Journal of Experimental Psychology:
Human Learning and Memory
1979, Vol. 5, No. 3, 229-240

Fact and Fantasy: The Roles of Accuracy and Variability in Confusing Imaginations With Perceptual Experiences

Marcia K. Johnson
State University of New York at Stony Brook

Carol L. Raye
Barnard College

Alvin Y. Wang and Thomas H. Taylor
State University of New York at Stony Brook

In Experiment 1, pictures were presented to subjects two, five, or eight times, and subjects were asked to imagine each picture two, five, or eight times. Subsequently, subjects estimated the number of times each picture had been presented. Their estimates of the frequency of these external events were influenced by imagination trials; this effect was greater for good imagers than for poor imagers. Experiment 2 involved a similar design in which subjects were asked either to imagine the same referent for a word or to imagine a different referent for a word on successive imagination trials. Consistency (same referent) did not increase the influence of imaginations on immediate judgments of external frequency. Thus, the results of Experiment 1 were attributed to the greater accuracy (as opposed to greater consistency) of good imagers' internal generations of the stimuli. Furthermore, variation (imagining different referents), like greater accuracy, increased the effects of imagination trials on immediate but not on delayed judgments of frequency. Possible mechanisms underlying these effects are discussed. In general, the two studies show that qualitative characteristics of completely covert generations influence their impact on estimates of the frequency of external events.

Early philosophers were explicitly concerned about the relationship between thought and external reality. Does thought reflect external reality? Does thought distort our picture of reality? Psychologists, in considering this dilemma, have sometimes emphasized reality at the expense of thought, and sometimes emphasized thought at the expense of reality. If pressed, probably most of us would say that both reality and thought contribute to our perception of the world. This resolution leaves open many questions however. For example, how do we distinguish fact from fantasy (Johnson, Note 1)?

For memory theorists, the problem is compounded. If it is difficult to separate the contributions of reality and thought in the present (Perky, 1910), it may be even more difficult to separate the *memory* representations of previous perceptual events and the memory representations of previous imaginations. Memory theorists do not have satisfactory ways of characterizing the difference between veridical and nonveridical memory or of dealing with the difference between a bit of harmless embellishment of events that does not greatly affect the veracity of the report and real distortions and extreme thought disorders that may reflect substantial failures to discriminate fact from fantasy in memory. Whether or not we only remember what we thought, most of us believe that there is a meaningful difference between memories for thoughts that reflected

This research was supported in part by an award to the second author from the Spencer Foundation, Teachers College. We thank Carol Sprafkin for her help in conducting Experiment 1. Part of these results were presented at the annual meeting of the Psychonomic Society, St. Louis, November 1976.

Requests for reprints should be sent to Marcia K. Johnson, Department of Psychology, State University of New York, Stony Brook, New York 11794.

real events as they were happening, thoughts that recreated these events, and thoughts that created events that never happened. It is important to try to discover the cues or mechanisms that allow people to distinguish between memories for previous perceptual experiences and memories for previous imaginations, and to discover the conditions under which this "reality monitoring" (Johnson, 1977) breaks down.

We previously reported a technique that seems quite reliably to produce confusion between externally and internally generated memories (Johnson, Taylor, & Raye, 1977). This provides one situation for testing hypotheses about variables that might affect the accuracy of reality monitoring, and thus might suggest something about the nature of the process. The basic paradigm involves varying independently the number of times events happen and the number of times subjects think about these events. Subsequently, subjects are asked to make frequency-of-occurrence judgments. For example, the subject might see a list of words; each word occurs either two, five, or eight times across trials. Interspersed throughout presentation trials are "generation" trials during which the subject is given high-associate cues for items and is instructed either to produce or to think about the target items. The subject thus generates each item either two, five, or eight times across generation trials. Later, the subject is asked to estimate how often each item was actually presented by the experimenter (external frequency).

Using this procedure, Johnson et al. (1977) found that judgments of external frequency increased as the number of internal generations increased, and called this the IFE effect (increase in apparent Frequency of External events as a consequence of internal events). In addition, there was some evidence that the magnitude of the IFE effect depended on what subjects did on generation trials. For example, there was a greater IFE effect when subjects wrote the items down than when they simply thought about them. Overt productions, because they involve seeing the generated word, may be more like presentations, and thus may be more likely to produce confusion. Events that are more similar to the target event should in general be more likely to affect its estimated frequency.

