emotion intensity. Note that this memory bias interpretation can also account for the findings reported by Schwarz et al. (1988), reviewed above. Specifically, their lowfrequency scale covered a longer time span than their highfrequency scale, thus increasing the likelihood that an intense event had occurred during that time span and would come to mind.

Current Research

The current studies provide a systematic test of the conversational analysis outlined above, explore its implications for selfreports of emotion, and contrast its predictions with predictions derived from a memory bias hypothesis. Specifically, we tested (a) whether reference periods of different length influence participants' interpretation of emotion terms and (b) whether these changes in interpretation result in corresponding differences in participants' reports of their own emotions. In addition, we (c) tested the implications of our analysis for the use of vague quantifiers, such as *hardly ever* or *very frequently*, and (d) explored the impact of these processes on subsequent comparative judgments as addressed below.

Reference Period and Question Interpretation: Reports of Frequency and Intensity

Study 1 demonstrates that the term *anger* is interpreted as referring to more extreme and less frequent experiences when the question pertains to a "typical year" rather than a "typical week." Studies 2 and 3 were designed to show that differences in the reports of anger across reference period conditions were due to changes in question interpretation rather than to differences in the likelihood that intense events are encountered when searching time periods of different lengths. Specifically, in Study 2, participants were asked to rate the frequency of anger and subsequently reported relevant examples of anger episodes. For some participants, the meaning of *anger* was left undefined, thus requiring them to infer the intended meaning of the term

from the reference period specified in the question. For other participants, a definition of the intended meaning of *anger* was provided. If reference-period-induced changes in anger reports are due to changes in question meaning, providing a definition should reduce the impact of the time frame specified in the question. If reference-period-induced changes are due to changes in the time period searched per se, providing a definition should have no impact. The data supported the conversational prediction as described below.

Finally, participants in Study 3 were asked to report the absolute frequency of anger experiences during the "last week" and the "last year" in counterbalanced order. We predicted that the interpretation of the term *anger* suggested by the first question would carry over to the second question, regardless of the time period searched to answer the second question. This should result in increased reports of anger episodes for the 1-year period when the first question pertained to "last week," and, conversely, in reduced reports of anger episodes for the 1-week period when the first question pertained to "last year." Again, the data supported this prediction, as described below.

Vague Quantifiers

How participants interpret the emotion term presented in a question not only bears on which experiences they will attempt to retrieve but should also influence which frame of reference they consider appropriate in assessing these experiences. In Study 3, participants were asked to report the frequency of anger episodes in numeric terms. Under this condition, the only term that requires disambiguation is anger. In Studies 1 and 2, however, participants were asked to report their anger frequency along rating scales anchored by the vague quantifiers hardly ever and very frequently. These anchors are themselves ambiguous: How often is "very frequently?" As a large literature on the interpretation of vague quantifiers has demonstrated, the numerical meaning of terms like very frequently is partially determined by the context (for comprehensive reviews, see Moxey & Sanford, 1992; Pepper, 1981). For example, the term frequent refers to a higher numerical value when it pertains to "frequent shootings in Hollywood Westerns" rather than to "frequent earthquakes in California" (Pepper & Prytulak, 1974). Such shifts in the meaning of vague quantifiers as a function of context can be understood in terms of norm theory (Kahneman & Miller, 1986). A different norm for the rate of occurrence pertains to the class "shootings in Westerns" than to the class "earthquakes in California."

If so, different reference periods may influence not only which emotional experiences participants consider relevant but also how they interpret the quantifiers *very frequently* and *hardly ever*. When participants infer from a long reference period that the question pertains to rare and intense anger, they should use the term *very frequently* to describe lower absolute frequencies than when they infer from a short reference period that the question pertains to frequent and minor anger. This possibility has differential implications for Studies 1 and 2.

Recall that in Study 1, participants were simply asked to report how often they felt angry during the specified reference period (1 week vs. 1 year). If their interpretation of *very frequently* is based on their differential interpretation of the intended meaning of *anger*, they may provide similar frequency ratings on a vague quantifier scale under both reference period conditions. Such similar ratings on a scale may be obtained despite the fact that any given experience is likely to have occurred more frequently during 1 year than during 1 week. Such a finding would indicate that the reference period influences both the interpretation of the emotion term and, indirectly, the interpretation of vague quantifiers used as part of the response scale.

It is important to note that this should not be the case for those participants in Study 2 who are provided with an explicit definition of the intended meaning of *anger*. As discussed above, the explicit definition eliminates the need to draw on the reference period to disambiguate question meaning. Under this condition, participants should provide higher frequency ratings when the question pertains to "last year" rather than "last week." This is because the explicit definition of the term *anger* undermines the indirect impact of the reference period on the interpretation of *very frequently* and evokes reliance on the same norm for both time frames (Kahneman & Miller, 1986). In combination, this pattern of findings would inform us about the inferential processes involved in answering questions on scales anchored by vague quantifiers when the target experience is defined and when the target experience is not defined.

