

Reality monitoring vs. discriminating between external sources of memories

CAROL L. RAYE

Barnard College, New York, New York 10027

and

MARCIA K. JOHNSON

State University of New York, Stony Brook, New York 11794

Subjects participated in small groups. Two subjects engaged in a "conversation" in which they alternated generating single words, and other subjects listened to the conversation. Later, speakers were better than other subjects at correctly identifying the origin of words. This held both for a situation in which the subjects had considerable freedom in what they generated (Experiment 1) and for a situation in which what they generated was constrained by specific cues (Experiment 2). The fact that discriminating between externally and internally derived memories was easier than discriminating between two external sources of memories is consistent with the idea that the classes of internally and externally derived memories differ in characteristic ways.

Memories originate from self-generated events such as thought, as well as from perceptual events. We have used the term "reality monitoring" to refer to the processes people use in deciding whether memories initially had an external or an internal source (Johnson, 1977; Johnson & Raye, in press, Note 1). The present experiments address the question of whether confusion between memories from internal and external sources is similar to confusion between memories from two external sources. That is, is reality monitoring subject to more or less error than monitoring the origin of information derived from two external sources? Our working model proposes that there are a number of general dimensions on which externally and internally derived memories characteristically differ and which may be used in the process of reality monitoring. Externally derived memories are proposed to have more information about time and place, more sensory information, and more meaningful details. Internally generated memories are proposed to have more information about cognitive operations or processing that took place when the memory was established. Thus, reality monitoring decisions could be based simply on the amount of certain types of information (e.g., if a memory has a high sensory component, classify it as external, or, if it has a great deal of cognitive operations information, classify it as internal). In contrast, the origin of two memories derived from external sources could not be distinguished along these general dimensions. Thus, memories from external sources would have to be discriminated on the basis of specific content (specific

sensory information rather than amount, etc.). It follows that, given similar situations, it should be easier to identify the origin of internally vs. externally derived memories than to identify the sources of memories derived from two external sources.

EXPERIMENT 1

Method

The subjects were members of a cognition laboratory course at Barnard College. They participated in small groups during the first phase of the experiment. Within a group, they were randomly assigned to play different roles; two were assigned to be speakers, two were assigned to be recorders, and two (and any extra subjects in that group) were assigned to be listeners. All subjects thought that the speakers and recorders were serving as "experimenters," that the listeners were the "subjects," and that the subjects would later be tested for their memory of the "conversation." The members of the group were asked to introduce themselves (most of the students already knew each other, at least by name), and the procedure was then explained. The instructions indicated that the lab assistant would give the group a topic of conversation to start them off, for example, "snow." Speaker A was then to give a highly related word such as "white." Speaker B was to give a word related to Speaker A's word (e.g., "black"), then Speaker A would reply (e.g., "coal"), and so on, alternating. The speakers were instructed to try to use common words and to try to respond quickly. Recorder A was instructed to write down all of the words spoken by Speaker A, in order, on a piece of lined paper. Similarly, Recorder B was instructed to write down the words given by Speaker B. The listeners were simply instructed to listen carefully to both speakers, and they expected to receive a memory test later. Four groups of subjects were run simultaneously in a large room. In all, there was a total of eight speakers, eight recorders, and 12 listeners (with one group having 2, two groups having 3, and one group having 4 listeners). Each group was given a different word to start them off ("school," "lizard," "television," "happiness"). The speakers continued alternating until each had given 15 words. The recorders' papers were then

This research was supported by a grant from the National Science Foundation.

collected, and the students spent approximately 1 h listening to a lecture.

During this retention interval, the lab assistant made up two 60-item recognition lists. Two groups were yoked such that the 30 items generated by the speakers in one group served as the distractors for the subjects from the other group and vice versa. The 60 items were randomly assigned to test order, written in two columns on a sheet of paper, and dittoed for use with the subjects in those two groups. A second recognition list was made up in a similar fashion for subjects in the other two groups. On the recognition test, subjects were instructed to indicate with the person's initial which speaker said each word. The subjects responded with an "N" if neither speaker said a particular word. They were paced through the recognition test at a 5-sec rate by having them move masks, which covered all but a single test item, in time with an auditory cue.

In summary, speakers were later asked to identify the origin of externally vs. internally generated information. Listeners and recorders later were asked to identify the origin of information from two external sources. Recorders were included to assess the possibility that expected differences between speakers and listeners might be the result of listeners' distributing their attention across both speakers while speakers paid more attention to what they themselves said. Recorders saw their job as essentially to make sure they recorded what one of the speakers said. Thus, recorders were like speakers, in that they produced half of the items with an overt response; but, whereas speakers actually generated the items, recorders simply recorded them.