An interesting question arises regarding variations in the characteristics of *completely covert* generations. One possibility is that thoughts that are more like perceptions also may be more likely to be confused with these perceptions. On the other hand, it may be that all thoughts that activate the general concept of a target will add equally to judged frequency; differences in the quality of completely internal generations might matter little.

Experiment 1

The present study was directed at this question. Pictures, rather than words, served as the external events. The subjects were divided into good and poor imagers on the basis of an imagery test. The major question was whether good and poor imagers would differ in the magnitude of the IFE effect. Assuming that the mental regenerations of good imagers are closer to actual perceptual experience, they should show a greater IFE effect compared to poor imagers.

In addition to comparing good and poor imagers, the nature of the external events was varied. Items were presented either in color or in black and white. This manipulation was based on the assumption that it might be easier to generate an accurate image of a black-and-white line drawing than to generate an image of a color drawing. The black-and-white line drawings were more schematic and captured the core features of an object with less detail. Good and poor imagers may differ less in their ability to regenerate schematic features of experience than in their ability to regenerate its details. Thus, color pictures were expected to show a greater difference between good and poor imagers in the magnitude of the IFE effect than black and white pictures.

Method

Materials and design. Color pictures of common objects (e.g., clown, banana, mailbox) were selected from children's picture books. An artist produced a black-and-white line drawing of each item by tracing the original. Both color and black-and-white versions were photographed and presented as slides during the experiment. Experimental items were selected by five judges on the bases of common, independent assignment of verbal labels, the quality of the slides, and a minimum of associative relationships among the items. Half of the subjects saw 36 different pictures in color, and half saw the same 36 pictures in black and white. On the basis of an imagery test (described below), the subjects were divided into good and poor imagers. Thus, there were two between-subjects factors: type of subject (good vs. poor imagers) and type of picture (color vs. black-and-white). All subjects saw each picture either two, five, or eight times and were asked to generate images of each of these pictures either two, five, or eight times. Number of presentations and number of image trials were within-subject factors.

Procedure. Subjects participated in groups of up to five and were told that this was a study of imagery. Blocks of presentation and image trials alternated for 10 blocks of each type. On any particular block of presentation trials, 18 pictures were projected on a screen at a 2.5-sec rate. On image trials, subjects read a word in a booklet corresponding to the name of a picture, tried to imagine the picture, and rated how similar they felt their images were to the picture. Subjects generated 18 different images on each block of image trials, paced at a 4-sec rate. An equal number of items were randomly assigned to each of the nine combinations of presentation and image frequency. There were four different random assignments of items to presentation and image frequency. For each of these, which items were presented or imagined on each trial was random, with the restriction that no item be imagined before it had been seen. An attempt was made to prevent any systematic relationship between such variables as average lag between repetitions, position of last occurrence, and particular combinations of presentation and imagination frequency.

Following this phase of the experiment, the names of the pictures were read to the subjects in random order at a 5-sec rate. For each name, they were asked to write down the number of times the picture had been projected on the screen. Subjects were not warned in advance that frequency judgments would be required.

After this frequency judgment task, subjects were given a picture imagery test. Thirteen new pictures were presented, again either in color or in black and white, corresponding to the subject's condition. After a generation trial in which subjects rated the accuracy of their images, a surprise test was given consisting of a series of short-answer and multiple-choice questions about details of 7 of the test pictures (24 questions were the same for color and black-and-white subjects, and 5 additional questions concerning color were asked of the color subjects). In this test, subjects were asked to imagine the relevant picture. Then all questions about each picture were asked in succession.

Next, all subjects were given a test of color imagery. A slide of a color patch was projected, followed by a test slide comprised of six similarly colored patches, one of which was identical to the original. Subjects were required to pick out the correct color in the array. There were 10 of these items. Finally, all subjects were tested for free recall of the pictures from the first phase of the experiment.

Subjects. Forty introductory psychology students participated in the experiment and were assigned randomly to conditions.

Results[1]

Good versus poor imagers. For analyses that were potentially most sensitive to differences related to subject type, good and poor imagers were defined by using an extreme groups procedure that included only the 14 subjects with the best scores and the 14 subjects with the worst scores on the picture imagery test. These cutoffs included the maximum number of subjects while eliminating any overlap in the picture imagery scores of the two groups. The mean test score for good imagers in color (maximum score = 29) and black and white (maximum score = 24) conditions combined was 15.07; the mean for poor imagers was 9.57, $t(26) = 2.50$. The scores of these subjects on the color imagery test were then compared. Good imagers on the picture imagery test also gave more correct answers on the color imagery test (4.00 and 3.21 for good and poor, respectively), indicating some correspondence between performance on the two imagery tests. In general, the range of scores on the color imagery test was not as great as on the picture imagery test; therefore, it was less sensitive to subject differences.

