BRIEF REPORTS

Response conformity in recognition testing

DANA M. SCHNEIDER and MICHAEL J. WATKINS
Rice University, Houston, Texas

In a recognition memory test for a just-studied word list, subjects responded positively or negatively to each test word in the presence of another subject, with the two taking turns to call out their responses. Responses given second tended to conform to those given first. This was so for responses to both targets and lures and following both positive and negative first responses, although the effect was sharply reduced for targets given a negative first response. The same pattern of results was obtained in a second experiment, in which an incognito confederate replaced one of the subjects and so brought one set of responses under full experimenter control. These experiments illustrate a powerful effect of social pressure on recognition responses.

Experimental psychologists reify memory and so seek its quiddity (see Watkins, 1990). From the start (Ebbinghaus 1964/1885), they have honored such immaculate conception with rituals designed to keep their procedures free from impurities. One such impurity is other people's memories, and so it is that to this day subjects in memory experiments are kept uninformed of one another's responses. An alternative conception is that such "impurities" are inherent to memory, and that to know memory is to know their effects (cf. Bartlett, 1932). From this standpoint, the effect that one person's memory might have on another's is not something to be reflexively avoided, but something to be explored in its own right.

Though often ignored by students of memory, the effect that the views and judgments of others have on our own has been of enduring interest to psychologists of other persuasions. In a seminal study by Asch (1956), subjects reported which of three comparison lines matched a reference line shown at the same time. People were tested in groups of 7 to 9, and they reported their responses aloud in the order in which they were seated. All but one were confederates of the experimenter pretending to be subjects, and on two-thirds of the trials they deliberately gave the same wrong answer. The real subjects, who reported second to last, conformed to the erroneous judgments, on average, 37% of the trials, with fully three-quarters of them conforming on at least 1 of the 12 rigged trials. Control subjects, who wrote their judgments pri-

vately, erred on fewer than 1% of the trials. Other experiments showed such dramatic conformity after as few as three errant judgments, though not after just one or two (Asch, 1955).

If people conform to group consensus in their perceptual judgments, it would be surprising if they did not similarly conform in their memory judgments, for few would trust their memories more than their perceptions (cf. Deutsch & Gerard, 1955). But conformity of memory judgments remains largely unexplored. Indeed, we are aware of just two pertinent studies. In one, Betz, Skowronski, and Ostrom (in press) questioned subjects about a story they had read, and following some of the responses showed what they claimed to be tallied responses of other subjects. In a subsequent cued recall test, the subjects' responses tended to reflect the dominant response of the respective "tallies." In the other study (Hoffman, Granhag, & Loftus, 1996), subjects judged whether, in an earlier part of the procedure, they had seen, imagined, or neither seen nor imagined each of a series of pictures. Prior to each judgment, they were given what was alleged to be the response of a previous subject and invited to use it in forming their own judgment. Their judgments tended to conform to the bogus responses. The research reported here extends these findings by demonstrating conformity to a single other response in a standard recognition procedure.

EXPERIMENT 1

Each subject's recognition memory for a long list of words was tested in the presence of a single other subject, with the two responding aloud. The subjects were led to believe that they had studied the same list, but in fact they had not. The correct response for the first respondent may or may not have been correct for the second. Conformity to both correct and incorrect responses was

This research was included in D.M.S.'s master's thesis and was supported by National Institute of Health Grant MH35873. We thank M. Engstrom for her talented performance as the confederate, A. Yanez for help with the conduct and scoring of Experiment 2, and H. Roediger III for encouragement and helpful comments. Correspondence should be addressed to M. J. Watkins, Rice University, Department of Psychology, MS 25, 6100 Main St., Houston, TX 77005-1892 (e-mail: watkins@rice.edu).

measured both for words that had been included in the study list (targets) and for words that had not (lures).

Method

Subjects. The subjects were 32 Rice University undergraduates.

Procedure. The subjects were tested in pairs. They were each given a study sheet and told to study the words on the sheet in preparation for a memory test. Using a pencil to keep place, they shifted their attention from word to word every 2 sec, as paced by a metronome.