Comparative Judgments

So far, we have addressed the impact of different reference periods on the interpretation of emotion terms and subsequent reports of the frequency and intensity of emotions provided under different response formats. As a final consideration, we turn to relative judgments of the type "Compared to most other people, how often do you get angry?" In previous research, Schwarz and colleagues asked participants to report the frequency of their behavior using scales with several numeric response alternatives, ranging, for example, from once a week to more than twice a day (e.g., Menon, Raghubir, & Schwarz, 1995; Schwarz, Hippler, Deutsch, & Strack, 1985). They observed that participants assumed that values in the middle range of the scale reflect the "average" or "usual" frequency in the population, whereas values at the extremes of the scale reflect the extremes of the distribution. On the basis of this tacit assumption, participants drew on their own location on the scale to infer their relative standing in the distribution. Hence, they inferred, for example, that twice a week is below average when the scale ranged from once a week to several times a day but above average when the scale ranged from once every other month to more than once a week. As a result, the comparison information provided by the numeric response alternatives has been found to influence a wide range of comparative judgments (see Schwarz, 1990, 1996, for reviews).

Study 2 extends this research from scales with numeric response alternatives to rating scales with vague quantifiers. More important, it investigates the consequences of defining the target behavior and thus evoking reliance on the same norm across time frame conditions (Kahneman & Miller, 1986). As discussed above, participants who are provided with an explicit definition of *anger* and are asked about a long reference period should check a value at the higher end of the frequency rating scale. As a result, these participants may infer from their placement on the scale that they are angry more often than others. If obtained, this pattern of findings would suggest that the inference processes discussed above bear not only on immediate reports of emotion frequency and intensity but also on participants' comparative judgments of their own emotionality.

Study 1

Study 1 examined whether participants' interpretation of emotion terms depends on the length of the reference period specified in the question. Participants first reported the frequency with which they get angry during a "typical week" or during a "typical year" on a scale anchored with the vague quantifiers *hardly ever* and *very frequently*. Subsequently, they were asked to indicate how serious, frequent, and intense the experiences are that the researchers had in mind when they worded the question. We predicted that participants would infer that the researchers are interested in experiences that are less serious, more frequent, and less intense when the reference period pertained to a week rather than a year.

Moreover, we predicted that these differences in interpretation would be obtained in the absence of any differences in the frequency reports provided along the rating scales anchored by vague quantifiers. This prediction reflects that events are judged against the norm that is relevant to the particular class of events (Kahneman & Miller, 1986). Hence, different reference periods should influence not only the extremity of the experiences participants consider relevant but also the frame of reference they employ in determining what is "frequent" given this class of experiences.

Method

One hundred eleven undergraduates at the University of Michigan answered a two-page questionnaire titled "Interpretation of Questions." The reference period was manipulated between subjects, and the respective question read, "How often do you get angry during a typical week (year)?" Participants made their responses on a 9-point rating scale anchored by the vague quantifiers hardly ever (1) and very frequently (9). The instructions for the next part of the questionnaire informed participants that the researchers "are interested in how you interpret ambiguous questions." Participants were asked to indicate what type of experiences the researcher wanted them to consider when answering the frequency question they had just completed. They rated the experiences along three dimensions: (a) the intensity of feelings of anger from very weak (minor irritation) to very strong (rage), (b) the seriousness of the event causing the anger from not serious (trivial event) to very serious (major life event), and (c) the frequency of anger-eliciting situations from very rare (once a year or less) to very frequent (daily or more often). Each rating was provided on a 10-point scale, with higher values indicating higher extremity.

Results

The mean responses to the anger frequency question and the interpretation questions are shown in Table 1. Also included in this table are standard deviations and effect size statistics (Rosenthal & Rosnow, 1991). As expected, the responses to the interpretation questions indicated that participants drew on the length of the reference period in inferring what the researchers wanted the question to refer to. A series of two-tailed t tests revealed that compared with participants in the week condition, participants in the year condition inferred that the researchers

Table 1

Evaluation of	'Intended	Meaning	and	Frequency	of Anger
as a Function	of Time	Frame			

	Ti				
	We	ek	Year		
Variable	М	SD	М	SD	size (r)
Intended meaning					
Strength	5.74	1. 94	6.37,	1.99	.16
Seriousness	5.32	1.79	6.37 _b	1.77	.29
Frequency	6.65,	1.64	5.33 _b	2.04	.34
Frequency of reported			-		
anger	3.77 _a	1.98	4.26 _a	2.03	.12

Note. Means in the same row that do not share subscripts differ at p < .05.

were interested in stronger feelings of anger (M = 6.37 vs. M = 5.74), t(109) = 1.70, p = .09; more serious events (M = 6.37 vs. M = 5.32), t(109) = 3.11, p < .005; and less frequent events (M = 5.33 vs. M = 6.65), t(109) = 3.76, p < .001, respectively.