Frequency judgments. Figure 1 presents the frequency judgments for good and poor imagers. Means for good imagers are in the left panel and, means for poor imagers are in the right panel. The separations between the lines in each panel indicate that increas-

[1] Unless otherwise indicated, reported comparisons were significant at the .05 level or better.

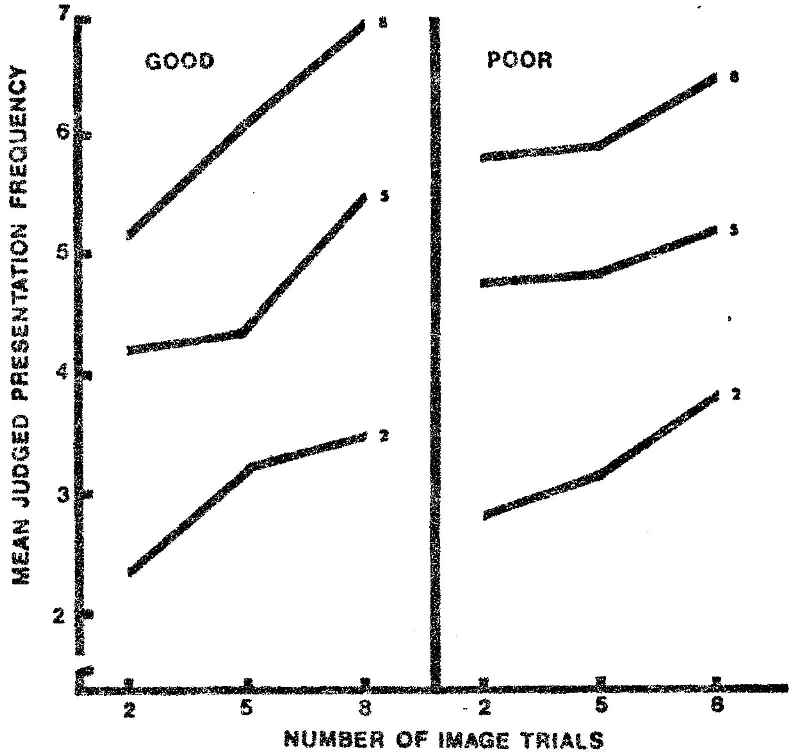

Figure 1. Judged presentation frequency as a function of the number of image trials. (Each line represents a different presentation frequency. The data are collapsed across type of picture [color vs. black and white], and good imagers are shown on the left and poor imagers are shown on the right.)

ing the number of actual presentations of pictures increased judgments of frequency of occurrence, $F(2, 48) = 116.49$, $MS_e = 1.56$. (See, e.g., Hintzman, 1969; Underwood, Zimmerman, & Freund, 1971). The effect of internally generated events on judged frequency of external events is indicated by the slopes of the lines; the more times subjects imagined an item, the more times they thought it had been presented, $F(2, 48) = 33.53$, $MS_e = .71$. This replicates the IFE effect found previously with words.

More important, the Type of Subject × Number of Image Trials interaction, $F(2, 48) = 3.53$, $MS_e = .71$, indicated that the IFE effect was greater for the good imagers (left panel) than for the poor imagers (right panel). The relationship between type of subject and the magnitude of the IFE effect is somewhat easier to see in Figure 2, in which the data have been collapsed across presentation frequency. There was no interaction between type of subject and number of presentation trials, $F(2, 48) < 1$, indicating that good and poor imagers did not differ simply in their accrual of frequency information in general. The triple-order Type of Subject × Type of Picture × Number of Generations interaction did not approach significance.

