AD-A250 281

MENTATION PAGE

Form Approved
OMB No. 0704-0188

REPORT DATE	3. REPORT TYPE AND DATES COVERED
	FINAL 01 May 88 TO 31 Jan 92

4. TITLE AND SUBTITLE

NORMS AND THE PERCEPTION OF EVENTS

5. FUNDING NUMBERS

G AFOSR-88-0206
PE 61102F
PR 6912
TA OR

6. AUTHOR(S)

Dr Daniel Kahneman

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Department of Psychology
University of California
Berke, CA 94720

8. PERFORMING ORGANIZATION REPORT NUMBER

AFOSR-TR- 92 0393

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Dr John F. Tangney
AFOSR/NL
Building 410
Bolling AFB DC 20332-6448

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The research was described in detail in preceding technical reports. This final report draws extensively on earlier ones, but it is somewhat selective, focusing on the more important and/or successful themes of the research. The report is organized as follows: 1. Studies of normality. 2. Further studies of contingent coding. 3. Processing of dimensional information in priming. 4. The language of counterfactuals. 5. Comparisons of intrapersonal and interpersonal norms. 6. Mental contamination. 7. Unintended comparisons. 8. Topic and referent in perceptual comparisons. 9. Anchoring.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
(U)	(U)	(U)	(U)

RESEARCH COMPLETED AND IN PROGRESS

The contract AFOSR-89-0206 was the main source of support for my research and that of my students from June 1988 to December 1991. During this period, one student whose work was supported primarily by the AFOSR contract obtained her PhD degree (Carol Varey), and two others are in advanced stages of their thesis research (Suzanne O'Curry and Maria Stone). Two undergraduates have completed publishable honors theses: one is in press (Craig Fox) and the other will be submitted shortly for journal publication (Paul Grant).

The focus of the research shifted gradually over the three years of the grant. The initial emphasis was on studies of surprise and spontaneous categorization, designed to provide new tests of norm theory. This line of work proved disappointing, and in the second and third years of the project the effort moved to a broder treatment of themes of norm theory, combining these themes with more successful lines of research on comparison processes, counterfactuals, the evaluation of consequences, and multiple computation.

The research was described in detail in preceding technical reports. This final report draws extensively on earlier ones, but it is somewhat selective, focusing on the more important and/or successful themes of the research.

The report is organized as follows:

1. Studies of normality
2. Further studies of contingent coding
3. Processing of dimensional information in priming
4. The language of counterfactuals
5. Comparisons of intrapersonal and interpersonal norms
6. Mental contamination
7. Unintended comparisons
8. Topic and referent in perceptual comparisons
9. Anchoring

1. STUDIES OF NORMALITY (see 1989 report)

Much effort in the first two years of the project was dedicated to studies in the ***normality paradigm***, a technique which provides a somewhat unusual measure of encoding and spontaneous categorization. The stimuli in our experiments are static images or short sequences of motion, presented on a CONRAC Model 7250, controlled by an IBM-AT, equipped with an Artist-plus graphics board and run on software developed in our lab.

The set of stimuli used in an experiment is defined by pairs of values on three attributes. Two training patterns (or norms) are defined by complementary values on these attributes. The remaining six possible combinations of the attribute values define test patterns. For example the two norms may be specified by values of shape (circle/triangle), color (pink/blue) and position (left/right of screen). Another example would be gender (face of boy/girl), expression (frown/smile) and label (nonsense syllables presented on the screen, e.g. UZU/GAR). Subjects are asked to observe training patterns, which are repeated with high frequency. They answer questions about the normality of an attribute on each test pattern. For each test pattern we ask two questions, on different trials, probing the two features that it shares with its nearest neighbor among the two norms. The dependent variable is the proportion of subjects who rate a particular attribute of a test trial as normal.

Consider an example in which the norms are Dark/Circle/Left and Light/Triangle/right. There are six test patterns, defined by the remaining conjunctions of the three binary attributes. Consider, for example, a test pattern consisting of a dark triangle on the left of the screen. The nearest norm (reached

by changing one feature) is the dark circle on the left. The two questions that will be asked for this test pattern are 'IS COLOR NORMAL?' and 'IS POSITION NORMAL?'.

Note that each of the two attributes that are ***not*** queried on a particular trial 'votes' for a yes or no answer to the question. In each case, the vote is split. (In the example above we do not ask 'IS THE SHAPE NORMAL?' because the other two attributes of the test pattern ***both*** vote 'No'). The balance of answers provides a measure of the relative weights of the two attributes in determining the response.

The sequence of events in an experiment is as follows. Subjects are first exposed to 12 observation trials in which the two norms occur with equal frequency. They are then shown one of the test patterns and are asked to make a dichotomous judgment of the normality of one of its attributes, by writing Y or N in an answer sheet. This is followed by four additional exposures of the training patterns (two for each), then another test pattern, then four more training trials, and so on until 12 test trials have been presented. In summary, the experiment consists of 116 trials: each of the two norms is presented 52 times, and each of the 6 test patterns is presented twice, paired on each occasion with a different question. The duration of such an experiment is approximately 4 minutes. We have run about 100 of these experiments so far, with 12-20 subjects in each.

Subjects were run in groups of up to four, and in most of our experiments four such groups were run. A session lasting 45 minutes could include up to ten separate experiments, using unrelated norms. The conjunctions of attributes defining the two norms were different for the different sub-groups participating in an experiment.

A natural way to interpret a question about the normality of an attribute is by expanding it. As it stands, the question is ambiguous because it can reasonably be expanded in several ways. Thus, the question 'IS THE COLOR NORMAL' could be understood as an abbreviation of 'IS THE COLOR NORMAL FOR THIS POSITION?' or 'IS THE COLOR NORMAL FOR THIS SHAPE?' or perhaps 'IS THE COLOR NORMAL FOR THIS CONJUNCTION OF SHAPE AND POSITION?'. The correct answers would vary accordingly. In practice, we allow only yes and no as answers to the questions, and hope to infer from the answers how the question was interpreted. The interpretation of the question, in turn, is expected to provide an indication of the role of different attributes in the spontaneous categorization (encoding) of patterns and events: in choosing to evaluate the normality of the position of the stimulus 'for' the shape, rather than for its color, the subject implicitly categorizes the event in one way rather than another.

In most of our research in this general design, position was one of the attributes varied throughout the series of experiments included in a session: in each of the experiments four of the stimuli were presented on the left of the screen, four on the right. In another study one of two nonsense-word labels was presented with each stimulus. We refer to such an attribute (position or label) as a ***medium*** for the study of the relations of the two other attributes manipulated in any given experiment.

To facilitate the analysis, we adopt a consistent convention in grouping and labeling the six measures. In each case we distinguish a primary dimension (labeled A), a secondary attribute (B) and the attribute of position (P). In the shape/color example introduced above, the A-attribute is shape and the B-attribute is color, by hypothesis.

The following examples assume that the training patterns are:

Pink circle on left Blue triangle on right

P? A+ This is the position question, when the A dimension votes 'yes' (the B-dimension votes 'no'). For example,

'IS POSITION NORMAL?' for a blue circle on the left.
P? B+ Same question, with B-dimension voting 'yes', e.g.,
pink triangle on left.

Note that the comparison of the answers to the two questions provides a fairly direct measure of the ***relative*** weights of shape and color, or of the strength of the tendency to interpret the question 'is position normal?' as 'normal for this shape' or 'normal for this color'. The question can be construed as a passive version of the sorting task often used in categorization research: instead of asking subjects to assign events to a location, we provide a location and require them to evaluate it.

The normality questions provide three independent tests of the relative importance of the two critical dimensions: the dimensions are pitted against one another in one set of questions, and their influence on one another is compared to that of position in the other sets. The original hypothesis was that the three tests would generally agree: the more important (less mutable) dimension was expected to control sorting or categorization, as indexed by the P? question. The facts turned out to be more complex than my simple notions of mutability and dimensional importance had suggested. Selected results are presented below. The first and second technical reports provide a more detailed discussion.

Color vs. Shape

There is a venerable hypothesis that adults and even young children find it more natural to classify objects by shape than by other attributes. In part as an attempt to validate the technique, we investigated this question in a number of experiments. Table 1-1 summarizes the experiments in which the critical attributes were shape and color.

Table 1-1

P?		P-		P+		
A+	B+	A?	B?	A?	B?	Condition
85	24	50	52	79	53	simple shapes -- distinct colors
91	12	34	50	91	53	complex shapes -- distinct colors
87	02	54	79	87	48	simple shapes -- similar colors
67	07	57	53	90	83	rectangles varying in orientation
50	32	39	39	86	71	rectangles varying in length
58	29	33	29	87	92	rectangles varying in aspect ratio
94	19	25	37	91	66	simple shapes differing by small feature

The first two rows of Table 1-1 illustrate the dominance of shape in two variations in which the shapes were distinctive geometric figures and the colors were highly discriminable. Two of the three manifestions of dominance are present: subjects are very likely to judge position normal if the shape of the test pattern corresponds to the shape of the training pattern usually shown in the same place. In the third test (the P+ questions) subjects are also significantly more likely to judge abnormal a color that is paired with the 'wrong' shape than a shape paired with the wrong color. However, there is no indication of dominance in the P- questions: when a training pattern is shown intact in the wrong place, subjects tend to assign the same ratings of normality to both attributes: the mean judgments are similar and the correlation is substantial.

The next rows of the Table summarize several experiments designed to clarify the role of discriminability of both attributes. Making the colors quite similar (though still easily distinguishable)

had little effect. On the other hand, the dominance of shape was clearly reduced when the shapes were rectangles -- they appeared in that case as color patches. The most important result in this series is the effect of varying shape by adding a small feature, as illustrated in Figure 2. This feature is less obvious and almost certainly less discriminable than the color difference between the norms, by standard measures, such as identification threshold or speed of same-different judgment. Although formal experiments will be needed to nail down the point, we are confident that neither discriminability nor an impression of within-attribute differences can explain the pattern of normality judgments. In general, the variations of within-attribute similarity, although they had some effect, did not reverse the dominance of shape over color.

The observations on shape and color collected so far suggest two conclusions: (1) The dominance of shape is a robust result, which may depend more on the ***individuality*** of the shapes than on their discriminability. (2) Color is subordinate to shape in these judgments, but is not nested within the shape attribute -- i.e., the judgment of the normality of color is not screened off from the attribute of position.

Sequences

Table 1-2 summarizes experiments in which the display consisted of a series of events. In the first of these experiments a letter was seen, which appeared to move to another location and simultaneously to change into another letter. We had expected that the second event in the sequence would be coded as subordinated to the first, and perhaps even as contingent on it. Nothing of the kind happened. The failure to obtain dominance in this simple situation is a significant result, because it eliminates a plausible interpretation of normality ratings as reflecting the confirmation or disconfirmation of expectations. The critical comparison is between the ratings of the first and of the second events in the P+ condition. In the A? case the first event occurs in its usual place and is followed by an unusual sequel; in the B? case, the second event does not correspond to the letter that just preceded it. However, subjects rated both events normal, indicating that the relation to position was more important than the sequence of expectations and confirmations.

Table 1-2

	P?		P-		P+		
	A+	B+	A?	B?	A?	B?	Condition
	27	27	58	42	96	92	simple sequence of two letters
*	69	31	81	77	69	50	character appearing in frame
*	60	17	60	63	53	37	sequence of two distinctive motions
*	72	03	31	34	66	56	complex motion --> simple motion

Objects and Motions: Conditions for Contingent Coding

Dominance of one attribute over another was quite often observed in the results presented so far, but contingent coding was striking by its absence. Contingent coding was defined in the section that introduced the normality technique by a particular pattern of answers on the P- questions: low on the A? question, higher on the B? question. Evidence of contingent coding is finally found where there was most reason to expect it, in judgments of the normality of objects and their actions. The precise conditions under which contingent coding is found -- in contrast to mere dominance -- now appear to be quite an interesting problem, which we plan to explore in further work.