Finally, the responses to the anger frequency question did not differ significantly across the reference period conditions (t = 1.28, ns). This is consistent with the assumption that the vague quantifiers hardly ever and very frequently were interpreted in the context of the inferred meaning of the term anger.

Discussion

The results of Study 1 indicate that the length of the reference period specified in the question influences the interpretation of the emotion term as suggested by a conversational analysis. When a frequency question pertained to a long reference period, participants inferred that the researcher was interested in experiences that are less frequent and more serious than when the same question pertained to a short reference period. Moreover, these differences in interpretation were obtained in the absence of significant differences on the frequency reports provided on a rating scale with vague quantifiers. The latter finding reflects that the time frame not only influences the interpretation of the emotion term but also determines the frame of reference used in mapping the frequency of the experience onto a set of vague quantifiers. We return to this issue in Study 2.

Study 2

In Study 1, participants were explicitly asked to infer the meaning of the emotion term that the researchers had in mind. This design leaves open to what extent participants would spontaneously arrive at these differential interpretations. If the predicted effects can be obtained even when participants are not prompted to interpret the meaning of the question, this would strengthen our assumption that inferences about the different extremity and frequency of events elicited by reference period information are a natural part of the question answering processes.

To address this issue, we again asked participants in Study 2 to report how frequently they were angry during a typical week versus a typical year and to subsequently report an example of the anger experiences they had. Their examples presumably represent the type of experiences they had in mind when answering the preceding frequency question. We expected that participants would provide examples of more intense anger when the frequency question referred to a year rather than a week.

In addition, Study 2 was designed to provide additional evidence bearing on the role of conversational processes and memory bias. According to the conversational interpretation, participants interpret the question as referring to more intense and less frequent episodes when it specifies a long rather than a short reference period. According to the memory bias explanation, rare and intense events are more salient than frequent and less intense ones and hence are more likely to come to mind. Yet, this can only be observed when the time span searched is long enough to allow for their occurrence. Hence, intense events are more likely to be reported when the question pertains to a longer reference period. To examine these explanations, we manipulated the ambiguity of the term anger. For some participants the meaning of the term anger was left undefined, whereas other participants were provided with a definition. If interpretative processes are crucial, the reference period should only affect participants' interpretation when the meaning of the term anger remains unspecified. In contrast, if memory search processes are crucial, the time frame of the question should influence which examples participants report regardless of the ambiguity of the emotion term.

Note also that the presence or absence of a definition of the emotion term should influence whether participants in the 1week versus 1-year conditions would report differential frequencies of anger on a scale with vague quantifiers. When no definition is provided, the frequency is evaluated relative to the norm pertaining to the class of events suggested by the length of the reference period (Kahneman & Miller, 1986; Pepper, 1981). In this case, the respective frequencies should be mapped onto the hardly ever to very frequently scale with reference to different, context-specific norms, resulting in similar, moderate reports under both conditions (as observed in Study 1). In contrast, an explicit definition of the emotion term evokes reliance on the same norm in both time frame conditions. In this case, participants should report higher frequencies when the question pertains to a year rather than a week, reflecting that any experience is likely to have occurred more frequently during 1 year than during 1 week.

If the reference period and definition manipulations have the predicted effects on frequency judgments, we may further expect these manipulations to influence subsequent comparative assessments of one's own emotionality, as discussed above. To explore this possibility, we asked participants, "Compared to other people, how often do you get angry?" To answer this question, participants may draw on their placement on the frequency rating scale, inferring that they are angry more often than others when they endorse a high value (cf. Schwarz et al., 1985). This should be particularly likely when the term *anger* is defined and the question pertains to 1 year, thus eliciting reports of frequent anger. Hence, we predicted that participants would infer a higher comparative frequency of anger when the question refers to 1 year rather than 1 week but that this effect would only be obtained when the meaning of *anger* is defined, paralleling the

above predictions for frequency reports on rating scales with vague quantifiers.

Method

Ninety-seven undergraduates at the University of Michigan responded to a one-page questionnaire. As in Study 1, the reference period was specified in the question: "How often do you get angry during a typical week (year)?" The 9-point response scale was anchored by the vague quantifiers *hardly ever* and *very frequently*.

In the no-definition condition, no additional information was provided, thus replicating Study 1. In the definition condition, a short paragraph served to clarify the meaning of *anger*. Specifically, participants were told to keep in mind that *anger* refers to a wide range of feelings and situations and were given examples of minor and major angerprovoking situations (e.g., waiting in a long checkout line vs. losing important documents).

After answering the anger frequency question, participants were asked to describe an example of a situation that made them angry. Following this report, they were asked, "Compared to most other people, how often do you get angry?" $(1 = much \ less \ than \ most \ others, 9 = much \ more \ than \ most \ others)$.