At this point it might be useful to consider two hypothetical outcomes that would both be consistent with the hypothesis that good imagers are more likely than poor imagers to confuse externally and internally generated memory traces. Suppose the probability that any particular image will be likely to be counted or to influence a judgment of presentation frequency is .50 for good imagers and .25 for poor imagers. If we assume that *image* traces that are counted are simply added to *presentation* traces that are counted and that memory for presentations is perfect, we would expect functions like those shown in the left half of Figure 3. That is, the judgments of good imagers would have a steeper slope as a function of number of image trials and would be higher overall. If, on the other hand, good and poor imagers have about the same idea of the average frequency of presentation of items, we would expect judgments to be corrected to reflect a judgment scale centered about this average. In this case, judgments would look more like the right half of Figure 3, indicating more confusion between presentations and images for good imagers, but about the same average anchor point for the scale in the two types of subjects. It seems plausible that good and poor imagers might start the frequency judgment task with similar assumptions about average frequency. A sense of average frequency could be derived without selecting or referring to particular memory traces—perhaps from a general memory of the overall structure of the task. For example, suppose the task took about 20 min. and the subject estimated he or she had spent about half that time seeing pictures and half that time imagining pictures. The 10-min. estimate of

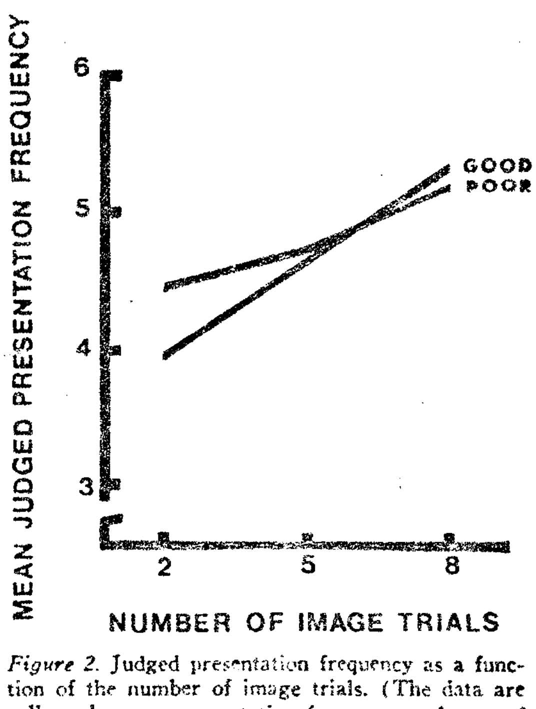

Figure 2. Judged presentation frequency as a function of the number of image trials. (The data are collapsed across presentation frequency and type of picture.)

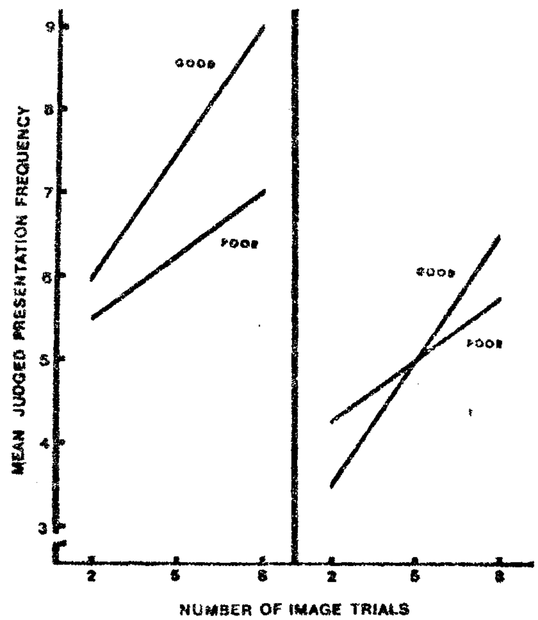

Figure 3. Hypothetical data showing judged presentation frequency as a function of the number of image trials.

viewing time could be divided by the product of the subject's estimate of the rate at which pictures were presented and the number of different pictures seen. This value could then serve as an anchor for relative judgments in the frequency of presentation task, as shown for the hypothetical data shown in the right half of Figure 3. (We have used an average of five presentations.) The point of this example is not to suggest that subjects consciously perform such calculations prior to the frequency judgment task, but to suggest that the anchor for the judgments might be based on information (such as time estimates) that might be similar for good and poor imagers. Thus, although a slope difference would be expected between good and poor imagers, an overall difference in absolute values for the presentation frequency judgments might but would not necessarily follow from differences in confusion.