Table 1-3 presents results for conditions involving objects, actions and changes. We briefly consider the conditions in turn.

Something resembling the expected pattern of contingent coding was obtained where the norms were schematic faces of a boy and a girl, one frowning and the other smiling. Phrasing the normality questions was awkward. We ended up asking "Is the character normal?" for the A? question and "Is the expression normal?" for the B?. Half of the responses to the former question in the P- condition were positive, indicating a tendency to judge the norm display as normal when it is presented intact in a new position. The proportion of positive responses to the B? question was very significantly higher, suggestive of contingent coding. However, the character was judged abnormal when paired with the wrong expression in the P+A? condition.

Table 1-3

P?		P-		P+		
A+	B+	A?	B?	A?	B?	Condition
97	10	47	87	27	17	faces and expressions
89	26	46	80	61	30	motion, ask by object
83	39	57	72	48	28	occluded, ask by object
96	23	62	92	38	35	motion (frozen), ask by object
77	31	54	88	46	31	occluded (frozen), ask by object
83	83	25	33	92	83	motion/color, ask by color
92	75	29	29	96	71	occluded/final color, ask by color
68	54	43	50	82	50	shape/motion, ask by shape
71	75	39	46	82	54	occluded, shape, ask by shape
72	72	31	44	97	66	shape/motion, large shapes that sit
87	84	37	56	81	59	shape/occluded, large shapes
77	53	40	70	87	30	shape+color/motion, ask by shape
80	53	53	73	100	37	shape+color/occluded, ask by shape
98	07	39	87	52	20	destination, ask by object
96	31	42	81	31	19	destinations (frozen), by object
93	43	29	71	93	32	shape/destination, ask by shape
93	27	47	77	93	20	shape+color/destinations, by shape
96	14	39	71	93	68	changing colors
87	03	37	56	91	50	shape/changing salient colors

Several experiments were carried out in an attempt to identify the conditions that produced the new pattern of results in the motion and occluded-motion experiments. As shown in the Table, one feature of the results depends on asking subjects about the normality of the ***object***, rather than of a particular attribute: low ratings were given in the P+A? condition, presumably because the object appeared abnormal when it behaved abnormally. This was not true when respondents evaluated the normality of 'the shape'. The more significant feature of the results is the discrepancy between the normality ratings for shape and motion in the P- condition: the dissociation vanished when the moving objects were distinguished by a single feature (either shape or color); it was restored when the objects were defined by a conjunction of features, although the normality of the shape (not the object) was judged.

A pattern of contingent coding was also obtained in another condition, in which objects appearing on one or the other side of the display (labeled 'starting positions' in the questions) moved to one or another

marked destinations, above and below the center of the display (labeled destinations). In the P- condition the object (or 'shape', or 'color' in different experiments) is consistently judged less normal than the destination. Evidently. the evaluation of the destination is conditioned on the object, not on its starting position.

Some evidence of contingent coding was also observed in a 'changing color' display, in which the norms are distinctive white shapes, which gradually take on distinctive colors. The situation is informationally equivalent to that investigated in the shape/color experiments described earlier, but the judgments indicate a stronger tendency to rate the shape by its position and to relate the color to the shape.

2. FURTHER STUDIES OF CONTINGENT CODING (from 1990 report)

The next series of experiments was intended to investigate the factors involved in dominance and in contingent coding, and at the same time to examine an alternative interpretation of our other results, which would explain judgments of normality in terms of discriminability or similarity.

Four conditions in the present series were an attempt to address directly the relationship between discriminability and normality. All these conditions used position as the third attribute. In conditions 1 and 2 the displays were static: attribute A was shape and B was size. In conditions 3 and 4 the displays were moving objects: attribute A was shape and B was direction of motion. In conditions 1 and 3 the two norm shapes were distinctly different and color was held constant across the two norms. In conditions 2 and 4 the two norm shapes were very similar and the colors differed but were similar. Thus, in conditions 2 and 4 a cluster of features defined attribute A in the design. The normality of A was probed by shape (e.g., "Is the shape normal?").

Discriminability data were collected in a pilot experiment with 6 subjects, to ensure that the differences between the norms in attribute A in conditions 2 and 4 (where this attribute is defined by a conjunction) were not more discriminable than the differences in attribute A in conditions 1 and 3, respectively. Subjects were instructed to assign the two norm stimuli to different response keys. A series of 50 trials was then presented, with a single stimulus shown on each.

All norm pairs actually used in the normality experiment satisfied the following condition: attribute clusters defining A were not more quickly distinguished from each other than single attributes defining A in the comparable condition. Reaction times for clusters were either equal to, or longer than, reaction times for the single-attribute comparison.

The normality results are shown in Table 2-1.

Table 2-1

	P?		P-		P+	
	A+	B+	A?	B?	A?	B?
1	100	0	75	93	79	7
2	93	0	79	89	82	11
3	64	32	75	79	75	29
4	82	25	86	82	75	36

Conditions 1 and 2 show dominance but not contingent coding, and also indicate that A is normal even when size is not. This replicates previous results for size and shape and extends the previous finding of

shape+color and size to cases with low discriminability. No differences were observed between conditions 1 and 2. The results for conditions 3 and 4 show a similar pattern to 1 and 2. There is no sign of the contingent coding that we expected in condition 4.

It seems that the link between A and B attributes is so strong in all conditions that normal pairing dominates position (the 3rd and 4th columns). However, abnormal values on B are not sufficient to make A abnormal (column 5).

Four other conditions investigated normality for a particular class of highly-individuated stimuli -- words. Again, all four conditions used position as the third attribute.

Conditions were as follows:
condition 5. A = word; B = size+color, prompted by size
condition 6. A = non-word; B = size+color, prompted by size
condition 7. A = word; B = color
condition 8. A = word; B = underlining type and color

The results are presented in Table 2-2.

Table 2-2

	P?		P-		P+	
	A+	B+	A?	B?	A?	B?
5	86	18	89	100	46	7
6	82	7	86	75	57	7
7	96	0	79	93	68	14
8	96	0	75	96	82	7

In all conditions the word dominated other attributes. This was also true for the non-word in condition 6, which indicates that familiar associations and meaning are not the mediators of normality in these conditions. As in conditions 1 to 4, abnormal secondary attributes do not make the word abnormal. There is some tendency for contingent coding in conditions 7 and 8, but there is still a strong tendency to respond that the word is normal if its secondary attributes are (more so if it is supported by a cluster of attributes, as in condition 5). We had expected that condition 8 would provide the most likely condition for contingent coding. The observed result, though in the expected direction, was much weaker than anticipated.

Our inability to get control of contingent coding was disappointing. I decided to set aside for the moment the pursuit of the normality measure and to focus on other experimental problems in the same general area.

3. PROCESSING OF DIMENSIONAL INFORMATION IN PRIMING
Kahneman, Gibbs and Treisman

In previous work undertaken in collaboration with Anne Treisman and Brian Gibbs, I have studied an effect that we labeled 'object-specific priming'. The target stimulus in most of our studies was a letter that was to be named as quickly as possible. The target was contained in one of several objects, e.g.,

outline squares. The essential feature of the situation was that the whole set of squares had just arrived from an original position -- the movement time ranged in different studies from 80 to 600 msec. While the squares were stationary in their initial positions and just before they started to move, letters briefly appeared in them. These are the primes. The main result of our study was that there was a priming effect of presenting the target letter in the initial display, but only if the prime appeared in the same square that later contained the target. Indeed, the standard result with letter stimuli (words are different) is that presenting the target letter in the 'wrong' object yield little or no benefit compared to a control condition in which the target is not primed at all. Hence the label 'target-specific priming'.

An obvious question about this priming effect is the level of encoding at which it arises. Applying a fairly standard diagnostic, Treisman and I conducted an experiment to test whether the object-specific priming effect is also case-specific. We varied the case of the prime and of the target independently, and observed that priming was diminished when the case varied between prime and target. Brian Gibbs followed up with a Master thesis in which he required subjects to respond to a particular feature of the stimulus (e.g., its shape, size or color), allowing the prime and the target to vary in response-irrelevant attributes. We considered these results equivocal, and decided to clarify the issue in a series of experiments, which was conducted in the fall of 1989.

The common feature of the experiments is that the displays consist of four white squares, which contain colored letters. A priming pattern is first shown around a fixation cross. It is then removed, and a target field is immediately shown. There are four possible positions of the target field -- computed by moving the whole pattern so that one of the four initial squares is centered on the fixation cross. The sequence of displays yields a powerful impression of coherent motion. Object-specific priming can be studied by comparing performance in several cases: (1) when the target matches the prime stimulus shown in the same object; (2) when the target matches the prime stimulus shown in another square; (3) when the target does not match any of the primes. Figure 3.1 illustrates the first of these cases. It is also possible to construct tasks in which the prime and the target are not physically identical, but differ in case, color, size or other attributes. The project was designed to study the effect of such manipulations of prime-target resemblance.

Size priming with shape/character varied

In this experiment the stimuli were two red capital letters (Y and O), in two sizes, 3.3 and 6.5 mm tall. Each letter was centered in a white square measuring 20.3 mm. The priming display always contained two large and two small characters. It was presented for 100 msec and was immediately followed by the target field (see Figure 3.1). The subject indicated the size of the target character marked by the cross-hairs, by pressing one of two keys assigned to different hands. Table 3.1 presents the reaction time for 'large' and for 'small' responses, as a function of the agreement between the target and the character presented in the 'same' square in the original display.

Table 3.1 -- Reaction time to size discrimination with irrelevant variation of shape/character

Agreement		Target Size		
		Large	Small	Mean
Size	Shape			
+	+	511	470	491
+	-	506	480	493
-	+	517	494	506
-	-	524	490	507

The results are unequivocal: there is a substantial object-specific priming effect (14 msec, t(15) = 4.52, p<.01) and not a trace of interaction with the shape of the stimulus.

Color priming with shape/character varied

The design of the experiment was the same as the preceding one. The subject now responded to the color of the character that appeared in the target position, by pressing a key. The possible colors were red and green. The temporal parameters were the same as in the previous experiment.

Table 3.2 -- Reaction time to color discrimination with irrelevant variation of shape/character

Agreement		Target Color		
		Red	Green	Mean
Color	Shape			
+	+	481	461	471
+	-	480	475	477
-	+	517	489	503
-	-	508	486	497

Again, the results are quite clear. There is a substantial object-specific priming effect (26 msec, t(15) = 6.36, p<.01) but the interaction of color and shape similarity is not significant (t = 1.54). There is no evidence that object-specific color priming is affected by the identity of the prime and target characters.

Letter priming with case variation and key response

The accumulation of evidence for independence in the processing of different dimensions of the stimul[illegible] was sufficiently impressive to justify a partial replication of the Kahneman-Treisman experiment sutdy of the effects of case identity on object-specific priming. The earlier experiment had been conducted with a different display, in which only two squares were shown in 'real' motion, and where the subject made a vocal response to indicate reading the letter. For the present experiment we adopted the display and design of the two preceding studies. There were four squares, and two possible target characters (G and D). The subject responded to the identity of the target letter by pressing a key. The exposure duration of the prime was 100 msec. The results are shown in Table 3.3.

Table 3.3 -- Reaction time to letter discrimination with irrelevant variation of case

Agreement		Target case		
		Upper	Lower	Mean
Letter	Case			
+	+	502	498	500
+	-	511	502	507
-	+	537	526	531
-	-	549	528	539

The now familiar pattern of results is observed again: a robust object-specific priming effect of 30 msec (t(11) = 4.02) when the prime and the target have the same case, 32 msec when the case varies (t=4.59). There is of course no trace of an interaction.

Letter priming with case variation, vocal response

We now decided to replicate the original case experiment, using a vocal response, in the four square display, in an attempt to identify the boundary conditions for the interaction of object-specific priming with case identity. The display conditions were the same as in the preceding study but the vocabulary of possible stimuli was explanded to 8 letters (B,D,G,H,N,R,Q,T), and vocal RT was measured. Table 3.4 shows what happened. The larger vocabulary allows a control condition in which the target letter is not presented at all in the priming display. This is useful, because the object-specific effects observed in the key-press experiments are the sum of object-specific priming (when there is a match between prime and target) and inhibition (in cases of mismatch). Results for this control condition are shown in the bottom row of the Table.