Results

Reported examples. Participants' examples of anger situations were evaluated by three independent raters ($\alpha = .77$). The raters used a 5-point scale (1 = somewhat annoying, 5 = extremely annoying). The means, standard deviations, and effect size statistics are shown in the top panel of Table 2. A 2 (reference period) \times 2 (definition) analysis of variance (ANOVA) performed on the combined ratings of the examples revealed the predicted reference period by definition interaction, F(1, 94) = 4.01, p < .05. Without a definition of the term *anger*, participants reported examples of less severe episodes when the preceding frequency question pertained to 1 week (M = 3.11) rather than to 1 year (M = 3.81), t(47) = 2.77, p < .01. In contrast, no difference between the week and year frames was obtained when the term *anger* was defined independently of question wording (t < 1).

Table 2

Severity of Reported Event, Frequency of Anger, and Comparative Judgment as a Function of Time Frame and Definition

	Time frame condition				
	Week		Year		
Variable	М	SD	М	SD	Effect size (r)
Severity of reported event					
Undefined	3.11,	0.89	3.81 _h	0.88	.37
Defined	3.28	0.77	3.25	0.95	01
Frequency judgment	-		-		
Undefined	3.80,	2.08	4.36,	1.70	.15
Defined	3.41	1.40	5.52 _b	1.69	.57
Comparative judgment	-		-		
Undefined	4.04	1.97	3.92,	1.78	03
Defined	3.50°	1.34	5.00 _b	2.00	.41

Note. Means in the same row that do not share subscripts differ at p < .05.

We also evaluated the possibility that the examples provided as part of the definition increased the accessibility of similar events in memory (Higgins, 1996), resulting in differential reports for this reason. Our data provide no support for this possibility. Only 7 participants listed examples that referred in some way to "waiting" or "losing things." Moreover, the number of participants who listed such examples did not differ across the definition and no-definition conditions. Finally, when the data were analyzed without these participants, all critical effects were replicated.

Frequency reports. The means, standard deviations, and effect size statistics of participants' frequency reports are shown in the middle panel of Table 2. A 2 (reference period) \times 2 (definition) ANOVA revealed a main effect of reference period, F(1, 96) = 13.63, p < .001. This main effect was qualified by the predicted reference period by definition interaction, F(1, 96) = 4.77, p < .05. When the term *anger* was not defined, participants reported similar frequencies (t = 1.05, ns), replicating Study 1. When the term was defined, participants reported feeling angry more often when the question pertained to 1 year rather than to 1 week, t(45) = 4.63, p < .001, consistent with our predictions.

Comparative judgments. Finally, a 2 (reference period) \times 2 (definition) ANOVA was performed on participants' comparative judgments, shown in the bottom panel of Table 2. This analysis again revealed the predicted reference period by definition interaction, F(1, 96) = 4.88, p < .05. When the term *anger* was not defined, participants' comparative judgments did not differ (t < 1), paralleling the absence of a reference period effect on their frequency reports. When the term *anger* was defined, participants reported getting angry more often than others when the frequency question pertained to 1 year rather than to 1 week, t(45) = 2.98, p < .01, again confirming our predictions.

Discussion

The results of Study 2 support the predictions derived from the conversational analysis. When the meaning of the term anger was not defined, participants were expected to infer that the question pertains to more severe events when presented with a reference period of 1 year rather than 1 week. The events that they reported as typical examples confirmed this prediction. This difference was eliminated, however, when the meaning of anger was defined independently of the reference period specified in the question. In contrast, a memory bias account would predict that participants would report more extreme events whenever they sample the extended period of 1 year, independent of whether the crucial term is defined or not (provided that the definition allows for a broad range of examples, which was the case). Moreover, the conversational account predicts the additional findings obtained in Study 2, findings on which a memory account is silent.

When the term *anger* was undefined, participants reported similar and moderate frequencies independent of whether the question pertained to 1 week or 1 year. These data presumably reflect that participants drew on differentially extreme experiences, as shown in their examples, but evaluated these experiences relative to different norms (Kahneman & Miller, 1986). In contrast, they reported higher frequencies in the 1-year than in the 1-week condition when *anger* was defined, consistent with the assumption that the definition evoked reliance on the same norm. Finally, they drew on their frequency report when they were asked to compare their own frequency of being angry to the frequency of others, reporting a higher comparative frequency of anger when the 1-year frame elicited a high frequency estimate. This latter finding parallels the previously observed impact of participants' placement on numeric frequency scales on comparative judgments (see Schwarz, 1990, for a review).

In contrast to Study 1, the above findings were obtained without asking participants to explicitly infer the meaning of the question across different time frames, thus indicating the applicability of the current findings to a typical research situation.