There was one other significant effect, and the only one involving the type of picture: an interaction between type of picture and number of presentations, $F(2, 48) = 3.83$, $MS_e = 1.56$. The means are shown in Figure 4. As can be seen, the slope of the func-

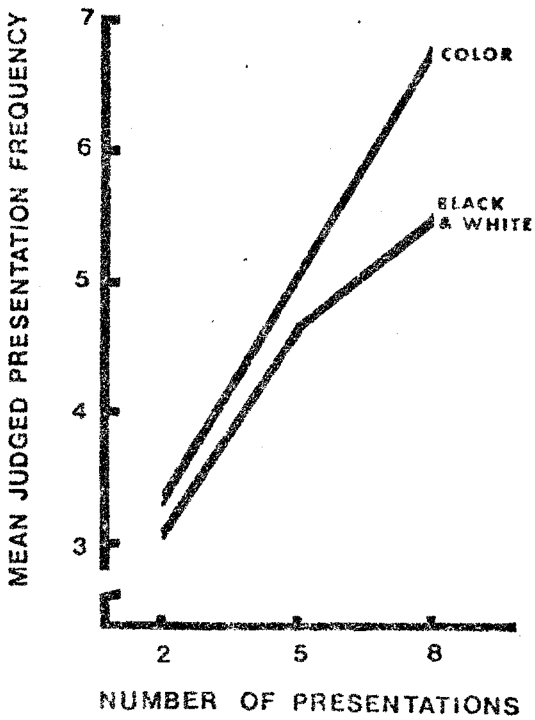

Figure 4. Judged presentation frequency as a function of the number of presentations.

tion relating judged number of presentations to actual number of presentations was steeper for the color pictures. Although not the major focus of the present study, this interaction suggests that discriminations among the relative frequency of external events may be better for more complex events.

Recall. The overall mean recall was 15.36. Mean recall as a function of number of image trials and number of presentations is shown in Table 1. Recall appeared to increase with number of presentations, but this effect was not great, $F(2, 48) = 2.43$, $p < .10$. In a previous study (Johnson et al., 1977, Experiment 2) with word stimuli, this same tendency ($p < .10$) was found. In the procedure used there and in the present study, items were presented in random order and were generated to cues in random order; there is little reason for the subject to note or generate interitem relationships. This may account for the relatively weak relationship between presentations and free recall, assuming that organization is the major factor in free recall.

In the analysis of the recall scores, the only significant effects were the main effect of number of image trials, $F(2, 48) = 11.00$, $MS_e = 1.13$, and the Number of Image Trials × Number of Presentations interaction, $F(4, 96) = 9.24$, $MS_e = .77$. In general, recall increased with number of image trials. This effect was most systematic for the highest presentation value (see Table 1). Image trials are, in effect, a type of covert recall test. Although there is some question about the underlying mechanism, the beneficial effect on later recall of previous recall trials is not surprising (e.g., Allen, Mahler, & Estes, 1969; Hogan & Kintsch, 1971; Thompson, Wenger, & Bartling, 1978). Interestingly, this effect was not found in Johnson et al.'s (1977) Experiment 2, in which words were used. The fact that there were no differences on the final recall test involving good and poor imagers suggests that the difference between good and poor imagers was not simply that good imagers gave more attention or effort to the stimuli and tasks.

Discussion

The fact that image trials inflated estimates of external frequency of pictures replicates previous findings that overt and covert recall trials inflated estimates of external frequency of words (Johnson et al., 1977). This is in contrast to the absence of effects on frequency judgments when subjects form images on study trials (Ghatala, Levin, & Wilder, 1973). One important difference between the present procedure and procedures used in other studies is that here the critical

Table 1
Mean Number of Items Recalled, Experiment 1

No. presentations	No. image trials: 2	5	8	Total
2	1.29	1.89	1.43	4.61
5	1.64	1.18	2.32	5.14
8	1.21	1.75	2.65	5.61
Total	4.14	4.82	6.40	15.36

internal generations occurred some time after the stimulus was presented rather than during or shortly after stimulus presentation. Similarly, increments in frequency judgments have also been obtained when uncontrolled covert rehearsals of earlier items are probably occurring during study (Proctor & Ambler, 1975; Shaughnessy & Underwood, 1973). It seems reasonable to suppose that the maximum effect from regenerating an event will be obtained once the initial event has left consciousness or working memory.

With respect to qualitative characteristics of covert regenerations, one of the initial predictions—that the difference between the magnitude of the IFE effect for good and poor imagers would be greater for color pictures—was not supported by these data. In retrospect, this was probably not a very powerful manipulation. Perhaps a better strategy would have been to contrast stimuli that differ more markedly in the ease with which they can be imagined accurately (e.g., a square vs. a clown).