Following the study phase, the experimenter commented that this was pilot research, and that she had not made individual response sheets for the upcoming recognition test but would instead call out the words. The subjects were to respond "yes" or "no" for each word, according to whether they accepted or rejected it as having been included in the study list. They took turns responding first, switching after each column of the experimenter's test list.

Design. Of 192 words selected from the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982), a collection of common bisyllabic words, a random 96 were designated as critical and the other 96 as fillers. Half of the filler words were included in the study list and half were recognition lures. The filler words were not counterbalanced with respect to target/lure status, their purpose being merely to maintain the deception that the subjects had studied the same list by increasing (to 75%) the proportion of test items for which the correct response was the same for each subject.

The experiment was designed around the status of the critical words, and specifically their status in the recognition test. For any given subject there were 12 words within each combination of three binary, within-subjects factors: target/lure status for one of a given pair of subjects, target/lure status for the other subject, and the order in which the two subjects responded. Thus, for a given subject, 12 targets to be responded to first were also targets for the other subject, 12 targets to be responded to second were also targets for the other subject, 12 targets to be responded to first were lures for the other subject, and so on. Four pairs of subjects were needed to rotate these 12-word subsets through the various combinations of conditions. There were three replications of this core design for a total of 16 pairs of subjects.

Study lists and recognition test. Four versions of the study lists were required to complete the counterbalancing. Each consisted of four columns of words. Each column included 12 critical words, of which 6 were also studied by the other subject within the pair, and 12 fillers. To lessen the

chances of a subject noticing from an inadvertent glance that their cosubject's study sheet differed from his/her own, the first and last three words in each column were the same for all versions, and so, necessarily, were filler words. The order of the remaining words within each column was random.

A single recognition test was used for all subjects. It included all 192 words. For each subject, 96 of the words (48 filler words and four subsets of 12 critical words) were from the study list and thus targets, and 96 (48 filler words and four subsets of 12 critical words) had not been included in the study list and so were lures. The written version of the test, which only the experimenter saw, was arranged in eight 24-word columns. To strengthen the subjects' belief that they had studied the same list as their cosubjects, the first two columns included only filler words, with 12 targets and 12 lures in each column. Each of the remaining six columns included 16 critical words (specifically, 4 of each combination of target/lure status for each subject within a pair) and 8 fillers (4 targets and 4 lures).

Results and Discussion

Overall, the proportion of test items accepted as targets (i.e., given a "yes" response) was .62 for targets and .27 for lures. The corresponding proportions for first responses, which formed base rates from which conformity was measured, were .62 and .25.

Of chief interest, second responses were sharply biased by first responses. With the data collapsed across target/lure status, the mean proportion of words accepted as having been included in the study list was substantially higher for words that had just been accepted (.59) than for words that had just been rejected (.36) by the other subject [$t(31) = 8.50, p = .00$]. This bias held for both targets (.74 vs. .55) and lures (.41 vs. .18) considered separately [$ts(31) = 4.87$ and $5.75, p = .00$ in both cases].

Two kinds of conformity score were computed for each combination of target/lure status and prior response by the other subject. The raw conformity score was simply the change from base (i.e., first response) acceptance rate in the direction of the other subject's (prior) response. For

Table 1
Proportion of Positive Second Responses and Mean Conformity Scores for
Targets and Lures: Experiments 1 and 2

	Proportion of Positive Responses*	Conformity			
		Score	Raw		Conditional Score†
			<i>t</i>	<i>p</i>	
Experiment 1					
Targets (base rate = .62)					
After <i>yes</i> response	.74	.12	3.60	.00	.32
After <i>no</i> response	.55	.07	2.35	.01	.11
Lures (base rate = .25)					
After <i>yes</i> response	.41	.16	4.24	.00	.21
After <i>no</i> response	.18	.07	2.66	.01	.28
Experiment 2					
Targets (base rate = .64)					
After <i>yes</i> response	.77	.13	3.49	.00	.36
After <i>no</i> response	.60	.04	1.32	.10	.06
Lures (base rate = .38)					
After <i>yes</i> response	.54	.16	3.25	.00	.26
After <i>no</i> response	.29	.09	2.38	.01	.24

Note—Conformity scores are defined in the text. All *ps* are one-tailed. *For Experiment 1, proportions are unweighted averages of the individual subjects' proportions.