Table 3.4 -- Vocal reaction time in letter naming with irrelevant variation of case

		Target case		
Agreement		Upper	Lower	Mean
Letter	Case			
+	+	479	481	480
+	-	481	473	477
-	+	491	490	491
-	-	494	483	489
Unprimed letter		492	484	488

The comparison with the control indication indicates that there is no trace of priming except when the prime and the target are shown in the same object. The results also show that there no significant inhibition is produced by presenting the target in the 'wrong' object. The object-specific priming is smaller than in some of our previous work, is the same when case is identical and when case is different (11 and 12 msec, respectively), and is significant in both cases ($t(11) = 2.75$ and $t = 3.30$, respectively). The results are quite consistent with the other experiments in this series, but diverge from those previously obtained by Kahneman and Treisman, which used a somewhat different display, where the object-specific priming was 21 msec when case was identical and 8 msec when it varied between prime and target. We are at the moment at a loss to explain the difference.

Categorization of characters with case varied

In the final experiment in this series, we returned to the key-press response. The subject's task was to press one key for letters in the first half of the alphabet (A,E,G vs N,Q,R). The priming display and the target display both consisted of two letters each from each category, one in upper and one in lower case. Except in the last condition of Table 3.5, the target letter was always present in the priming display, sometime in the same case, sometime in a different case. The results are shown in Table 3-5.

There is significant priming when the target letter that is to be categorized has been presented in the same object, both when case is the same (16 msec, $t(19) = 3.55$) and when case varies (13 msec, $t = 2.26$). The effect of case identity is not significant ($t = 1.00$). There is a small but probably reliable advantage of showing the target in a square that previously contained another letter in the same category: the overall difference between rows 3,4,5 and rows 6,7,8 averages 6 msec, $t(19) = 2.12$, $p<.05$. However, the advantage of priming by the same letter is significantly greater (for the comparison of rows 1,2 to rows 3,4, $t(19) = 4.76$).

Table 3.5 -- Categorization time with variation of case

Agreement within target object	Target letter in priming field	
1) Same letter	same case	531
2) Same letter	different case	535
3) Same category	same case	551
4) Same category	different case	543
5) Same category	absent	546
6) Different category	same case	551
7) Different category	different case	553
8) Different category	absent	555

The findings of this experiment further confirm object-specific priming (or interference). They also provide evidence that the effect is produced in part by pooling of response tendencies or by high-level categorization -- the category priming effect observed here, although quite small, is theoretically significant. The results also indicate that there is something special about case -- a conclusion also suggested by other findings in the reading literature. It could have been argued that the only thing that the upper and lower case representations of a letter have in common (if physically dissimilar) is that they map onto the same response. But merely mapping onto the same response could not explain cross-case priming, because the different letters in a category also map onto a response, in the present experiment. The upper and lower case versions of a letter appear to be 'the same', for the purpose of priming, just as a green and a red version of the letter would be. The absence (or weakness) of within-category priming must be interpreted together with the total independence of dimensions processing observed in the other experiments of this series. Taken as a set, these findings suggests that priming occurs at the level of what Treisman calls 'feature maps'.

4. THE LANGUAGE OF COUNTERFACTUALS: 'ALMOST' AS AN INDICATOR OF PROPENSITY AND PROXIMITY (Kahneman and Varey, 1991)

One of the central tenets in norm theory (Kahneman and Miller, 1986) is that the normality of an event is assessed by comparing it to the norms that it evokes retrospectively. The treatment of counterfactuals is a central problem in that theory. For the past year Carol Varey and I have been engaged in the study of a particular class of counterfactual assertions. Many situations are aptly described by such phrases as 'Team A almost won', 'Tom almost died', 'Joan almost got married to Ted'. Use of the word 'almost' to describe achievements that came close to happening is an example of ***spontaneous*** generation of counterfactual alternatives to the actual outcome. The near-outcome is so readily available that the counterfactual is not expressed as a counterfactual conditional with a specified antecedent. We call these assertions ***close counterfactuals***, and the attempt to explore what can be learned from them about intuitive notions of probability and causality has been a focus of my effort this year under the AFOSR contract. Much of the effort involves conceptual analysis, but we have also run several questionnaire studies eliciting intuitions about appropriate uses of 'almost'. A paper describing some of the results of these studies appeared in the ***Journal of Personality and Social Psychology***.

A treatment of the psychology underlying close counterfactuals turns out to be inextricably linked with an investigation into some aspects of causality and probability. Counterfactual assertions normally invoke causal beliefs and assign degrees of probability or plausibility to unrealized outcomes. Accounts of causality, in turn, often invoke counterfactual beliefs (for example, about what would have happened in the absence of a putative cause) as well as notions of conditional probability. Finally, notions of objective probability often rest on intuitions about causal systems. The present studies are concerned with a psychological study of this nexus of issues.

Our approach combines some simple phenomenological observations and a basic linguistic inquiry into the conditions under which close counterfactual assertions are appropriate. The genre is not unknown in psychology. Studies of what people mean when they say that 'John went to the restaurant', or when they use the words 'can' and 'try' have contributed significantly to our understanding of how people think about events and actions. In the present studies we examine the use of the word 'almost' to explore how people think about counterfactuals, probability and causation.

We restrict our discussion of 'almost' to cases in which the actual outcome X, or the near-outcome Y, is an ***achievement*** (see Lyons, 1977) -- a change of state that occurs at a particular moment, usually as the culmination of a longer causal episode. We analyze the beliefs that a speaker expresses by the assertion that an individual almost died, or almost missed a deadline, and examine what such beliefs can teach us about the cognitive representation of uncertain events and of causal propensities.

Students at the University of California at Berkeley served as subjects. They were recruited by posters displayed outside the student union offering a small payment for immediate completion of a questionnaire. Respondents were given instructions followed by approximately fifteen questions. An illustration is given below:

> In the following questions you are asked to rate statements on a scale from "Appropriate" to "Very Peculiar". A set of statements is presented for each question. You are to rate whether the last statement fits well with those that preceded it.
>
> (1) At the end of a long game of chance, John could have won the whole pot if a die that he rolled showed a six. The die that he rolled was loaded to show six 80% of the time. John rolled it and it showed a two. John almost won the whole pot.
> Appropriate ___ Somewhat Peculiar ___ Very Peculiar ___
>
> (2) Tom almost died but in fact he was never in real danger.
> Appropriate ___ Somewhat Peculiar ___ Very Peculiar ___

Some of the questions were paired with similar questions in a between-subjects design. For example, one variant of example 1 provided the same scenario, but asked subjects to judge the statement 'John almost threw a six'. Some subjects were also asked to make within-subject comparisons. An example follows:

> (3) John played in a game of chance involving six die throws. He would have won the whole pot if he had thrown six sixes in a row. He threw five sixes and a five.
>
> Fred played in a game of chance involving five die throws and a coin toss. He would have won the whole pot if he had thrown five sixes and tossed heads. He threw five sixes and tossed tails. Which of the following is more appropriate:
> a. John almost won the whole pot.
> b. Fred almost won the whole pot.
> c. Both are equally appropriate.

We next briefly discuss some major conclusions of our analysis of close counterfactuals, illustrating them with selected examples of the data we have collected.

The objective stance. close counterfactuals are treated as a matter of objective fact, in the sense that their truth or falsity does not depend on the beliefs of any individual or community. The event that

almost happened did not really happen, and in that sense does not belong to reality -- but the fact that it almost happened is treated as real, not as a mental event such as a fantasy or an imagining.

(4) Everyone thought Phil almost died.... but in fact he was never in real danger.
Appropriate 69% Very peculiar 10% (N = 29)

(5) Tom almost died.... but in fact he was never in real danger.
Appropriate 7% Very Peculiar 66% (N=29)

An objective attitude similar to that which is applied to counterfactual statements is also adopted when people talk of causes -- these are viewed as facts about the world, not as subjective events. An objective attitude also characterizes many probability statements -- when probability is taken to describe a disposition or causal propensity of a system rather than a state of belief. (Contrast 'the probability that the ball drawn from the urn would be red was .60' with 'the probability that the Nile would be longer than the Amazon was .60'.)

Propensities and dispositions. We draw a distinction between two kinds of assessment of the probability of a particular outcome at the end of an event episode. A ***disposition*** for the focal outcome is the probability of the focal outcome as assesses prior to the initiation of the episode. A ***propensity*** for the focal outcome is the probability of the focal outcome as assessed from event cues during the course of the episode.

The key observation about close counterfactuals is that strong prior dispositions are not sufficient to support the statement that an outcome almost occurred. Event cues supporting a strong propensity are required. This is illustrated by the following examples:

(6) John rolled a die that was loaded to show six 80% of the time. John rolled it and it showed a two.... John almost threw a six.
Appropriate 6% Very peculiar 62% (N = 32)

(7) Tom almost registered for the tournament. He would have won if he had played... Tom almost won the tournament
Appropriate 10% Very peculiar 62% (N = 40)

Proximity, progress, and sensitivity to obstacles. People are sensitive to a dimension that is commonly described as the distance between states of the world at different points in time. The representation of causation as movement through space and as the overcoming of obstacles along the way is involved in a rich family of metaphors -- 'coming close' is one of many. We have examined some of the factors that control impressions of distance, including the number of intervening causal stages, the decisiveness of the intervening events and the possible obstacles in the path to the focal outcome.

One series of questions focused on cases in which an individual 'wants X' or 'considers doing X'. We were interested in identifying cases in which such intentional states would support the statement that the individual 'almost got X' or 'almost did X'. Some examples follow

(8) Martin considered getting married to Meg. Martin almost married Meg
Appropriate 14% Very peculiar 34% (N = 29)

(9) Neil considered not getting married to Amanda. Neil almost didn't marry Amanda
Appropriate 62% Very peculiar 19% (N = 32)

(10) Fred considered stealing his child's savings. Fred almost stole his child's savings.
Appropriate 30% Very peculiar 16% (N = 32)

(11) Ned considered breaking into a bank vault. Ned almost broke into a bank vault
Appropriate 18% Very peculiar 44% (N = 32)

Consideration of an action supports the assertion that it was almost performed only when (1) a relatively small number of steps intervene between the thought and the action; (2) consideration may be assumed to suggest a possible desire to perform the action; and (3) when the individual who considered the action could reasonably be thought to be capable of it. In a romantic relationship, either individual has the power to terminate it and thinking about breaking up may imply dissatisfaction. An individual who considers marrying someone, or even clearly wishes to marry that person, may be quite far from being able to carry out the intention. Our subjects' responses clearly differentiate these cases. Subjects are also sensitive to the fact that much more remains to be done, beyond mere consideration, for the project of breaking into a bank vault than for stealing one's child's savings.

Conclusions

On the basis of the data collected in our surveys and general linguistic intuitions, we claim support for the following conclusions:
(1) Counterfactuals, causes and (some) probabilities are treated as facts about the world, not as constructions of the mind.
(2) The absence of perfect hindsight indicates that people attribute inherent uncertainty to some causal systems -- what happened is not treated as necessary or inevitable.
(3) Probabilities of outcomes can be assessed on the basis of advance knowledge (dispositions) or of cues gained from the causal episode itself (propensities). The distinction is critical to the use of 'almost', which requires the attribution of a strong propensity to the counterfactual outcome.
(4) Cues to propensity are the temporal or causal proximity of the focal outcome, and any indications of accelerated progress.
(5) A general schema of causal forces competing over time is applicable to many achievement contexts.
(6) Dispositions that are not supported by event cues will be neglected in retrospective judgments of outcome probability.
(7) Conversational pragmatics allow more latitude in the acceptance of 'almost' when the speaker is emotionally involved in the near-outcome.