However, the major prediction was supported. The frequency estimates of good imagers were more influenced by the number of times they imagined the pictures than were the frequency estimates of poor imagers. Of course, it is not unlikely that simply presenting the names of the pictures on image trials would affect estimates of picture frequency. Presumably, the pictures and the words both activated some of the core features that define a particular concept or event. However, if this were the only source of the IFE effect, good and poor imagers should have produced similar judgment functions, since they had equal experience with the names of the pictures. Additionally, the fact that there was no difference in the effect of number of presentations on the frequency estimates of good and poor imagers suggests that the two types or subjects do not just generally differ in the ways they acquire and use frequency information. Therefore, the obtained results strongly suggest that the accuracy of internal regenerations of external events is important in producing confusion between the two. It should be noted that accuracy could, for example, refer to a match between two "pictorial" memories (the perceptual trace and the imaginal trace), a match between two "propositional" memories, or a match at some even more abstract level between the content of two memories represented somewhere in somewhat different formats.

Experiment 2

Experiment 2 was addressed to the possibility that the differences in accuracy of imaginations of good and poor imagers did not produce the differences in their frequency estimates, but that these differences were the result of a variable correlated with accuracy. Accuracy and consistency of imaginations may be confounded. That is, if an image is an accurate representation of a picture on successive trials, the representation is probably consistent from trial to trial. Furthermore, some minimum number of activations of a particular imagination may be required before the memory of it is sufficiently strong to act as an externally derived representation. Although poor imagers might be inaccurate but consistent, a more likely hypothesis is that poor imagers are also less consistent in the internal representations they generate. For example, a poor imager, in trying to improve the quality of an image, might have been more likely to change it on successive trials. Thus, it may not have been greater accuracy of good imagers, but rather (or also) greater *consistency* of their internal representations that was important in producing the results obtained in Experiment 1.

Experiment 2, therefore, was an attempt to investigate the relationship between consistency in internally generated representations of external events and judgments of external event frequency. To avoid the problem of confounding variability of successive imaginations and similarity between imaginations and the external event, words were presented as external events and subjects were asked to generate images for the referents of the words. Half of the subjects were asked to generate the same image for an item on each generation trial, and half were asked to generate a different image for an item on each generation trial. Since the external events were printed words and the internal events were idiosyncratic visual images of referents for the words, it would be difficult to argue that any image was more similar physically to the external event than any other.

Thus, the major question of Experiment 2 was whether the degree of consistency of representations on imagination trials affects frequency judgments and, if so, whether such an effect provides a rival to the accuracy hypothesis in explaining the outcome of Experiment 1. In addition, Experiment 2 explored the effects of consistency on both delayed and immediate judgments of situational external frequency and provided evidence regarding changes in the IFE effect over time. Of course, the variability in images generated by subjects instructed to generate different referents for words in Experiment 2 was assumed to be much greater than the hypothesized variability in poor imagers' images in Experiment 1. In addition, it was not necessarily expected that the variability would be of exactly the same nature. However, the present procedure was adopted to maximize the opportunity to observe effects attributable to variability in completely covert generations.

Method

Materials and design. The materials were derived from those used by Johnson et al. (1977). Thirty-six items were formed by using the names, or shorter paraphrases, of 36 Battig and Montague (1969) categories as cues and one high-frequency instance from each category as to-be-remembered items. On study trials, cue-item pairs were presented; on image trials, cues alone were presented and subjects were asked to generate an image ("mental picture") of a referent for the target item. As a cover task, all subjects rated the vividness of their images on image trials. Subjects assigned to the image-same condition were asked to generate the same image for a target each time it was imagined. Subjects assigned to the image-different condition were asked to generate a different image for a target each time it was imagined. Within subjects, three levels of presentation frequency (2, 5 or 8) were combined factorially with four levels of image trials (0, 2, 5 or 8). The zero-image condition had not been included in previous studies because relative judgments rather than absolute accuracy was of interest. However, at this point we thought it important to establish that internally generating items inflates estimates of event frequency relative to a zero condition.