†Conditional conformity scores are for pooled data.

example, if a subject accepted 11 of the 48 (i.e., .23) critical lures when responding first and 4 of the, say, 23 (i.e., .17) that the other subject had just rejected, he/she would be given a raw conformity score of $(.23 - .17 =) .06$ for this condition. Note that positive scores indicate a shift in the direction of the other subject's prior response. Table 1 shows the mean of the subjects' raw conformity scores for each of the four conditions. As is also shown in Table 1, for none of the conditions was conformity likely to have arisen wholly by chance.

Comparing the raw conformity scores across conditions is tricky. The problem is that opportunity to conform depends on base rate. For example, a subject with a .10 base acceptance rate for lures who conformed completely to prior rejections would be given a raw conformity score of only $(.10 - .00 =) .10$. A better idea of the relative degrees of conformity is given by the *conditional* conformity score, which is the raw conformity score divided by the highest possible raw conformity score. Conditional conformity scores of 0.00 and 1.00 thus represent zero and complete conformity, respectively. These are shown for the pooled data in the far-right column of Table 1. Unfortunately, with sparse data the estimates for individual subjects are unstable. To take an extreme example, the conditional conformity score in the rejected lure condition for a subject with a lure acceptance base rate of 0.00 would be indeterminate. Therefore, the conditional conformity scores were not subjected to inferential analyses. Nevertheless, the means for the pooled data clearly suggest less conformity in the rejected target condition than in any of the others.

The results of this experiment demonstrate that the judgments made by the second of two respondents in a recognition test tend to conform to those of the first respondent, whether or not the first respondent was correct. Thus far, we have assumed that such response conformity is due to the second respondent being influenced by the first. It is conceivable, however, that there was no such influence at all, and that the observed conformity was an artifact of a variability among the words in their proneness to be judged as having occurred in the study list. If, for whatever reason, some words tend to invite acceptance and others rejection, then subjects would tend to conform in their judgments even if tested in isolation.

The data were therefore reanalyzed with second responses conditionalized, not on first responses, but on what first responses would be if correct (i.e., on target/lure status for the first respondent). Unlike the subject-mediated first responses, target/lure status was controlled by the experimenter and hence not subject to item selection effects. The overall acceptance rate among second responses was discernibly greater for the first respondent's targets (.50) than for the first respondent's lures (.42) [$t(31) = 3.53, p = .00$], indicating that response conformity was not entirely due to item variability. That second responses conformed much less to the correct than to the actual response of the first respondent is hard to interpret. Certainly, this result would be expected if conformity to first responses were partly an artifact of item vari-

ability effects, but it would also be expected in the absence of such effects, for target/lure status for the first respondent could bias the second responses only to the extent that first responses are correct.

In short, this experiment has shown marked conformity of recognition responses to those of a single other person. Some of this conformity could have arisen from a possible inherent variability in the tendency of the test items to elicit positive recognition responses, but the evidence for a direct influence of the first respondent on the second warranted investment in a second, somewhat more costly, experiment.

EXPERIMENT 2

Experiment 2 differed from Experiment 1 in two main ways. First, only one person in each pair was a subject. The other pretended to be a subject but was in fact a confederate of the experimenter. This deception allowed complete experimenter control over one set of responses, thus precluding any contaminating effect of item variability. Second, each *yes/no* recognition response was rated according to a 3-point confidence scale.

Method

Subjects. The subjects were 24 Rice University undergraduates. None had participated in Experiment 1.

Procedure. The procedure was essentially the same as in Experiment 1, though with a confederate replacing one of the subjects of each pair. One minor difference was that the study list was displayed, in two 48-word halves, by overhead projector, with the experimenter reading the words aloud at a rate of one every 2 sec. The recognition test differed from that of Experiment 1 in that the respondents supplemented each *yes/no* decision with a rating of their confidence in their decision according to a 3-point scale (1 = *maybe correct*, 2 = *probably correct*, and 3 = *definitely correct*). The confederate's *yes/no* decisions were controlled by the direction of the experimenter's gaze—an experienced researcher having, in pilot research, failed a challenge to identify this signal. The confederate's confidence ratings were determined by the first letter of the test word. The subject and the confederate reversed their order of responding from time to time, as directed by the experimenter.