Results and Discussion

Two measures were calculated for every subject: The first, recognition errors, was the sum of the misses (calling a word from the conversation a new word) and the false positives (attributing a new word to either of the speakers); the second measure was the percent correct identification of origin, the total number of words correctly attributed to each of the two speakers divided by the total number of words correctly identified as from the conversation. The means of both of these measures for the three conditions are shown in Table 1.

Since the recorders were basically a differential attention control group, the first comparisons were made between recorders and listeners. These two conditions did not differ on either measure; these conditions were then combined and contrasted with speakers. Thus, the analyses of primary interest compared subjects who had external and internal sources of information with subjects who had two external sources of information. While it appears that overall recognition of an item as

old was better for the speakers than for the other two conditions, this comparison was not significant [$F(1,26) = 1.18$].

For the present purposes, the measure of primary interest was the percent correct identification of origin. According to the model, subjects discriminating between internal and external sources were expected to show better identification of origin than subjects discriminating between two external sources of information. This prediction was confirmed [$F(1,26) = 2.93, p < .10$].

This result can be taken as support for the idea that memories for external events as a class differ from the class of memories for internal events. These differences potentially provide cues that allow the origin of memories for internally generated and externally derived events to be distinguished more easily than the memories for two comparable external events. For example, a difference in the average amount of cognitive operations information produced from perception vs. generation (with more created in generating) could serve as a discriminative cue available in external-internal discriminations that is not available in external-external discriminations.

Our working model of reality monitoring also allows for the possibility that qualitative, as well as quantitative, information is assessed in a more extended reasoning phase of reality monitoring. Applying this to the present situation, assume that the semantic content of the externally and internally derived memories differ in some way. Internally generated responses should be more idiosyncratic than external events. Internally generated events are not necessarily more meaningful than external events, but they are more likely to refer to biographical information or information relevant to the "self." The presence of a difference in average idiosyncrasy value for external and internal items could provide a discriminative cue: Would I have said this? Is this a high-frequency response for me? Does this response have some special meaning for me?

EXPERIMENT 2

In a second conversation study, the content (as well as idiosyncracies) of the conversation was controlled. A subject that we will call the "director" asked the speakers questions such as "What is a word that is the opposite of fast, beginning with s and ending with w?" The director indicated which of the two speakers was to answer, and the speakers' responses were essentially determined by the questions, which, obviously, left little room for idiosyncratic contributions to the content of the conversations. Again, there were recorders and listeners, and the listeners thought they would receive a memory test. All subjects, including the directors, were later given a recognition test that required them to indicate who said what. Of primary interest was whether

Table 1
Recognition Errors (RE) and Percent Correct Identification
of Origin (PC) for Experiment 1

	RE	PC
Speaker	3.25	90
Recorder	6.12	81
Listener	4.58	81

Note—Recognition errors include misses + false positives.

there would still be superior discrimination between internal and external items, compared with external items from two sources, when the idiosyncratic component of the internally generated information was eliminated.

Method

The subjects were members of a cognition laboratory course at Barnard College. None of these subjects served in Experiment 1. They participated in four small groups, as in Experiment 1. Within each group, they were randomly assigned to roles. There were a total of eight speakers, eight recorders, nine listeners, and four directors. The cover story used was similar to that of Experiment 1, and thus only the listeners were explicitly instructed to expect a memory test. The directors were given a "script" that indicated the questions and which speaker was to answer each. There were two types of questions, opposites (e.g., "old-new") and category instances (e.g., "crime-murder"). The director first indicated which speaker was to answer and then indicated the type of question (opposite or category), the first item (e.g., old), and the first and last letter of the desired answer (e.g., "opposite of old that begins with n and ends with w"). There were 24 questions of each type, randomly distributed throughout the sequence, with the restriction that they be distributed evenly throughout. Speakers answered in random order, with the restriction that each speaker answer an equal number of each type of item. Again, each recorder wrote down the responses of one of the speakers, and at the end of this phase of the experiment, the recorders' papers were collected and the subjects spent approximately 1 h in lecture and various other lab activities. Then all subjects received a recognition test consisting of all 48 target responses randomly intermixed with 24 distractors (12 opposites and 12 category instances from categories that were not present in the first phase of the experiment). The subjects were asked to indicate for each item whether it was said by Speaker 1 or Speaker 2 or neither. The subjects were paced through the recognition test at a 5-sec rate by having them move masks in time with an auditory cue.