Study 3

Studies 1 and 2 demonstrated that the reference period of frequency questions influences the interpretation of emotion terms and hence the type of episodes people consider when providing their reports. Moreover, Study 2 showed that this effect is eliminated when the emotion term is explicitly defined. Study 3 provides an additional test of the meaning shift hypothesis and addresses how the implicit definition conveyed by the reference period of a preceding question affects the interpretation of subsequent questions. Suppose participants are first asked how often they were angry during the last year (week) and are subsequently asked how often they were angry last week (year). We propose that the first question evokes an interpretation of anger in terms of major (minor) annoyances and that this interpretation carries over to the subsequent question, reflecting that the first interpretation becomes part of the cumulative common ground of the research conversation (Clark & Schober, 1992; Schwarz, 1996). If so, participants should report lower weekly frequencies when the 1-week question follows the 1-year question, reflecting that the 1-year question establishes an interpretation of anger in terms of more severe events, which are less frequent. Conversely, participants should report higher annual frequencies when the 1-year question follows the 1-week question, reflecting that the 1-week question establishes an interpretation of anger in terms of less severe events, which are more frequent. This prediction, generated by the conversational inference interpretation, is incompatible with the prediction generated by a memory search interpretation (Thomas & Diener, 1990). The latter account suggests that the reported frequency of anger experiences should vary solely as a function of the length of the reference period.

To test these hypotheses, we asked participants to report the frequency of anger for the last week and the last year in counterbalanced order. In contrast to Studies 1 and 2, their frequency reports were assessed in an open response format and participants were asked to provide a numeric answer.

Method

Ninety-two University of Michigan undergraduates were randomly assigned to two conditions of a self-administered questionnaire. Two participants did not provide complete data, leaving 90 participants for the analyses. All participants were asked to report their frequency of anger for the reference periods of 1 year and 1 week, and the order of both questions was varied. In the week-year condition, the questions read as follows: What would you say, how often have you been angry in the past week? _____ times.

Now thinking about the entire past year, including the past week: How often have you been angry in the past year? ____ times.

In the year-week condition, the questions read as follows:

What would you say, how often have you been angry in the past year? _____ times.

Now thinking about the past week only, how often have you been angry in the past week? $_$ times.

Results

A comparison of participants' answers to the respective first question shows that they reported an average of 4.16 anger experiences in response to the 1-week question and an average of 90.76 anger experiences in response to the 1-year question. Table 3 illustrates that the estimate for a week is 2.4 times higher than the value of 1.74, which would be expected by dividing the yearly estimate by the number of weeks in a year (52). Similarly, the estimate for a year is 2.4 times lower than the value of 216, which would be expected by multiplying the weekly estimate by the number of weeks in a year (52). This between-subjects discrepancy again indicates that the different reference periods elicited different question interpretations.

More important, as shown in Table 3, the interpretation of *anger* elicited by the reference period of the first question carried over to the second question, despite the question's change in time frame. Because of skewed distributions, the analyses were performed on log-transformed frequencies. A series of t tests revealed that participants reported lower anger frequencies for the past week when this question was preceded by the question about the past year ($M_{log} = 0.83$) than when it was not ($M_{log} = 1.24$), t(88) = 2.45, p < .02. Similarly, they reported higher anger frequencies for the past year when this question was preceded by the question was preceded by the question about the past week ($M_{log} = 4.16$) than when it was not ($M_{log} = 3.54$), t(88) = 2.05, p < .04.

Table 3

Mean Log-Transformed Frequencies, Mean Absolute					
Frequencies, and Expected Values for Anger Experiences					
as a Function of Time Frame and Question Order					

	Log- transformed frequency		Mean	•		
Question order	М	SD	absolute frequency	Expected value		
		Past v	week			
Week-year	1.24	0.79	4.16	$4.16 \times 52 = 216^{a}$		
Year-week	0.83	0.79	2.24	$2.24 \times 52 = 116^{a}$		
Effect size (r)	.25					
		Past y	year			
Week-year	4.16	1.33	132.90	$132.9/52 = 2.55^{\circ}$		
Year-week	3.54	1.56	90.76	$90.76/52 = 1.74^{\circ}$		
Effect size (r)	.21					

^a Expected value for 1 year. ^b Expected value for 1 week.

Discussion

Consistent with the previous findings, participants' absolute frequency estimates in response to the first question asked indicate that they inferred different experiences to be the target of the question. Accordingly, their estimates for 1 year were less than half of the estimate one would extrapolate from their weekly estimates, reflecting that the 1-year question addresses less frequent and more severe experiences than the 1-week question. More important, the interpretation elicited by the first question carried over to the interpretation of the second question. Hence, participants provided lower weekly estimates when the preceding 1-year question conveyed an interest in rare and extreme events and provided higher annual estimates when the preceding 1-week question conveyed an interest in frequent and minor events than when the respective question was asked in the first position.