Design. As before, 192 words were chosen from the Toronto Word Pool (Friendly et al., 1982), but this time with the distribution of their initial letters (the confederate's confidence cues) constrained so that a third began with a letter between A and F, a third with a letter between G and N, and a third with a letter between O and Z. A random half of the words in each of these three sets were designated critical and the others fillers.

The design was the same as in Experiment 1, but with the confederate's *yes/no* decision substituting for the target/lure status for the "other" subject, and with confidence rating added as a fourth factor. For any given subject, four critical words were assigned to each of the $(2 \times 2 \times 2 \times 3 =) 24$ combinations of the factor levels, the choice of these words being rotated across the 24 subjects. For the 12 combinations involving targets, these four words were distributed one to each quarter of the study list. Each quarter of the study list also included 12 filler words. The order of the words within each quarter was random except that filler words were placed in the first six positions of the first quarter and last six positions of the last quarter. Two versions of the study list counterbalanced target/lure status.

As in Experiment 1, all 192 words were included in the recognition test, and the responses to only the critical words were scored. The first 48 test words were fillers. Of these, the first and second 24 each included 12 targets and 12 lures. Half of the subjects responded first for the first 24 words and second for the next 24; the other half responded second for the first 24 words and first for the second 24. Of the remaining words,

each successive block of 18 comprised, in random order, 12 critical and 6 filler words, with 1 critical word corresponding to each combination of target/lure status, confederate *yes/no* response, and confederate confidence rating, and with 1 filler word corresponding to each combination of target/lure status and confederate confidence rating—the confederate's *yes/no* responses to filler words being always correct. The subject and confederate switched order of responding after each block.

Results and Discussion

As shown in Table 1, the results of preliminary analyses in which confidence ratings were disregarded were akin to those of Experiment 1. Thus, for the combined target/lure data, the proportion of positive responses was higher following a positive (.65) than following a negative (.44) confederate response [$t(23) = 4.58, p = .00$]. This relation held for both targets (.77 vs. .60) and lures (.54 vs. .29) considered separately [$ts(23) = 3.75$ and 4.39 , respectively, $p = .00$ in both cases]. Moreover, the overall average of the pooled conditional conformity scores was identical to that of Experiment 1 (.23), and the pattern of conditional conformity scores across the conditions was similar. In particular, that for targets after confederate rejections was noticeably lower than those for any other condition.

For the analyses of primary interest, each *yes/no* decision was combined with its confidence rating to form what may be called a recognition rating, expressed on a scale from 1 to 6 (1 = *No, definitely*, 2 = *No, probably*, 3 = *No, maybe*, 4 = *Yes, maybe*, 5 = *Yes, probably*, and 6 = *Yes, definitely*). The overall mean recognition rating was 4.20 for targets and 3.17 for lures; that for first responses was 4.13 for targets and 3.17 for lures.

The key finding, shown in Figure 1, was that recognition ratings increased fairly steadily as a function of the confederate's prior recognition rating. The linear component of this relation was unlikely to have arisen by chance: With the data for targets and lures combined, $t(23) = 6.40$,

$p = .00$; for the targets alone, $t(23) = 4.87, p = .00$, and for the lures alone, $t(23) = 6.63, p = .00$. The linear component of trend was discernibly greater for lures than for targets [$t(23) = 3.80, p = .00$].

The linear components of trend clearly owed much to the confederate's *yes/no* responses as opposed to her confidence ratings. However, additional analyses of the combined target and lure data showed a positive residual linear trend in the subjects' recognition ratings across the three levels of the confederate's recognition rating within the *yes/no* responses—that is, in the subjects' recognition ratings after *no* (3) or *yes* (1), *no* (2) or *yes* (2), and *no* (1) or *yes* (3) confederate responses [$t(23) = 3.45, p = .00$].