5. COMPARISON OF INTRAPERSONAL AND INTERPERSONAL NORMS

Two separate projects were concerned with the relative weights of different norms in comparisons. Craig Fox and I studied the role of these norms in reports of satisfaction with various domains of life. Paul Grant and I conducted several experiments to find out if people simultaneously apply interpersonal and intrapersonal norms to the evaluation of a single performance, or choose between these norms.

Norms in Judgments of Satisfaction
(Fox and Kahneman, in press)

A basic finding of well-being research is that objective circumstances and actual achievements are poor predictors of satisfaction with financial status, grades, physical condition, and other life domains (Argyle, 1987). Instead, satisfaction is mainly determined by an explicit or implicit comparison of the current state to some reference norm or standard. One tradition of research has emphasized the role of social comparisons in determining feelings of satisfaction or relative deprivation (e.g. Festinger, 1954). Another tradition has emphasized comparisons to an adaptation level, which is mainly determined by the individual's personal history (e.g., Brickman and Campbell, 1971; Helson, 1964). A study by Emmons and Diener (1985) compared the importance of the two norsm, by comparing the correlations of judgments of interpersonal and intrapersonal satisfaction with global assessments of satisfaction, for several domains. They observed that the correlations were higher with interpersonal comparisons. Surprisingly, this pattern was present in private domains, such as love life and intimate friendships, where comparisons are unlikely.

It is intuitively appealing that global variables should be predicted and explained by their more specific constituents. In the context of well-being research, this intuition suggests that global satisfaction with life should be explained by satisfaction with various life domains, and that satisfaction with each domain should be explained in turn by more specific measures, such as evaluations of inter- and intrapersonal comparisons. However, the constructionist perspective suggests caution. In this approach, difficult judgments are made by using the most accessible relevant information and by relying heuristically on simpler judgments or on other accessible cues such as current mood (Schwarz and Clore, 1983). This analysis suggests the perverse hypothesis that the correlation between judgments of social comparison and of global satisfaction may be especially high in domains where people know little about others. In such cases, of course, subjective social comparison is an ad hoc construction that plays little or no part in the causation of satisfaction.

STUDY 5-1

The first study consisted of a partial replication of the Emmons-Diener survey, with one new measure: we asked respondents to evaluate the ***importance*** of recent changes and of social comparison in their previous ratings of satisfaction. From the set of domains studied by Emmons and Diener we selected three "public" domains in which we expected social comparison to be highly accessible (physical attractiveness, grades, and housing) and two relatively "private" domains in which information about others is likely to be more ambiguous (friends and love life). Our hypothesis was that social comparison would be considered more important in the public than in the private domains, while correlations would show the opposite pattern.

Method. The sample consisted of 149 students (95 men, 52 women, 2 unreported) registered at U.C. Berkeley.

Results and Discussion. Table 5-1 lists mean importance ratings for social comparison and for change, Pearson correlations of these variables with satisfaction and with each other, and standardized beta weights for the prediction of satisfaction. The pattern of beta weights closely replicates the results of Emmons and Diener. The finding that the beta weight for social comparison is especially high for love life is also replicated.

TABLE 5-1
Mean importance ratings, standardized beta weights and correlations for five domains of satisfaction

	Importance		Std. ß-weight		Correlation Coefficient		
DOMAIN	SOC	CHG	SOC	CHG	SOC-SAT	CHG-SAT	SOC-CHG
Friends	2.66	3.58	.45	.28	.54	.41	.29
Love Life	2.58	3.39	.85	.25	.82	.53	.45
Grades	3.32	3.92	.64	.41	.59	.40	.12
Housing	2.82	3.26	.49	.35	.62	.57	.54
Attract	3.09	3.30	.53	.19	.59	.32	.30

The new results of the experiment concern the importance ratings, which suggest a different story. Respondents consistently indicated that they had attached more importance to change than to social comparison in rating satisfaction. The difference was separately significant beyond the .01 level for every domain except physical attractiveness.

For each subject we also computed the difference between importance ratings of social comparison and of change. We then averaged these differences separately across public and across private domains. Consistent with our prediction, the mean difference favoring change was greater for the private domains (averaging .84 for love life, friends) than for the public domains (averaging .42 for grades, attractiveness, housing) ($t(145) = 4.41$, $p<.01$). Thus, Study 5-1 supports the hypothesis that the correlations between global satisfaction and ratings of social comparison do not necessarily reflect the relative importance of the latter variable.

STUDY 5-2

The judgment model of well-being (Schwarz and Strack, 1991) suggests that the reference norm to which people compare their state is labile (see also Kahneman and Miller, 1986). As a consequence, we should expect evaluations of satisfaction to vary with the momentary salience of different standards of comparison, which can be influenced by topics raised earlier in the survey. This idea suggested an additional test of the main hypothesis of this article. We proposed earlier that social comparisons in private domains of life (e.g., love or friendship) are sometimes inferred from (or anchored on) global satisfaction. This heuristic is most likely to be used, we assume, when the salience of global satisfaction is high. Salience can be enhanced, for example, by asking subjects to evaluate satisfaction just before they evaluate social comparison. Thus, we expect an order effect on the correlation between global satisfaction and social comparison, but only in private domains where direct cues for social comparison are lacking.

Questions about satisfaction preceded questions about social comparison and about recent changes in the Emmons-Diener study, as they did in Study 5-1. This sequence is appropriate if the goal is to avoid suggesting to subjects that they use particular constituent judgments in evaluating the global questions. However, if there is a possibility that the specific judgments are affected by the global ones, or by one another, then order must be varied. We therefore conducted a survey using six different forms, representing the six possible orderings of the sections dealing with satisfaction, social comparison, and recent change.

Method. The sample consisted of 125 undergraduates (63 men, 60 women, 2 unreported) registered in an introductory psychology class at San Jose State University. Six survey forms were used, representing all possible orders of the ratings of global satisfaction, social comparison, and change.

Results. We computed two correlations for each pair of measures, grouping together the three forms for which the order of these measures was the same (e.g., satisfaction judgments precede social comparisons for three of the six orderings: satisfaction-social-change, satisfaction-change-social, and change-satisfaction-social). The pairs of correlations are shown in Table 5-2, along with a test of statistical significance for the difference between the correlations.

TABLE 5-2

Correlation of items within domains as a function of question order.

DOMAIN	QUESTION ORDER					
	soc...sat		sat...soc	chg...sat		sat...chg
grades	.614	ns	.600	.566	ns	.386
attractiveness	.690	a	.530	.505	ns	.577
housing	.755	a	.580	.575	ns	.664
friends	.408	b	.760	.538	ns	.473
love life	.555	b	.868	.620	ns	.623
N	62		59	60		61

	soc...chg		chg...soc
grades	.380	ns	.422
attractiveness	.608	ns	.436
housing	.671	ns	.538
friends	.576	ns	.509
love life	.440	a	.669
N	62		63

Our hypothesis predicted an order effect for only two of the fifteen comparisons included in Table III: the correlations between global satisfaction and social comparison for the two private domains (love life and friends). The results are striking: the two correlations for which a difference was predicted are the only ones for which the difference is significant ($p<.01$). These results confirm the idea that global satisfaction provides an optional (not obligatory) heuristic for social comparisons, where more direct information is scarce.

General Discussion

The results of both studies demonstrate the power of a judgmental analysis of measures of well-being, as well as the pitfalls of drawing causal inferences from correlations between these measures. In Study 5-1 we found that the subjective importance that respondents assign to social comparison is (relatively) lowest in the private domains, while the correlation between social comparison and global satisfaction in these domains is notably high. In Study 5-2 we found that this high correlation can be substantially reduced when social comparison is assessed before global satisfaction. This manipulation presumably

Table 5.3 Results of Experiment 5.3

Case#	Intrapersonal Mean	Intrapersonal Var.	Interpersonal Mean	Interpersonal Var.	Unspecified Mean	Unspecified Var.	p	Model Var.	F
1	-1.86	0.48	1.73	0.56	-0.96	3.22	0.75	2.96	0.92
2	-1.05	0.43	1.60	0.61	-0.26	2.93	0.70	1.96	0.67
3	-1.83	0.72	1.45	0.34	-1.43	1.80	0.88	1.81	1.03
4	1.96	0.56	-1.26	0.28	1.30	2.58	0.80	2.20	0.85
5	2.04	0.56	-1.18	0.97	1.78	1.81	0.92	1.40	0.77
6	2.96	0.21	-0.73	0.65	1.87	2.57	0.81	3.28	1.27
7	1.08	0.66	2.76	0.18	1.13	0.85	0.97	0.75	0.88
8	0.78	0.66	2.74	0.19	1.30	1.58	0.73	1.27	0.80
9	0.21	0.58	2.20	0.46	0.87	0.85	0.67	1.45	1.71
10	-2.04	0.32	-0.32	1.04	-1.74	1.38	0.82	0.90	0.65
11	-1.75	0.49	-0.22	0.52	-1.48	0.99	0.82	0.86	0.87
12	0.09	0.25	-2.23	0.54	-0.39	0.88	0.79	1.20	1.36
13	-0.40	0.04	-2.00	0.35	-0.35	0.51	0.84	0.60	1.17

Table 5.4 Results of Experiment 5.4

Case#	Intrapersonal Mean	Intrapersonal Var.	Interpersonal Mean	Interpersonal Var.	Unspecified Mean	Unspecified Var.	p	Model Var.	F
1	-2.04	0.41	1.85	0.13	0.09	2.99	0.45	4.02	1.34
2	-1.00	0.36	1.35	0.33	0.61	2.34	0.32	1.55	0.66
3	-2.00	0.70	1.50	0.36	-0.26	1.93	0.50	3.62	1.88*
4	1.52	0.42	-1.39	0.46	-0.09	1.36	0.45	2.56	1.90*
5	2.13	0.29	-1.32	1.06	0.83	1.78	0.62	3.40	1.90*
6	2.43	0.51	-1.53	0.78	0.22	3.36	0.44	4.56	1.36
7	2.04	0.48	1.36	0.23	1.70	0.67	0.50	0.49	0.72
8	1.67	0.51	1.30	0.41	1.39	0.70	0.25	0.48	0.69
9	0.77	0.81	2.44	0.47	1.52	1.62	0.55	1.38	0.85
10	0.65	0.49	2.67	0.22	1.96	1.13	0.35	1.26	1.11
11	0.30	0.30	2.53	0.25	1.61	1.46	0.41	1.49	0.98
12	-1.68	0.94	-0.44	0.36	-0.91	1.36	0.38	0.97	0.71
13	-1.35	0.23	-0.44	0.25	-1.22	0.45	0.86	0.34	0.76
14	-0.14	0.57	-2.00	0.67	-0.96	1.04	0.56	1.50	1.40
15	0.09	0.08	-2.00	0.56	-1.00	1.54	0.48	1.43	0.93
16	0.91	0.45	-1.87	0.48	-0.61	1.25	0.46	2.42	1.94*

-- Model Var. refers to variance predicted by the probability model, e.g., the variance that would be expected if the choice hypothesis is true..
-- F is composed of the model variance (column 9) over the variance observed in the unspecifie condition (column 7).
-- * indicates significance at the .05 level.
-- Five judgment cases were excluded from experiment 1 and three cases were excluded from experiemnt 2 because p could not be estimated.

reduces the tendency of respondents to rely on global satisfaction as a heuristic for judging social comparison. Ratings of well-being cannot be understood in terms of a simple psychophysical read-out from a well-defined subjective dimension onto a response scale. The alternative view is that the task of assessing one's well-being is a difficult one, and that an answer is produced by opportunistic reliance on cues that are suggested by the question iself, by previous questions in the survey, and by the circumstances of the moment (Schwarz and Strack, 1991).