General Discussion

Reference Periods Influence Question Interpretation

Researchers who collect self-reports of emotional experiences typically assume that the understanding of emotion terms is widely shared within a culture. Whereas this holds true for distinctions among different emotions, each emotion term itself covers a wide range of experiences that vary in terms of intensity and frequency. In most studies, researchers refrain from specifying which kind of experience they have in mind, in part because the meaning may not seem problematic and in part because any specification may interfere with participants' subjective definition of their experience. Ironically, however, other features of the question, such as the specified reference period or a set of frequency response alternatives (e.g., Schwarz et al., 1988), may provide information that participants draw on to infer the intended meaning of the question. As cooperative communicators, participants need to determine what the questioner has in mind, or else their answers may not provide the requested information (Grice, 1991; Hilton, 1995; Schwarz, 1994, 1996; Sperber & Wilson, 1995; Strack & Schwarz, 1992). In the case of frequency questions, one important piece of information that individuals may draw on is the length of the reference period. In general, a long reference period conveys that the questioner is interested in events that may happen relatively rarely, whereas a short reference period conveys an interest in more frequent events. In most cases, frequent events are less extreme than rare events. As a result, the same question elicits reports of different events and experiences as the present studies illustrate.

In Study 1, participants inferred that the researcher was interested in less serious and more frequent events when the question pertained to how often they are angry during a typical week rather than during a typical year. Study 2 confirmed these meaning shifts and showed that participants reported less extreme examples of anger when asked about a 1-week rather than a 1-year period. Moreover, this effect was eliminated when the meaning of the term *anger* was explicitly defined. Finally, Study 3 showed that these differences in interpretation affect the absolute number of anger experiences reported. When asked about anger during the past week, participants reported frequencies that would have resulted in an annual estimate of roughly 2.5 times the frequency they reported when asked about the past year. Moreover, the interpretation established by the first question asked carried over to the interpretation of subsequent questions as discussed above.

In combination, these findings are consistent with the conversational inference interpretation offered here and are difficult to reconcile with a memory bias account (Thomas & Diener, 1990). The latter account holds that extreme events are more accessible in memory and are hence more likely to come to mind, provided that the reference period is long enough to allow for their occurrence. Although this is certainly true, this process should be relatively independent of whether the meaning of anger is explicitly defined or not. Yet, in Study 2, reports of more extreme events for longer reference periods were only obtained when the term was undefined. Moreover, the changing reference periods used in Study 3 allowed participants to sample different periods for each question asked. Yet, the interpretation established by the reference period of the first question asked carried over to subsequent questions as one would expect given the cumulative nature of establishing common ground in conversations (Clark, 1985; Clark & Schober, 1992). We hasten to add, however, that the two accounts are not mutually exclusive. In fact, they share the basic logic that extreme events are rare and hence relatively unlikely in any given short reference period but relatively likely in a long reference period. The key difference is that a conversational account emphasizes deliberate inferences about the intended meaning of the question in addition to the sampling of different reference periods explicitly required by the question.

The present findings have a number of obvious methodological implications that bear, in nonobvious ways, on the theoretical conclusions that we may draw from participants' concurrent and retrospective self-reports.

First, emotion questions are inherently ambiguous and require considerable interpretational work, which makes them highly susceptible to contextual influences of the research instrument. This is of less concern when researchers assess on-line ratings of ongoing emotional experiences (e.g., Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993) or ask for appraisals of a preselected episode (e.g., Smith & Ellsworth, 1985) or the evaluation of a specified focal event (e.g., Mesquita & Frijda, 1992). In these cases, the immediate experience or the presented event is the focus of the conversation and likely to be taken as the referent of the question, consistent with the notion that the most obvious referent is assumed to be the intended one (Sperber & Wilson, 1995). To assess emotions in a more general way, however, experimenters often ask participants to rate how frequently or how intensely they experienced an emotion during some specified time period (for reviews, see Larsen & Fredrickson, in press; Watson & Clark, 1997). In these cases, the same emotion term may cover a wide range of different experiences. If the intended meaning is not specified by the researcher, it needs to be inferred from other features of the research instrument.

Second, one such feature is the length of the reference period used, and different reference periods result in reports of different experiences and events. Hence, self-reports pertaining to different reference periods cannot be directly compared with regard to either the frequency or the intensity of the event. With regard to frequency estimates, researchers have long been aware that reports for short reference periods consistently lead to higher estimates than reports for more extended reference periods, a

finding that holds across a wide range of behaviors and experiences (see Sudman et al., 1996, for a review). This observation has typically been attributed to memory loss and biased estimation procedures (see Fiedler & Armbruster, 1994), and these factors undoubtedly play a role. Yet, if participants infer from a short reference period that the researcher is interested in a frequent experience and from a long reference period that the researcher is interested in a rare experience, participants' estimates may be less biased than has been assumed. Instead, they may often deliberately report on experiences of objectively different frequency. Similar considerations apply to the differential intensity of the reported experiences. Although intense experiences are undoubtedly more memorable (see Banaji & Hardin, 1994, for a review), the higher intensity of experiences reported for extended reference periods may again reflect changes in question interpretation as was the case in the present studies. Even though further research is needed to identify the relative contributions of differential memory loss and question interpretation to these phenomena, the present findings highlight that memory is not the only, and not necessarily the most crucial, contributor.