The main conclusion from this experiment is that conformity of recognition responses can arise from the direct influence of 1 respondent on another. The close similarity of the results of this experiment to those of Experiment 1 suggest that response conformity observed in Experiment 1 owed little if anything to item variability effects.

GENERAL DISCUSSION

Cognitive psychologists like to set their own agenda, but over the last decade or two a growing interest in false memories and alleged recoveries of repressed memories has spilled from the courtroom and the public domain into their research sancta (see Lindsay & Read, 1994). One result has been a flurry of research on the role of social setting in the remembering process. Research on response conformity is a facet of this trend. The experiments reported here illustrate this phenomenon with a standard recognition procedure. They show that, given the chance, people will conform in substantial measure to another person's recognition responses.

Two respects in which the present study differed from most other studies of how memory responds to social pressure are worth noting. First, the pressure preceded the subject's response by a mere second or so rather than by minutes or even days, thus precluding the possibility of subject confusion about the source of the pressure. Second, it showed that people could be induced to increase, as well as to decrease, their response accuracy.

Figure 1. Mean recognition ratings for targets and lures as a function of the confederate's prior response: Experiment 2. N, *no* response; Y, *yes* response; 1, *maybe*; 2, *probably*; 3, *definitely*.

In both of our experiments, responses to targets that had been rejected by other respondents proved relatively immune to social pressure. Thus, the conditional conformity score for those targets to which the other respondent had given a negative response was only 41% (Experiment 1) or 21% (Experiment 2) of the average of the scores for the other three conditions. This finding could be said to make intuitive sense, for of the four combinations of subject response and target/lure status, positive responses to targets might be expected to be given with the most conviction and thus be least subject to revision in the face of a contrary judgment. And indeed, analysis of the confidence data of Experiment 2 confirmed this assumption: The proportion of responses judged *definitely correct* was considerably higher for positive responses to targets (.36) than for positive responses to lures (.10) or for negative responses to either targets (.06) or lures (.14).

A meta-analysis of perception experiments of the sort reported by Asch (1956) has suggested that response conformity in perceptual judgments has diminished across the years (Bond & Smith, 1996). Although the data for a comparable analysis of response conformity in memory judgments is lacking, it is clear from our and other recent experiments (Betz et al., in press; Hoffman et al., 1996) that memory responses are highly malleable. Conformity of perceptual judgments has been referred to as the "majority effect" because the conformity is typically to a clear majority. The experiments reported here show that no such majority is needed to induce conformity of memory judgment. Whether such a robust "extraneous" effect will tempt a reevaluation of the prevailing reified conception of memory is, however, another matter.

REFERENCES

- ASCH, S. E. (1955, November). Opinions and social pressure. *Scientific American*, **193**, 31-35.
- ASCH, S. E. (1956). Studies in independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs*, **70**, (9, Whole No. 416).
- BARTLETT, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- BETZ, A. L., SKOWRONSKI, J. J., & OSTROM, T. M. (in press). Shared realities: Social influence and stimulus memory. *Social Cognition*.
- BOND, R., & SMITH, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin*, **119**, 111-137.
- DEUTSCH, M., & GERARD, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal Psychology*, **51**, 629-636.
- EBBINGHAUS, H. E. (1964). *Memory: A contribution to experimental psychology*. New York: Dover. (Originally published 1885)
- FRIENDLY, M., FRANKLIN, P. E., HOFFMAN, D., & RUBIN, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation*, **14**, 375-399.
- HOFFMAN, H. G., GRANHAG, A., & LOFTUS, E. F. (1996). *The role of conformity in the creation of false memories*. Manuscript in preparation.
- LINDSAY, D. S., & READ, J. D. (1994). Psychotherapy and memories of childhood sexual abuse: A cognitive perspective. *Applied Cognitive Psychology*, **8**, 281-338.
- WATKINS, M. J. (1990). Mediationism and the obfuscation of memory. *American Psychologist*, **45**, 328-335.

(Manuscript received May 1, 1996;
revision accepted for publication June 24, 1996.)