Table 2
Mean Judged Presentation Frequency, Experiment 2

Item	No. presentations: 2	5	8	*M* total
Immediate	5.69	7.25	8.23	7.06
Delay	4.73	5.41	5.75	5.30
M total	5.21	6.33	6.99	—

Three items were randomly assigned to each of the 12 combinations of presentation and image frequency and, as described in Experiment 1, were ordered randomly on blocks of presentation and test trials. There were four such sets of materials assigned randomly to subjects. On any particular block of presentation trials, 18 cue-item pairs were presented, and on any particular block of image trials, subjects were asked to generate images of from 11 to 16 different items. Ten presentation blocks alternated with 10 image blocks. Following this, a tape of the 36 target items (without cues) was played and subjects were asked to write a number corresponding to the number of times the word had actually been presented on study trials. Subjects in the immediate test conditions performed this judgment task in the same session (immediately following the above procedure); subjects in the delayed test conditions performed the judgment task when they returned after 1 week. To avoid subject-selection effects, all people signed up for and returned to a second session.

Procedure. Subjects participated in groups of up to four. Instructions emphasized the importance of following the imagery instructions, and subjects were not warned that there would be a frequency judgment task. Materials were presented via mimeographed booklets. The rate was 2 sec on study and 3 sec on test trials. Five sec per item were allowed on the frequency judgment task.

Subjects. Twenty introductory psychology students were assigned randomly to each of the two image conditions (same and different), and a random half of subjects in each condition were tested immediately and the other half were tested after a 1-week delay.

Results

For each subject, the mean of the judgments at each combination of presentation and image frequency was computed. Table 2

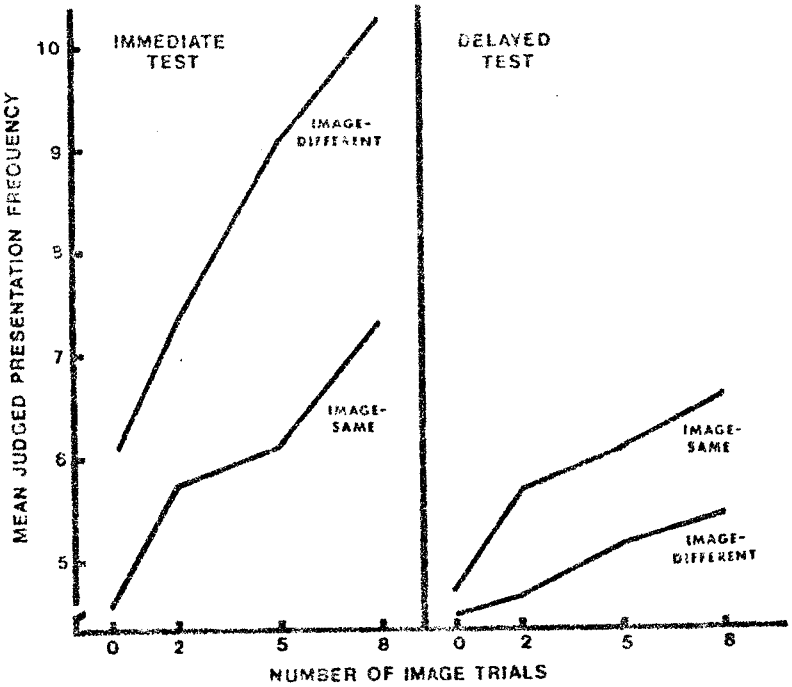

Figure 5. Judged presentation frequency as a function of the number of image trials. (The data are collapsed across presentation frequency. Subjects tested immediately are shown on the left, and subjects tested after 1 week are shown on the right.)

shows the mean of these frequency judgments as a function of number of presentations and retention interval, collapsed across image condition and number of image trials. Both the main effect of presentation frequency, $F(2, 72) = 77.61$, and the interaction of presentation frequency and retention interval, $F(2, 72) = 13.96$, $MS_e = 1.66$, replicate previously reported findings that peoples' sensitivity to the relative frequency of external exents declines over time (Underwood et al., 1971).

Figure 5 presents mean judged presentation frequency as a function of the number of image trials and the image condition, collapsed across presentation frequency. First, judgments increased with the number of image trials, $F(3, 108) = 69.43$, $MS_e = 2.10$, extending the previously reported (Johnson et al., 1977; the present Experiment 1) influence of covert thoughts on estimated external event frequency to the types of thoughts investigated here. In addition, there was an interaction between the number of image trials and retention interval, indicating the influence of internally generated occurrences on judgments of external event frequency declines with time, $F(3, 108) = 9.28$.