Results and Discussion

Generally, the speakers produced the desired answers in Phase 1. Failing to respond or producing a response different from that expected averaged only 2.88 and .38 responses/group for category instances and opposites, respectively. These items were then omitted for all members of the relevant group of speakers, recorders, listeners, and director when the recognition data were scored. As will be clear from Table 2, subjects were very near ceiling for the category items, and therefore category and opposites were analyzed separately.

Table 2
Recognition Errors (RE) and Percent Correct Identification of Origin (PC) for Experiment 2

	Category Instances		Opposites	
	RE	PC	RE	PC
Speaker	2.12	99	6.25	83
Recorder	3.00	96	7.88	73
Listener	2.00	88	5.00	69
Director	1.75	80	6.50	58

Note—Recognition errors include misses + false positives.

For the category items, compared with listeners, both speakers and recorders were better at identifying the source of items they called old [$F_s(1,28) = 9.36$ and 5.60 , respectively; $MSe = 55.15$]. Although speakers were a bit better than recorders, both groups were very near ceiling and did not differ ($F < 1$). The directors were worse than the speakers [$F(1,25) = 18.15$], recorders [$F(1,25) = 13.77$], and listeners [$F(1,25) = 3.49$; $p < .10$] at identifying the origin of the information. In general, recognition errors were rare, as can be seen in Table 2, and there were no differences among the conditions [$F(1,25) = .48$, $MSe = 4.22$].

Performance on the opposites was well below ceiling and therefore provided the more sensitive comparison among the conditions. As can be seen in the right half of Table 2, the opposites produced exactly the same pattern as that found in Experiment 1. As in Experiment 1, the speakers were significantly better at identifying the origin of an item they recognized as old than were the recorders and listeners [$F(1,25) = 4.27$, $MSe = 172.35$, $p < .05$], and the recorders and listeners did not differ [$F(1,25) < 1$]. As with the category items, the ability of the directors to identify the origin of opposites was quite low. The relevant $F_s(1,25)$ for directors compared with speakers, recorders, and listeners were 9.77 , 3.72 ($p < .10$), and 2.25 ($p > .10$), respectively.

The recognition errors (misses plus false alarms) for the opposites are also shown in Table 2. Remember, for this analysis, a response was counted correct if the subject indicated an item was old, even if the subject was mistaken about who said it. As in Experiment 1 and with the category items of the present experiment, the conditions did not differ in ability to identify an item as old [$F(3,25) = 1.41$, $MSe = 8.30$]. It is particularly interesting that the directors were as good as the other subjects on recognition, even though they tended to be so much poorer on the identification of the source of the items, suggesting that simply attending to the items does not guarantee correct identification of origin. The dissociation of memory for occurrence and identification of origin indicates they are not based on exactly the same information or mechanisms.

The finding in Experiment 2 that speakers were better than listeners and recorders at identifying the origin of opposites rules out, we think, the possibility that the entire difference in Experiment 1 could be attributed to differences across conditions in the idiosyncratic value of the content of the items. In fact, the magnitude of the difference was almost exactly the same as it was in Experiment 1, suggesting that idiosyncratic cues did not add much. Of course, we would not want to offer this as a general conclusion, since we would expect idiosyncratic value to be one of the cues used in the reasoning processes involved in reality monitoring. However, the present results are consistent with the

possibility that, independent of the personal significance of what an individual says, the cognitive operations that go into generating information persist in memory and become potential cues as to the origin of that information (Durso & Johnson, in press; Raye, Johnson, & Taylor, 1980).

REFERENCE NOTE

1. Johnson, M. K., & Raye, C. L. *A working model of reality monitoring*. Paper presented at the annual meeting of the Psychonomic Society, San Antonio, Texas, November 1978.

REFERENCES

- DURSO, F. T., & JOHNSON, M. K. The effects of orienting tasks on recognition, recall, and modality confusion of pictures and words. *Journal of Verbal Learning and Verbal Behavior*, in press.
- JOHNSON, M. K. What is being counted none the less? In I. M. Birnbaum & E. S. Parker (Eds.), *Alcohol and human memory*. Hillsdale, N.J: Erlbaum, 1977.
- JOHNSON, M. K., & RAYE, C. L. Reality monitoring. *Psychological Review*, in press.
- RAYE, C. L., JOHNSON, M. K., & TAYLOR, T. H. Is there something special about memory for internally generated information? *Memory & Cognition*, 1980, **8**, 141-148.

(Received for publication April 23, 1980.)