Third, studies that compare concurrent and retrospective reports necessarily confound the concurrent versus retrospective nature of the report with the length of the reference period: Whereas concurrent reports pertain to short periods, with 1 day being the upper limit, retrospective reports pertain to more extended ones. On the basis of the above reasoning, retrospective assessments should therefore result in reports that suggest that the experience under study is more intense but less frequent than is reflected in concurrent reports. The available data are consistent with this implication. As noted in the introduction, Thomas and Diener (1990; see also Parkinson et al., 1995) observed that retrospective reports of emotions reflected a higher intensity than concurrent reports. Similarly, McGonagle, Kessler, and Schilling (1992) observed that daily reports of marital disagreement suggested a frequency of disagreements that was about twice as high as reported retrospectively with a question referring to the period of a month. Although memory biases and motivational factors are likely to contribute to these discrepancies, another likely contributor is again participants' differential interpretation of question meaning as a function of the reference period used. Unless we take this possibility into account, we may profoundly overestimate the extent to which individuals rewrite their personal histories over a period of several weeks.

Finally, it is worth noting that different interpretations of the observed effects result in different methodological recommendations for improving the accuracy of retrospective reports. If discrepancies between concurrent and retrospective reports are mostly due to memory factors, we need to introduce conditions that refresh participants' memory, such as event calendars or personal timelines that invite participants to recall important events that can serve as a useful context for their recall task (see Sudman et al., 1996, for a review). If the discrepancies reflect differential question interpretation, however, such interventions may be less successful than a simple clarification of question meaning.

Reference Periods Indirectly Influence the Interpretation of Vague Quantifiers

The current studies also bear on participants' interpretation of vague quantifiers, which raises additional complications for the interpretation of frequency reports. Specifically, our findings indicate that the specified reference period indirectly influences which frame of reference, or norm (Kahneman & Miller, 1986), participants consider appropriate in mapping the absolute frequency of an event onto a rating scale anchored by vague quantifiers. As a result, the term *very frequently* denotes a lower absolute frequency when participants assume that the question pertains to rare events than when they assume that it pertains to frequent ones (see Pepper, 1981, for a review). This has important implications for the assessment of frequencies.

Not surprisingly, participants' absolute frequency reports, assessed in Study 3, indicated that minor anger is more frequent than major anger and that both types of anger are more frequently experienced during a year than during a week. Yet, these plausible differences could not be observed when participants provided their frequency reports along rating scales anchored by vague quantifiers, reflecting their differential interpretation of the frequency terms (Study 1 and the no-definition condition of Study 2). However, when participants were provided with an explicit definition of anger (Study 2), thus ensuring that all participants drew on the same norm, their ratings did reflect that any experience is more frequent during a typical year than during a typical week.

The observed impact of reference periods and explicit definitions on the use of vague quantifiers adds to the list of complications associated with the use of this popular measurement format. Throughout, reports based on vague quantifiers reflect not absolute frequencies but subjective assessments of these frequencies relative to a norm. The specific norm used is context dependent and can vary as a function of the behavior or experience under study, the frame of reference established in the research situation, and participants' personal history. Consequently, reports based on vague quantifiers cannot be directly compared across behaviors, across research instruments, or across participants (for detailed discussions, see Moxey & Sanford, 1992; Pepper, 1981). Accordingly, researchers are well advised to assess frequency reports in absolute terms. Although such reports are error prone, an absolute response format avoids many of the interpretational ambiguities that plague the use of vague quantifiers.

Concluding Remarks

Over the past decade, researchers have increasingly explored the implications of conversational processes for the study of social judgment, attribution, and decision making (for reviews, see Bless, Strack, & Schwarz, 1993; Hilton, 1995; Schwarz, 1994, 1996). The emerging findings highlighted that research participants are cooperative communicators who do their best to provide the information (they think) the researcher is asking for. Unfortunately, what kind of information that may be is often insufficiently specified, and participants have to draw on contextual features of the research conversation to disambiguate the meaning of the questions asked or the nature of the task presented to them. The accumulating evidence indicates that conversational inference processes contribute to a considerable extent to some of the more puzzling biases and shortcomings of social judgment (Schwarz, 1996). The present findings suggest that research into personality and individual differences may also benefit from a closer examination of conversational processes in research settings (see also Funder, 1987). Unless researchers become sensitive to the differences in meaning conveyed by different question wordings and research procedures, we run the risk of misinterpreting the information participants provide.