Multiple Norms (Grant and Kahneman, in preparation)
Paul Grant's project was concerned with people's judgments of behavior in the presence of multiple frames of reference. Norm theory (Kahneman and Miller, 1986) suggests two such frames which can be used to judge an actor's behavior: the first is to locate the person's behavior relative to an interpersonal norm or frame of reference; the second is to locate the person's behavior relative to an intrapersonal norm or frame of reference. Thus, to judge the riskiness of a friend's bet at the track, the interpersonal comparison would pick out the riskiness of her bet relative to the bets of others, while the intrapersonal comparison would pick out the riskiness of this bet with respect to her previous bets. Given these two frames of reference, the question can be asked: if frame of reference is not specified, what form will peoples' judgments of behavior take? Previous research (see Schul & Szyf, 1991) suggests two hypotheses: (1.) People mix the two standards when judging an actor's behavior (Mixture hypothesis), (2.) People choose one of the standards to judge the actor's behavior (Choice hypothesis). In all, four experiments have been conducted exploring these two possibilities. Each will be described in turn.

Experiment 5-3
An experiment was run in which subjects in three conditions made judgments of new behaviors by target actors. Two questions are addressed: (1.) do people have to choose between the standards or do they use both (mixture) in rendering their judgments of behavior? (2) which standard has a more pervasive effect upon judgment?

Method
Seventy-seven University of California undergraduates participated in the experiment in order to fulfill a course requirement. Seven of the subjects did not follow the instructions and were deleted from the statistical analysis.

Stimulus materials consisted of nine examples. Each example centered around a particular activity -- for example, competitive sports, tips after a meal at a restaurant, performance on a math quiz, etc. -- and involved the behavior of three individuals. Three background behaviors and one target behavior were created for each person in each example; all behaviors were expressed in quantitative terms -- batting average, number of sales, etc. The first person's behavior was always high, the third person's behavior was always low, and the second person's behavior was always intermediate; thus, no overlap between the behaviors of the three persons was allowed. Each actor's three behaviors constitute an intrapersonal scale; the aggregate of nine behaviors constitutes the interpersonal scale. Target behaviors were chosen keeping in mind the fact that each behavior takes on simultaneous values on both scales, and that these values are typically different. For example, a behavior that is high interpersonally may well be low intrapersonally. In all, there are nine possibilities for target behaviors.

The placement of target behaviors in examples was balanced with respect to the two scales, given the constraint that person A's target was always high interpersonal, person B's target was always medium interpersonal, and person C's target was always low interpersonal. To insure that the subjects paid attention to all the data presented to them, a preliminary task was developed for each example. Since one has to look at all three of an actor's behaviors to find her middle score, subjects were asked to pick out the median score for each target actor. This task has the added advantage of having subjects pay special attention to the key reference points for both the interpersonal and intrapersonal distributions.

Design. A manipulation of instructions created three groups. Subjects in the intrapersonal condition were instructed to judge target behaviors by comparing to the actor's previous behavior; subjects in the interpersonal condition were instructed to judge the target behaviors by comparing to the previous behavior of the group; subjects in the unspecified condition were not given instructions as to how to judge the target behaviors. Evaluative judgments were made on a seven point semantic differential scale.

Procedure. The instructions informed the subject that a series of examples would be presented, that each example would contain a summary of an activity such as bowling or competitive sales, that behavior of three individuals would be given for each activity, and that two tasks would need to be performed for each example. The middle-value task was presented first and required the subject to locate the middle score (median) in each actor's distribution of behaviors. The second task was termed the judgment task and required the subject to rate a new behavior from each of the three actors. A new behavior was given for each actor and subjects were to rate it by checking the adjective best completing a stem sentence. It is here that the independent variable was implemented, as the stem sentence was varied by condition. If subjects were placed in the unspecified condition the following stem completion appeared:
Alfred's performance on the fourth afternoon was
[] Very Good
[] Good
[] Fairly Good
[] Nothing Special
[] Rather Bad
[] Bad
[] Very Bad

In the interpersonal and intrapersonal conditions the stem completion task was the same as above except that a relative clause was added to the beginning of the sentence. The interpersonal clause was "compared to the scores of the group." The intrapersonal clause was "compared to his (or her) previous performance."

Results. Table 5-3 lists the interpersonal, intrapersonal, and unspecified means and variances for each target judgment case. Also listed is a p-value for each judgment, which is a measure of the relative weighting of the two standards (e.g., an estimate of the probability of an intrapersonal judgment being made in the unspecified condition), and a model variance estimate based on a combination of the means and variances of the interpersonal and intrapersonal groups (e.g., a prediction of what the variance of the unspecified group should be if the choice hypothesis is true). Finally, an F-ratio is listed for each judgment case. This ratio is composed of the model variance over the variance observed in the unspecified group.

The p-values range from a low of .67 to a high of .97, with the average p-value equal to .81. In all cases, the variance of the unspecified group is considerably greater than the variance in either the interpersonal or intrapersonal groups. In general, these data can be interpreted to suggest that people choose between interpersonal and intrapersonal standards when judging another's behavior. In four of the cases they used the intrapersonal standard outright, rejecting interpersonal comparison completely. In the other twelve, 80% judged intrapersonally and 20% judged interpersonally.

Experiment 5-4

The purpose of the next study was to determine the influence of the mid-value orienting task utilized in the first study. It is possible that this task may have encouraged the predominant use of the intrapersonal standard in subjects' judgments of behavior. To see if this was the case, a new orienting task was developed. In this task, subjects were asked to order all nine scores in each example form highest to lowest and to write down the second, fifth, and eighth highest ones. Notice that subjects write down the

exact same scores in this new "2,5,8 task" as they would in the mid-value task (this is due to the fact that the three distributions in each example do not overlap). By focusing subjects' attention on all nine scores, this new task should have the effect of emphasizing the interpersonal frame of reference more than the intrapersonal frame of reference. Thus, if the orienting task is influencing subsequent judgments of behavior, then judgments following the 2,5,8 task should have lower p-values than judgments following the mid-value task. Conversely, p-values should remain the same if the orienting task has no influence.

Method

Subjects. Sixty-nine University of California undergraduates participated in the experiment in order to fulfill a course requirement. Subjects were run in several sessions.

Materials, Design, and Procedure. Everything was the same as in experiment one except for the new orienting task. At the top of each example subjects were instructed as follows:
2nd, 5th, 8th Task.

Ordering all nine from highest to lowest,please list the 2nd, 5th, and 8th highest scores:

2nd_____ 5th_____ 8th_____

Results. Table 5-4 shows that the p-values have indeed come down. In Experiment 5-4, p ranges from .25 to .86, with the average p being .48. Thus, subjects clearly judged more interpersonally in the present study than in Experiment 1. However, the effect of the 2,5,8 task seems to be less pronounced than the mid-value task, as p averages about .5. P would have had to average .25 to match the .75 effect of the mid-value task. Table 5-4 also reveals evidence that subjects mixed the two frames of reference. Indeed, in 4 of the 18 cases F reaches significance and allows for a rejection of the choice model.

In sum, Experiment 5-4 suggests that the mid-value task biases subjects' subsequent judgments toward the intrapersonal frame of reference. Moreover, the alternative 2,5,8 task produces less of a bias, even though subjects search for the same scores as in the mid-value task. In addition, the presence of judgments that combine the two frames of reference suggests the following hypothesis: The orienting task activates, or primes, one of the frames of reference (mid-value primes intrapersonal; 2,5,8 primes interpersonal); however, regardless of task, attributing a score to an individual activates the intrapersonal frame of reference. Thus, when the mid-value task is used, very little consideration of the interpersonal standard will be seen, since it has not become activated. This account does not, of course, explain why 10 - 20% of the subjects in experiment one judged interpersonally.

Experiment 5-5

The purpose of Experiment 5-5 was to test the interpretation of the interpersonal instructions. It seems possible that subjects might take interpersonal information into account when making this judgment, even though they have been explicitly instructed to judge intrapersonally. Experiment 3 tests this possibility by introducing a manipulation of the interpersonal scale. If interpersonal information is covertly influencing overt intrapersonal judgments, then it should make a difference where in the interpersonal distribution the target actor appears. That is, the same target behavior should be rated differently if the actor is at the top of the interpersonal scale than if he is in the middle, since an intrapersonally poor behavior will be interpersonally fair if he is at the top of the distribution, but interpersonally poor if he is in the middle. Two versions of the intrapersonal questionnaire were devised, such that for each example the background and target behaviors for two of the actors were the same between forms, and one actor was different between forms. The different actor was either higher or lower interpersonally than the other two. The point was to see if a target behavior is rated the same when the actor is interpersonally the best of the three (designated actor A), as when he is interpersonally in the middle (designated actor B).

Subjects. 50 University of California undergraduates participated in the study as a part of a course requirement. All subjects were run in individual sessions.

Materials, Design, and Procedure. The materials were as in the previous two studies. In each example, the original background and target behaviors were compressed slightly to make room for a fourth actor's behaviors This was done so as not to extend the range absurdly in several of the examples (for example, a baseball average of .140).

Results. Without question the results do not support the hypothesis of interpersonal pollution. The means of subjects' ratings of target actors common across the two conditions were subjected to t tests. Of the eighteen, only one achieved significance at .05 level (the critical value is t =1.69).

Experiment 5-6

Experiment 5-6 tested the idea that reversing the judgment task of the first experiment might lead to more mixing of the frames of reference. Just as interpersonal and intrapersonal norms can be used as judgment standards, they can also be used to generate new behaviors given an evaluative description. So, if I am told that Bill shot a "good" round of golf, I can generate what his score must have been to deserve that description.

56 paid subjects participated as a part of a series of unrelated experiments which were run together.

Again, the same 9 examples were utilized from experiment 1. The 2,5,8 orienting task was used in place of the mid-value task, because it seems to have a less biasing effect on subsequent judgments. The background behaviors were the same as in Experiments 1 and 2. In place of target behaviors were evaluative descriptions of behavior on a fourth occasion. These descriptions were chosen to match the target behaviors that were used in Experiments 1 and 2. As in Experiments 1 and 2, three groups of subjects were created -- interpersonal, intrapersonal, and unspecified groups. Subjects in the unspecified condition were given the following judgment task:
Alfred's performance in the fourth game was Nothing Special. He must have shot a score of _____ .

Subjects in the intrapersonal and interpersonal conditions were given the following judgment task with a relative clause added to the beginning of the sentence: "compared to his previous scores," and "compared to the scores of the group," respectively.

Results. The numerical results of Experiment 4 were subjected to the same probability model as the ratings of Experiments 1 and 2. In the High/Low and Low/High cases, p ranges between .59 and .87., with the average p across the six cases being .79. These results look more like experiment 1 than 2. Thus, in the reverse task, people appear to be choosing between frames of reference.

6. MENTAL CONTAMINATION

Literature Review and Theoretical Analysis (Kahneman and Varey)

Several sources of evidence suggest that intentional control of mental processes is not always as easy as it may appear. In fact, the intention to perform a particular mental operation commonly activates other operations in addition to the specifically intended one. The proliferation of such unintended computations creates a problem of control that is often manifested in slowed responses, in contaminated responses, or in outright errors.

Together with Carol Varey, I am currently engaged in a review of contamination effects in the cognitive and social psychology literatures. We distinguish between two broad categories of effects arising from unintended computations. When responses are made along an ordered scale, the contaminated response reflects a ***compromise*** between answers arising from the intended and the unintended processes. In

these situations the ***outcome*** of the intended process is affected by unintended processing. When the response is a categorical choice, the results of the unintended process provide either ***conflict*** with, or ***support*** for, the result of the intended process, and crosstalk produces ***delayed or speeded responses***, or errors.

A prototypical example of compromise effects is the phenomenon of anchoring in judgment: the processing of the anchor as a suggested solution to a problem typically leads to a response that is pulled toward the irrelevant and uninformative value. The Stroop effect is a paradigmatic illustration of conflict effects due to an unnecessary mental operation. In the Stroop task, subjects are asked to name the ink-color that a word is written in. Subjects are slower to name the ink-color when the written word is itself a conflicting color word. This effect is not simply a reduced efficiency resulting from performing two processes at once since different words have different effects. The color naming process is slowed down relative to reading a neutral word. And, in fact, a congruent color word results in faster color naming.