References

- Banaji, M. R., & Hardin, C. (1994). Affect and memory in retrospective reports. In N. Schwarz & S. Sudman (Eds.), Autobiographical memory and the validity of retrospective reports (pp. 71-88). New York: Springer Verlag.
- Bless, H., Strack, F., & Schwarz, N. (1993). The informative functions of research procedures: Bias and the logic of conversation. *European Journal of Social Psychology*, 23, 149-165.
- Clark, H. H. (1985). Language use and language users. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 2, pp. 179– 232). New York: Random House.
- Clark, H. H., & Schober, M. F. (1992). Asking questions and influencing answers. In J. M. Tanur (Ed.), *Questions about questions* (pp. 15– 48). New York: Russell Sage Foundation.
- Fiedler, K., & Armbruster, T. (1994). Two halfs may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality and Social Psychology*, 66, 633-645.
- Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, 101, 75-90.
- Grice, H. P. (1991). Logic and conversation. In S. Davis (Ed.), Pragmatics: A reader (pp. 305-315). Oxford, England: Oxford University Press.
- Higgins, E. T. (1996). Knowledge: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), Social psychology: Handbook of basic principles (pp. 133-168). New York: Guilford Press.
- Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin*, 118, 248-271.
- Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4, 401-405.
- Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review*, 93, 136-153.
- Knäuper, B. (1998). Filter questions and question interpretation: Presuppositions at work. *Public Opinion Quarterly*, 62, 70-78.
- Larsen, R., & Fredrickson, B. (in press). Measurement issues in emotion research. In D. Kahneman, E. Diener, & N. Schwarz (Eds.), Foundations of hedonic psychology: Scientific perspectives on enjoyment and suffering. New York: Russell Sage Foundation.
- McGonagle, K. A., Kessler, R. C., & Schilling, E. A. (1992). The frequency and determinants of marital disagreements in a community sample. Journal of Social and Personal Relationships, 9, 507-524.
- Menon, G., Raghubir, P., & Schwarz, N. (1995). Behavioral frequency judgments: An accessibility-diagnosticity framework. *Journal of Consumer Research*, 22, 212-228.
- Mesquita, B., & Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin*, 112, 179-204.
- Moxey, L. M., & Sanford, A. J. (1992). Context effects and the communicative functions of quantifiers: Implications for their use in attitude research. In N. Schwarz & S. Sudman (Eds.), Context effects in social and psychological research (pp. 279-296). New York: Springer Verlag.
- Parkinson, B., Briner, R. B., Reynolds, S., & Totterdell, P. (1995). Time frames for mood: Relations between momentary and generalized ratings of affect. *Personality and Social Psychology Bulletin*, 21, 331– 339.
- Pepper, S. (1981). Problems in the quantification of frequency expres-

sions. In D. Fiske (Ed.), New directions for methodology of social and behavioral science: Problems with language imprecision (Vol. 9, pp. 25-41). San Francisco: Jossey-Bass.

- Pepper, S., & Prytulak, L. S. (1974). Sometimes frequently means seldom: Context effects in the interpretation of quantitative expressions. *Journal of Research in Personality*, 8, 95-101.
- Rosenthal, R., & Rosnow, R. L. (1991). Essentials of behavioral research: Methods and data analysis (2nd ed.). New York: McGraw-Hill.
- Schwarz, N. (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. S. Clark (Eds.), *Review of personality and social psychology: Vol. 11: Research methods in personality and social psychology* (pp. 98-119). Newbury Park, CA: Sage.
- Schwarz, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. In M. Zanna (Ed.), Advances in experimental social psychology (Vol. 26, pp. 123-162). New York: Academic Press.
- Schwarz, N. (1996). Cognition and communication: Judgmental biases, research methods and the logic of conversation. Hillsdale, NJ: Erlbaum.
- Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*, 49, 388-395.
- Schwarz, N., Strack, F., Müller, G., & Chassein, B. (1988). The range of response alternatives may determine the meaning of the question:

Further evidence on informative function of response alternatives. Social Cognition, 6, 107–117.

- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, 48, 813– 838.
- Sperber, D., & Wilson, D. (1995). Relevance: Communication and cognition (2nd ed.). Oxford, England: Blackwell.
- Strack, F., & Schwarz, N. (1992). Communicative influences in standardized question situations: The case of implicit collaboration. In K. Fiedler & G. Semin (Eds.), *Language, interaction and social cognition* (pp. 173-193). Newbury Park, CA: Sage.
- Sudman, S., & Bradburn, N. M. (1983). Asking questions. San Francisco: Jossey-Bass.
- Sudman, S., Bradburn, N., & Schwarz, N. (1996). Thinking about answers: The application of cognitive processes to survey methodology. San Francisco: Jossey-Bass.
- Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology*, 59, 291– 297.
- Watson, D., & Clark, L. A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of Personality As*sessment, 68, 267-296.

Received August 25, 1997 Revision received March 4, 1998

Accepted March 26, 1998