Our review explores these and other contamination effects in depth, addressing cognitive variants of Stroop effects, such as the confusions between metaphorical and literal truth, and between truth and validity, as well as manifestations of 'unintended thought' in social perception. In the last year and a half, the grant has supported several experimental research programs in contamination. Karen Jacowitz and I conducted a large study of anchoring effects in judgment; Carol Varey wrote her dissertation on a new source of crosstalk effects; with Anne Treisman and Maria Stone I began a new line of studies on crosstalk between concurrent relational tasks. Further research on crosstalk effects is planned for the extension period.

Crosstalk and Contamination in Cognitive Processes -- Dissertation research by Carol Varey

This dissertation investigated the problem of the control of cognitive operations. If a person wishes to perform an operation, ***A***, how effectively can she prevent herself from performing operation ***B*** in addition to, or instead, of ***A***? What operations are likely to be performed inadvertently, and why?

The Introduction reviewed several examples in the psychological literature that show that the result of an unintended process can have important consequences on the intended process. The term ***crosstalk*** refers to the response timing effects and errors that arise from conflict (or collaboration) between intended and unintended processes. A Theoretical Framework section considered these crosstalk effects in the light of three possible sources for unintended operations: habitual cognitive operations, recently-performed operations, and concurrent operations.

This theoretical framework for conceptualizing crosstalk suggested the possibility of effects not previously investigated in the literature. Two such effects, called computational momentum and stimulus inertia, were investigated in a series of four experiments. The first effect, ***computational momentum***, is the tendency for people to continue to perform a mental operation after it is no longer relevant. Thus, tasks that were intended only to be performed on earlier stimuli are also performed on currently-relevant stimuli, creating crosstalk with the currently relevant task. The second effect, ***stimulus inertia***, reflects the tendency to perform the current operation upon memory traces of stimuli that were processed earlier.

The investigation of computational momentum and stimulus inertia requires an experimental paradigm in which the subjects' task changes frequently. Effects of computational momentum are shown when performance on the intended operation is affected by the answer to the previous operation applied to the current stimulus. Such effects may be evinced by slowed or speeded responses dependent upon the irrelevant answer, or by changes in error rate dependent upon the irrelevant answer. Similarly, effects of

stimulus inertia are shown when performance (speed or accuracy) on the intended operation is affected by the answer to the current operation applied to a previous stimulus. Two paradigms allowing frequent changes of task were used: feature verification and "same"-"different" judgments.

Experiments 6-1 and 6-2 used a feature-verification paradigm. Subjects were presented with simple visual displays such as three red triangles at the top of the terminal screen, or two blue squares at the left of the screen. In any single display the elements all shared the same color and shape, they were all in the same quadrant on the screen, and there were two, three, four or five elements. Each display was defined by a conjunction of four features (color of elements, shape of elements, number of elements, and screen position of display), with each feature chosen from a set of four possible values. Subjects were presented with a question probing a particular feature value, for example "Blue?" to which they responded by hitting the key marked "Y" for Yes, or the key marked "N" for No. In Experiment 1, subjects performed the same task for five displays, after which a new question appeared and was in turn applied to five displays, and so on. In Experiment 2, a new question appeared with each display.

An illustration will serve to explain how crosstalk effects can be examined in this paradigm. Suppose that the subject intends to answer the question "Blue?", and that her previous question was "Triangle?" Computational momentum is evinced by differences in the response to "Blue" depending on whether or not the current display shows triangles. Stimulus inertia, in contrast, is shown by differences in the response to "Blue?" according to whether or not the previous display (the target of the "Triangle?" question) was blue or not.

In Experiment 6-1, there were clear effects of conflict between the computational momentum (CM) answer and the answer to the current (intended) question. These effects were present in both RT and error rates. As predicted, these effects were strongest for the first and second displays following a new question, as shown in Table 6-1:

Table 6-1. Effects of computational momentum on RT for each display in Experiment 6-1 (n=22).

	Correct answer	CM answer: No	CM answer: Yes
display 1			
	No	653	667
	Yes	637	594
display 2			
	No	485	493
	Yes	461	451
display 3			
	No	490	496
	Yes	445	450
display 4			
	No	484	490
	Yes	451	446
display 5			
	No	496	496
	Yes	459	446

The answer to the irrelevant stimulus inertia (SI) question also had effects on RT and error rates, although in this case responses to the current question were faster and more accurate when the SI answer was yes, irrespective of the current answer (see Table 6-2). Although subjects may have computed the irrelevant stimulus inertia answer, an alternative explanation for this result is that when a feature appears in a display it semantically primes the related probe, thus facilitating responses to it.

Table 6-2. Effects of stimulus inertia on RT for display 1 Experiment 6-1 (n=22).

		SI answer	
		No	Yes
Current answer	No	659	638
	Yes	628	616

The computational momentum and stimulus inertia effects were markedly larger than the effects of the previous response (see Table 6-3). Also, the faster responses when the previous response was compatible were obtained at the cost of greater errors. In other experiments compatibility with the previous response has been found to influence RT. However, the paradigm of varying questions allows the effects of the previous response response to be unconfounded from the effects of the previous question. It appears that repeating the question may be a more important factor in "response-priming" effects.

Table 6-3. Effects of previous answer on RT for display 1 Experiment 6-1 (n=22).

		Previous answer	
		No	Yes
Current answer	No	658	662
	Yes	620	611

The CM effects in Experiment 1 may have occurred because the questions remained relevant for five trials, or because the question had to be committed to memory. In Experiment 6-2, these explanations were tested by presenting the question simultaneously with the relevant display, thus eliminating the memory requirement, and changing the question with each display, thus eliminating any benefits to be derived from a processing habit developed over displays. Again, compatibility effects of computational momentum were observed (see Table 2.4).

Table 6-4. Effects of computational momentum on RT, Experiment 6-2 (n=18).

		CM answer	
		No	Yes
Current answer	No	878	894
	Yes	852	840

The response to the stimulus inertia question also had an effect on RT, but in this experiment responses were faster and more accurate when the answer to the stimulus inertia question was No (see Table 2.5).

Table 6-5. Effects of stimulus inertia on RT, Experiment 6-2 (n=18).

		SI answer: No	SI answer: Yes
Current answer	No	877	896
	Yes	848	852

The remaining experiments used a "Same"-"Different" paradigm to investigate computational momentum. In Experiments 3a and 3b, subjects were first shown one of the questions "Same Color?", "Same Shape?", or "Same Number?". Then they were presented simultaneously with two simple visual displays, one on the left of the screen and one on the right (for example two green crosses on the left, and four white circles on the right). If the displays matched on the probed dimension, subjects responded by pressing a key marked "S" for Same. Otherwise they responded with "D" for Different. As in Experiment 1, subjects responded to five displays for each question. In this paradigm, evidence for computational momentum is shown by an effect of the CM answer (say, shape same or different) on the current answer (say, color same or different). Table 6-6 shows that CM effects are large and appear to be maintained across all five displays.

Table 6-6. Effects of computational momentum on RT for each display, Experiment 6-3a (n=20).

		CM answer: Diff	CM answer: Same	
relevant similarity	Diff	783	836	stim 1
	Same	709	686	
relevant similarity	Diff	609	605	stim 2
	Same	588	552	
relevant similarity	Diff	602	620	stim 3
	Same	567	550	
relevant similarity	Diff	608	622	stim 4
	Same	567	539	
relevant similarity	Diff	629	646	stim 5
	Same	590	566	

It was necessary to test whether these results were due to computational momentum, or were an artifact arising from a tendency for subjects to process ***all*** similarity dimensions, regardless of whether the dimension was recently probed. This was investigated in Experiment 3a by comparing the effects of irrelevant shape similarity for cases in which shape was the previously-probed dimension, with cases in which it was not. In Experiment 3b only the color and number probes were used. This allows us to see whether there is any effect of crosstalk from a dimension that is never probed. As table 6-7 shows, the compatibility effects of irrelevant shape similarity are much larger when shape was the previous question (i.e. shape is the CM dimension).

Table 6-7. Effects of irrelevant shape answer on RTs in Experiments 6-3a and 3b.
Columns (1) and (2) are from Experiment 6-3a (n = 20); Column (3) is from Experiment 6-3b (n = 20).

		(1) irrelevant shape is CM dimension		(2) irrelevant shape is not CM dimension		(3) irrelevant shape is never probed	
		Shape Diff	Shape Same	Shape Diff	Shape Same	Shape Diff	Shape Same
Color relevant:							
Color	Diff	566	591	569	620	612	628
Color	Same	513	502	515	516	548	559
Number relevant:							
Number	Diff	702	749	696	696	738	722
Number	Same	696	608	669	611	723	666
means:							
	Diff	634	670	633	658	675	675
	Same	604	555	592	563	636	613

Experiment 4 extended the feature version of the "Same"-"Different" paradigm to investigate cross-modal crosstalk. Subjects were given "Same Tone?" or "Same Color" as a probe, then the first color was presented accompanied by a tone, followed by the second color-tone pair. As in Experiment 3a, computational momentum was examined as a possible modifier of concurrent crosstalk effects. Results showed that the effects of irrelevant similarity were much larger when the irrelevant dimension was probed in the previous question (see Table 2.8). Again, conflict with the computational momentum answer led to slower responses than responses supported by the computational momentum answer.

In summary, all the experiments showed that the result of the computational momentum process affected the speed and accuracy of responses to the relevant question. The effect was observed in both feature-verification and "same"-"different" paradigms. Crosstalk occurred when the CM question probed a different modality from the currently-relevant question, as well as when both questions referred to a visual dimension. Experiment 6-2 showed that computational momentum effects do not appear solely as a result of a set of repeated applications of a particular operation, since a single trial will suffice. Nor is committing the task to memory prior to the relevant trials a necessary condition for computational momentum, since the effect is still evident when the task and the stimulus are displayed together. Thus it appears that even after a single execution of a task people have a tendency to repeat the same operation, and the results of the unnecessary operation contaminate the intended process. Future research is planned to investigate these effects further.

Table 6-8. Effects of irrelevant-modality answer on RT across all displays, Experiment 4 (n=19).

	(1) Other dimension probed in previous trial		(2) Same dimension probed in previous trial	
	irrelevant answer		irrelevant answer	
relevant dimension	Diff	Same	Diff	Same
Tone:				
Tone Diff	382	425	400	413
Tone Same	400	369	377	361
Color:				
Color Diff	376	358	356	344
Color Same	325	318	338	302

7. UNINTENDED COMPARISONS

Kahneman, Treisman and Stone

We have started research on the influence of unintended comparisons of irrelevant objects on subjects' ability to carry out a comparative task. In a paradigm we devised, subjects are presented with 4 objects on the screen. Their task is to compare the rightmost and the leftmost object and while disregarding the two middle objects. Several experiments in this general framework were conducted, and final results are available for most of them. We observed Stroop-like interference from the outcomes of operations performed on irrelevant stimuli in some cases, but not in others. This technique allows us to study the natural relationships that exist between the various types of comparisons.

Experiment 7-1 and 7-2

In experiment 7-1, subjects were presented with four items on the screen. The objects were two vertical lines and two digits flanked by two oblique lines. Subjects' task was to compare the oblique lines and press a key to indicate the shorter line (left is the left was shorter, and right if the right was shorter). Three conditions were possible for each type of the interfering stimuli (digits or vertical lines). If the relationship between the two middle objects made no difference, then the "compatible" condition would be no different from the "incompatible condition for both digits and lines. In experiment 7-2, the noise (interference) stimuli were the same as in Experiment 7-1, but the outside stimuli were digits, and the task was to press a key for the smaller of the two outside digits. Eighteen subjects participated in both experiments, the order of the experiments was counterbalanced. The results are presented in the following tables:

Table 7-1: Effects of irrelevant line and digit stimuli on subjects' performance in line comparison task (n=18)

Interfering stimuli---lines

	compatible	incompatible	control
reaction times	735	765	704
error rates	7.9	11.2	8.4

Interfering stimuli---digits

	compatible	incompatible	control
reaction times	688	741	721
error rates	6.0	8.8	8.9

Table 7-2: Effects of the irrelevant line and digit stimuli on subjects' performance in digit comparison task (n=18)

Interfering stimuli---lines

	compatible	incompatible	control
reaction times	482	505	488
error rates	1.9	3.5	2.2

interfering stimuli---digits

	compatible	incompatible	control
reaction times	521	528	515
error rates	4.9	5.6	4.5

In Experiment 7-1 (line comparison) irrelevant lines and irrelevant digits both cause interference. In Experiment 7-2 (digit comparison), only lines cause interference. The interaction of task x type of interfering stimuli is significant ($t(17)=2.0$, $p<0.05$)

What would be the reason for this interaction? One possibility may be that for the stimuli of the same type (digits or lines) some form of effortful selection has to occur before a comparison is made (that is, subjects need to decide which digits are relevant to the comparison or which lines are). Once such selection occurs, the probability of further processing of the irrelevant items is diminished. Stronger response interference will occur when this type of selection is not needed (as when the task involves digits, but the interfering stimuli are lines or when the task involves lines, but the interfering stimuli are digits), since the irrelevant items are likely to be processed further before they are actively suppressed.

Experiment 7-3

Experiment 7-3 was run to demonstrate that subjects did carry out a comparison of the middle digits in the Experiment 7-1. It could be argued that the effect observed in that experiment was produced by interference/facilitation from single digits, rather then from pairs of digits. That is, whenever subjects saw a small digit (1, 2 or 3) next to the short line, they were more likely to respond to it, regardless of what the other digit was. To rule out this possibility, we ran a control experiment in which a digit and a letter appeared in the middle, flanked by two oblique lines. Subjects' task was still to respond to the shorter of the two lines. If subjects were influenced by the absolute values of the digits, we may expect a correlation between reaction time and the value of the digit (1 to 9) that appears on the same side as the shorter line.

Table 7-3: Effects of a single irrelevant digit appearing on the same side as the shorter line on line length comparison task (n=12)

digit	react.time
1	813
2	777
3	815
4	784
5	820
6	840
7	803
8	812
9	779

No correlation is found. A comparison of reaction times for low numbers (1-3) and for high numbers (7-9) yields an insignificant $t(11) = 0.61$.

Experiment 4

In the next experiment, we decided to explore the natural similarities that might exist between digit and letter comparisons. Subjects were presented with four objects on the screen. The outside objects were always letters, and the two middle objects were always digits. The subjects' task was to decide which letter appears earlier in the alphabet. Our goal was to find out if this operation of comparison is similar to deciding which digit is numerically lower.

Table 7-4: Effects of irrelevant digit stimuli on subjects' performance in letter comparison task

interfering stimuli---digits (n=12)

	compatible	incompatible	control
reaction times	922	933	918
error rates	6.5	5.1	5.6

None of the differences between the conditions in this experiment are significant. It appears that in general, the task of deciding which letter appears earlier in the alphabet does not activate the unwanted comparison of digits. However, we suspected that comparisons invovling immediately successive letters could be different. To test this hypothesis, we ran an experiment in which the letters relevant to the task were always sequential, and the interfering digits were only sometimes sequential. The task was still to decide which letter appears earlier in the alphabet, and the display was identical to the previous experiment.

Table 7-5: Effects of irrelevant digit stimuli on subjects' performance on letter task with sequential letters. Letters are sequential; interfering stimuli are sequential or non-sequential digits

condition	reaction time
non sequential, compatible	924
non sequential, incompatible	913
sequential, compatible	903
sequential, incompatible	917
control	944

The difference between the compatible and the incompatible conditions was not significant for either sequential or non-sequential digits. However, an interesting effect was observed. Substantial and significant interference was observed in the control condition, in which the two noise items were identical. Thus, detection of identity appears to share coding with detection of sequence.

Experiment 7-6

In Experiment 7-6, subjects were presented with two three-letter abbreviations for the months of the year separated by two digits in the middle. (For example, JUN 5 7 JAN). The task was to press a key for the month that appears earlier in the calendar year and to ignore the digits. Same types of conditions as in the previous experiments were used.

Table 7-6: Effects of irrelevant number stimuli (1 to 12) on subjects' performance in months comparison task (which month comes earlier in the year)

interfering stimuli---numbers 1 trough 12 (n=6)

	compatible	incompatible	control
reaction times	1004	1033	1015
error rates	6.4	6.4	5.4

Even with this small sample size, the difference between compatible and incompatible conditions approaches significance $t(5)=2.48$. It appears that in this case, the outcome of the digit comparison did interfere with subjects' performance in the task. That is, sequence of months is encoded on a way similar to the sequence of numbers.

Experiment 7-7

In this experiment, we were interested if semantic judgments of size would be interfered with by either digit or line stimuli. Subjects were presented with two animal names flanking either two digits or two lines (for example, bunny 2 7 roach). Animal names were restricted to 7-5 letters in length, and the difference in size between animals was non- disputable. The following animal names were used: flea, roach, snail, mouse, bunny, sheep, horse, rhino, whale. Subjects' task was to press a key indicating the smaller of the two animals. They were instructed to use average size for each animal in comparison. The results are presented in the following table:

Table 7-7: Effects of irrelevant line and digit stimuli on subjects' performance in animal comparison task (which of the animals is physically smaller). (n=16)

Interfering stimuli---lines

	compatible	incompatible	control
reaction times	927	946	948
error rates	5.0	6.0	5.2

Interfering stimuli---digits

	compatible	incompatible	control
reaction times	944	965	944
error rates	5.4	6.7	4.6

A nonsignificant trend in the right direction is present in this experiment. Neither digits nor lines interfere strongly with subjects' ability to make size comparisons. It is not clear why we failed to obtain clear indications of interference in this experiment, which resembles experiment 7-5.

8. Topic and Referent in Perceptual Comparisons
Dissertation research conducted by Maria Stone

Human thought is selective. This claim is not controversial as long as the thought involves only one object to the exclusion of others. Picking out a single figure from a background or concentrating on a specific object or person in order to retrieve their characteristics from memory are such uncontroversial cases. If linguistic description is warranted, the subject of the sentence will frequently correspond to this selected "topic" of thought.

However, there are many situations in which human thought appears to be about not just one, but exactly two objects and a relationship between them. One example is comparison. In language, different roles are assigned to the two objects involved. One of them becomes the subject (topic) of a sentence, and the other becomes the object, or referent. What is the cognitive significance of this assignment of roles? One possibility is that the thought is about the relationship and/or difference between the objects, and that the assignment of roles arises only when the thought is processed for communication. The other is that the thought is not about the difference, but about one of the objects and its relationship to the other object. In this case, the distinction between the topic and the referent is cognitive as well as linguistic. This research explores the cognitive consequences of directional comparisons.

Maria Stone's previous research examined how the topic can be designated in linguistically neutral comparisons. The experiments described in an earlier report explored the link between attention and the selection of the topic of comparison. This year, the focus of research was on distinguishing the kind of processing the topic and the referent receive in perceptual comparisons. Two aspects of this distinction have been proposed.
1. The topic is said to "control the agenda" for comparison; e.g., the features of the topic get mapped onto the features of the referent, but not vice versa. This should have several empirical consequences. (a). When the topic has more unique features than the referent, it appears more different from the referent than when the referent has more unique features than the topic. This asymmetry was studied by Tversky (1977) and Agostinelli et al. (1986). It was also utilized in the six experiments described in a previous report, which studied the factors that determine the topic of comparison.

(b). For some stimuli, there is a specific natural order in which the features of an item are encoded (eg., letters in words). When two such items are compared directionally, the order in which the features will be checked off should correspond to the order of the features in the topic item.
(c) If the common features group together (due to proximity or similarity) in the topic, but not in the referent, finding them should be easier than when they group together in the referent, but not in the topic.
2. In the process of comparison, the topic is encoded relatively, whereas the referent is encoded absolutely. The results of this encoding should be noticeable when:
(a). The topic or the referent are repeated in a new comparison.
(b). In the memory for the topic and for the referent.

Overview of the new experiments:
A). Demonstrating that the topic "controls the agenda" of comparison:

Several experiments were conducted to demonstrate that the order in which the features of the two objects are compared is determined by the order of features in the topic object. Five-letter nonsense strings of consonants were used. One of the strings was designated as the topic of comparison using some of the manipulations that were effective in the previously reported experiments. The subjects' task was to write down the letters that the strings had in common. The strings were randomly generated, and always had three letters in common and two unique letters each. The order in which the common letters appeared in the two strings was randomly determined, and was often (but not always) different. Subjects were expected to report the common letters in the order in which they appear in the topic string.

In the first experiment, the first string was presented for 2000 msec., then a mask of "XXXX" was presented for 170 msec, then a long interval (1000 msec), and, finally, the second string was presented for 2000 msec. The results of previous experiments suggest that the first string should become the topic of comparison in this situation, i.e., the subjects will report the common letters in the order in which they appear in the first string. The results confirm this prediction--subjects were more likely to report the common letters in the order in which they appear in the first string than in the order in which they appear in the second string. The entire experiment consisted of 20 trials, and on average, on 8.2 trials the order of the reported letters was consistent with the order of common letters in the first string, compared with only 4.3 trials for the order consistent with the second string.

A second manipulation was designed to assign the role of topic to the item shown last on a trial. Two strings were shown on each trial, one in capitals and one in lower case. The strings remained on the screen for the duration of the trial. A third string, added 2000 msec later, could be either in capital or in small letters. The subjects' task was to compare the two strings in the same case. Previous results suggested that in this situation the third string would be the topic of comparison. As before, the hypothesis is that the order in which the common letters appear in the report should correspond to their positions in the topic string. This prediction was confirmed. This experiment also consisted of 20 trials, and the order of reported letters was consistent with the order of the common letters in the last string on 7.3 trials, compared with 3.4 trials for the order consistent with the string presented earlier, (n=12).

In a third experiment, only one string appeared initially on the screen, followed 2000 msec later by another string. The two strings remained on the screen together for another 1000 msec. The order of the reported letters was consistent with the order in the first letters on 4.9 trials, and with the order of letters in the second string on 4.8 trials (n=36). It appears that in this experiment, subjects were not consistently selecting the same string as the topic.

One problem with this paradigm is that the task is very difficult, and performance therefore strategic, rather than spontaneous and automatic. Exposure parameters had to be adjusted to allow adequate performance, which also meant that the strings stayed on the screen long enough to allow multiple eye

movements, and possibly several checks and rechecks of each string. The obtained results may be due to subjects' strategies, rather than to the spontaneous allocation of the role of a topic to one of the objects. New experiments are planned that will use three-letter nonsense strings with only two letters in common, thus making the task easier. The timing parameters will be changed to speed up the presentation. Both the hypothesis about the order in which the features are compared (b) and the hypothesis about the role of grouping (c) will be tested, using the new stimuli.

B). Demonstrating that the topic is encoded relative to the referent, and that the referent is not encoded on the same way.

The present analysis implies a difference between the coding that the topic and the referent are assigned as the result of their comparison. The topic is assumed to be encoded relative to the referent, whereas the referent is encoded absolutely. A new paradigm was designed to demonstrate this. On each trial, subjects were presented with two letters or two digits. One of the items was flashing, and thereby designated as topic. Subjects had to decide whether the flashing item was smaller (for digits) or earlier in the alphabet (for letters). On some trials, either the flashing or the stationary item was repeated from the previous trial. The item could be associated with the same response as on the previous trial, or with the opposite response. Since the topic (flashing item) is encoded relatively, its repetition with the repeated response should be significantly faster than its repetition with the opposite response. Since the referent (stationary item) is encoded absolutely, there shoul be no difference between repeating the referent with the same or with a different response. Results are presented in the following two tables.

Table 8-1: Effects of stimulus and response repetition in the letter comparison experiment. (mean response times for each condition (n=15))

type of perceptual repetition

response	none	top-top	ref-ref	ref-top	top-ref
same	1066	1024	1101	1186	1002
diff	1075	1183	1043	1133	1048

Table 8-2: Effects of stimulus and response repetition in the digit comparison experiment.
mean response times for each condition (n=17)

type of perceptual repetition

response	none	top-top	ref-ref	ref-top	top-ref
same	763	751	762	802	774
diff	794	841	802	774	793

No general benefit of perceptual repetition was observed for either letters or digits. In fact, conditions with no perceptual repetition were faster both for digits ($t(16)=2.93$, $p < 0.01$) and for letters($t(14)=1.99$, $p < 0.10$). For digits, but not for letters, a small benefit of response repetition was present ($t(16)=2.95$, $p < 0.01$). In both experiments, subjects are slower when the topic (flashing) item is repeated with a

new response than when the topic (flashing) item is repeated with the old (repeated) response ($t(16)=4.5$, $p < 0.005$ for digits; $t(14)=2.83$, $p < 0.01$ for letters). The effect of repeating the topic is smaller (for digits) or apparently absent (for letters). The difference between the effects of repeating topic or referent is significant both for digits ($t(16)=2.44$, $p < 0.025$) and for letters ($t(14)=3.74$, $p < 0.005$)

The results so far support the hypothesis that the topic is encoded relatively (as being smaller or larger, earlier or later in the alphabet), whereas the stationary (referent) item is not encoded in this fashion. When the relative codes assigned to a topic on two successive trials are in conflict, interference occurs. Since the referent is not encoded relatively, no interference is observed when a new response is paired with a repeated referent.

C. The coding of topic and referent. The hypothesis that emerges from earlier work is that the topic is encoded relative to the referent, whereas the referent is encoded absolutely. A new paradigm was designed to demonstrate this. On each trial, subjects were presented with two letters or two digits. One of the items was flashing, and the other was displayed continuously. In the letter experiment, subjects were asked to decide whether the flashing letter was earlier or later in the alphabet than the other letter. In the digit experiment, subjects were asked to decide whether the flashing digit was numerically smaller or larger than the remaining digit. On some trials, either the flashing or the stationary item was repeated from the previous trial. The item could be repeated with the same response as it appeared on the previous trial, or with the opposite response. If the topic (flashing item) is encoded relatively, its repetition with the repeated response should be significantly faster than its repetition with the opposite response. If the referent (stationary item) is encoded absolutely, there should be no difference between repeating the referent with the repeated response and repeating it with the opposite response. The results confirmed this prediction.

Is there a general tendency to respond to the topic and not to respond to the referent or do the effects observed in earlier experiments occur at the level of the specific response? In the experiment summarized in Table 8-3 comparison and naming trials alternated. On comparison trials, subjects saw two letters, one above the other. One letter was flashing. Subjects were instructed to press one key if the flashing letter was earlier in the alphabet, and another if it was later. This type of trial was always followed by a naming trial. Subjects were presented with the red and the green letter, and asked to name the red letter and disregard the green letter. The red or the green letter or both could be repeated from the previous (comparison) trial. Results of this experiment are presented in Table 3.

Table 8-3: Effects of stimulus repetition in letter comparison/letter naming
experiment.
mean reaction times (n=10)

	type of stimulus repetition					
none	top-red	top-grn	ref-red	ref-grn	tr-rg	tr-gr
669	638	682	643	648	636	692

Difference between baseline and stimulus repetition conditions in letter comparison/naming experiment.

type of stimulus repetition					
top-red	top-grn	ref-red	ref-grn	tr-rg	tr-gr
+31*	-14?	+26*	+20?	+32*	-23*

* significant differences ? close to significant differences

The amount of negative or positive priming observed in each experimental condition is represented in the bottom portion of Table 3. Repeating the topic as an item to be responded to (red) results in facilitation (31 msec., $t(9)=3.04$, $p<0.05$), and repeating it as an item to be ignored (green) results in small and so far nonsignificant amount of inhibition (-14 msec., $t(9)=-1.08$). The results look quite different for the referent. Repeating it as either green or red produces some faciliation (26 msec for red, $t(9)=2.55$, $p<0.05$; and 20 msec for green $t(9)=1.9$, $p<0.10$). Repeating topic as red and referent as green does not produce any more facilitation than simly repeating topic as red (32 msec., $t(9)=3.7$). A condition in which the referent becomes red and the topic becomes green shows inhibition (-23.4 msec., $t(9)=-2.86$, $p<0.05$). It appears that naming responses are facilitated both for the topic and for the referent. In addition, ignoring to topic is difficult, while ignoring the referent is easy. The topic appears to be generally selected for response, while the only response to the referent that is inhibited is the specific response required in the comparison.

9. ANCHORING EFFECTS

Kahneman and Jacowitz

The phenomenon of anchoring occurs when some initial value exists that a subject uses as a starting point for determining a response to a stimulus. Most often in the research to date, the anchor value has been a number that appears somewhere in the question or in the introduction or instructions. Then, subjects can adjust this value in the direction that they feel is appropriate in order to generate their actual response. In general, researchers have found that subjects do not make sufficient adjustments, so their final judgment is "anchored" to the initial value.

Many researchers have studied anchoring effects on judgment tasks and those factors that make them more or less likely to occur. Markovsky (1988) proposes three conditions for anchoring to occur: 1) the judgment is indeterminate, 2) an anchor exists, and 3) the anchor is salient. In addition, a potential anchor is more likely to be used as such if it is in a format that is compatible with the response scale.

A factor that might reduce anchoring effects is the degree of knowledge that subjects have about a topic and their confidence in their judgments. Although this has been suggested (e.g. Plous, 1989), no empirical support has demonstrated that susceptibility to anchoring is inversely related to confidence. In this study, we tried to provide direct empirical support for this relationship.

In order to test whether high confidence reduces anchoring effects, we needed to have a method for measuring anchoring. There are certain logical constraints on how to measure anchoring. For instance, at least two different anchors are needed for each question, as well as an unanchored group in order to compare the distributions of responses with and without anchors. The second purpose of this research is to provide an index that represents a measurement of the amount of anchoring in the responses to numerical judgments. The index value is determined by finding the difference between the means of groups exposed to high and low anchors. This difference is then divided by the difference between the anchor values. The index represents a measurement of the amount of motion toward the anchor values. For example, if the difference between the means is the same as the difference between the anchor values, that would indicate perfect anchoring and the index value would be one. If there is no difference between the means of the high and low anchor groups, then apparently the different anchors had no effect. In such a case, the index will equal zero which means that no anchoring has occurred. As the difference between the means increases, the high and low anchors are having more of an effect on the distributions. As a result, the index value will increase.

In order to be able to determine what would be appropriate high and low anchor values, we first obtained a distribution of unanchored responses to each of our 15 questions. The anchors that we used for the experimental groups were the 15th and 85th percentile responses from the unanchored

distribution. Because the subjects in the pretest and experimental groups were taken from the same population, we would expect the distributions to be similar if the anchors had no effect. However, if the anchors did have an effect, we would expect the distributions to shift so that the distribution of responses in the high (low) anchor condition would in general be higher (lower) than in the unanchored condition. We would also predict that highly confident subjects would be less affected by the anchors than less confident subjects.

Subjects were 156 students at the University of California, Berkeley. They completed the questionnaire as partial fulfillment of a course requirement in an introductory psychology class.
Subjects were asked to give their best estimates in response to 15 questions. Then, they were asked to rate their confidence in their answer on a ten point scale on which 0 was labeled "not at all confident," 5 was labeled "moderately confident," and 10 was labeled "extremely confident." Questions included some measurements such as the height of Mount Everest and some quantities such as the number of nations that are members of the United Nations.
Pretest subjects (N=53) were asked the questions directly. Anchor values for each question were chosen as the 15th and 85th percentile responses from the distribution of the pretest subjects' responses.

Experimental subjects (N=103) answered pairs of questions. The first question asked whether the quantity in question was greater or less than an anchor value. The second question was identical to the pretest questions which asked for a specific answer. There were two versions of the questionnaire, each with half high anchors and half low anchors.

In order to provide a measurement of anchoring, an index of motion toward the anchor was developed. The index for each question was defined to be the distance between the medians obtained with the high and low anchors divided by the distance between the high and low anchor values. An index value of 0 would indicate that no motion toward the anchor occurred because the two medians are identical. Greater values of the index indicate a higher degree of anchoring effects because the medians are farther apart.

To test the hypothesis that the degree of anchoring *is inversely proportional to the level of confidence*, the correlations between the index values and the mean and median confidences were calculated separately for the unanchored and anchored groups. For the unanchored groups, the correlation with the mean confidence was r=-.675 (r^2=.455) and the correlation with the median confidence was r=-.741 (r^2=.549). For the anchored groups the relationship was even stronger. The correlation with the mean confidence was r=-.818 (r^2=.669) and the correlation with the median confidence was r=-.840 (r^2=.705).

To further examine this relationship, low confidence subjects were separated from high confidence subjects for each question using a median split and separate index values were calculated. For all but one question, the index value is lower for the high confidence than low confidence subjects (see Table 9.1). Thus, highly confident subjects were less affected by the anchors than were less confident subjects. To test whether the distributions of responses were significantly affected by the high and low anchor values, Mann-Whitney tests were performed for each question. All of the differences were highly significant.

References

Agostinelli, G., Sherman, S.J., Fazio, R.H., & Hearst, E.S. Detecting and identifying change: Additions vs. deletions. Journal of Experimental Psychology: Human Perception and Performance, 12, 1986, 445-454.

Brickman, P., and D.T. Campbell. Hedonic relativism and planning the good society, in M.H. Appley (Ed.), Adaptation-Level Theory: A Symposium, New York: Academic Press, 1971.

Emmons, R.A., and E. Diener. Factors predicting satisfaction judgments: A comparative examination. Social Indicators Research 16, 1985, 157-167.

Festinger, L. A theory of social comparison process. Human Relations 7, 1954, 230-243.

Fox, C., & Kahneman, D. Correlation, causation and inference in surveys of life satisfaction. Social Indicators, 1992.

Helson, H. Adaptation level theory: An experimental and systematic approach to behavior. New York: Harper & Row, 1964.

Kahneman, D. Reference points, anchors, norms, and mixed feelings. Organizational Behavior and Human Decision Processes 51, 1992, 296-312.

Kahneman, D., Treisman, A., & Gibbs, B.J. The reviewing of object files: Object-specific integration of information. Cognitive Psychology 24, 1992, 175-219.

Kahneman, D., & Varey, C. Notes on the psychology of utility. In J. Roemer and J. Elster (Eds.), Interpersonal Comparisons of Well-Being, New York: Cambridge University Press, 1991.

Kahneman, D., & Snell, J. Predicting utility. In R.M. Hogarth (Ed.), Insights in Decision Making, Chicago: University of Chicago Press, 1990.

Kahneman, D., & Miller, D.T. Norm theory: Comparing reality to its alternatives. Psychological Review, 1986, 93, 136-153.

Lyons, J. Semantics, Volume 2. Cambridge, England: Cambridge University Press, 1977.

Markovsky, B. Anchoring justice. Social Psychology Quarterly 51, 1988, 213-224.

Schul, Y., & Szyf, M. Implicit theories of satisfaction: Concurrent influences of temporal and interpersonal standards. Unpublished manuscript, 1991, Hebrew University, Jerusalem.

Schwarz, N., & Clore, G. Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. Journal of Personality and Social Psychology 45, 1983, 513-523.

Schwarz, N., & Strack, F. Evaluating one's life: A judgment model of well-being. In Strack, F., Argyle, M., and Schwarz, N. (Eds.), Subjective Well-Being, Oxford: Pergamon Press, 1991.

Tversky, A. Features of similarity. Psychological Review 84, 1977, 327-352.

Tversky, A., & Kahneman, D. Loss aversion in riskless choice: A reference-dependent model. Quarterly Journal of Economics (Nov. 1991), 1039-1061.