Of primary interest with respect to interpreting the results of Experiment 1 was whether the image-same condition would show more impact than the image-different condition from imagining the stimuli. As can be seen from the left panel of Figure 5, the outcome was the opposite of that predicted by the consistency explanation for Experiment 1. In the immediate test, there was a significant interaction between imagination instructions and number of image trials, $F(3, 54) = 2.88$, $MS_e = 2.92$. Subjects who varied their image of an item showed a greater IFE effect than subjects who maintained a consistent image. One possibility is that the subjects in the image-different condition, to be sure of generating a new image on any particular image trial, reviewed their past images more than did subjects in the image-same condition; such a review process would increase overall fre-

Table 3
Mean Judged Presentation Frequency, Experiment 2

No. presentations	No. image trials 0	2	5	8	*M* total
2	3.61	5.00	5.47	6.78	5.22
5	5.10	5.91	6.63	7.69	6.33
8	5.85	6.57	7.71	7.83	6.99
M total	4.85	5.83	6.60	7.43	—

quency of images and thus could increase frequency judgments.

Interestingly, however, there was also a significant interaction between image condition and retention interval, $F(1, 36) = 4.31$, $MS_e = 77.16$. This interaction reflects the fact that whereas events seemed more frequent to the image-different condition in the immediate judgment test, events seemed less frequent to them in the delayed test. There would be no reason to expect an interaction of this type if the results of the immediate test were produced by a review process in the image-different condition. As discussed below, this interaction seems more consistent with the assumption that the groups differed in the variability of their images.

There was also an interaction between number of presentations and number of image trials, $F(6, 216) = 3.10$, $MS_e = 1.20$. The relevant means are shown in Table 3 and suggest that the interaction was produced by a ceiling effect on judgments operating at the highest combination of presentation and image frequencies (the 8–8 condition). This interaction does not change the general picture that increasing the number of image trials increased presentation judgments at all levels of presentation frequency.

In general, the estimates of presentation frequency were higher in Experiment 2 than in Experiment 1. This could have occurred for any one of a number of reasons (e.g., the subjects were not drawn from the same pool, the materials involved different concepts of not necessarily equivalent background frequency, concreteness, etc., or there may have been differences in the estimations of the appropriate anchor values for the frequency judgments in the two cases). As is discussed later, we think one of the more interesting possibilities is that it is harder to separate memories of images from memories of perceptions in the case of words than in the case of pictures; if so, this might in itself tend to produce higher overall estimates of frequency of occurrence for words than pictures, all other things being equal. Of course, the present data do not allow us to decide among these possibilities.

Discussion

If the major difference between good and poor imagers in Experiment 1 were the greater variability in the poor imagers images, instructing subjects to make different images across trials in Experiment 2 should have reduced the effects of imagination trials relative to instructing subjects to make the same image across trials. In the groups tested immediately (and thus comparable in retention interval to Experiment 1), just the opposite was found; judgments of the image-different subjects were more affected by number of generations than judgments of the image-same subjects. Therefore, it is clear that the pattern of results in Experiment 2 does not support the alternative hypothesis that differences in the variability in imaginations of good and poor imagers produced the interaction between type of subject and number of image trials in Experiment 1. Thus, the accuracy hypothesis remains viable.

It would be premature to attribute the results of Experiment 1 to purely perceptual characteristics of the internal event. It could be that the good imagers had better abstract or propositional knowledge of the content of the pictures, rather than better pictorial representations (e.g., see Pylyshyn, 1973). But regardless of how the content of memories is best represented, these results indicate that one factor in producing confusion between memory representations of perceptions and the memory representations of imaginations may be the faithfulness with which an imagination matches the memory of original perceptual experience. We may fairly often confuse thoughts with imaginations of relatively simple events, such as whether we imagined a particular statement during a conversation or actually heard it. One thing that may protect us from more debilitating confusion is the generally schematic and nondetailed character of our imaginations. Perhaps we normally do not confuse fact and fantasy too much because we generally do not or cannot generate faithfully the complexity of many experiences.

Together, these two experiments show that variability as well as accuracy in imaginations affects judgments of external event frequency, perhaps (as discussed later) by somewhat different mechanisms. Given this, insofar as good imagers imagine more accurately and poor imagers imagine more variably, the magnitude and direction of differences in frequency judgments between the two may depend on whether accuracy or variability is playing the larger role in a given situation.

There is probably no single basis for the operation of frequency information (either appropriate or inappropriate). For example, frequency estimates of an event could be based on the general strength or availability of a general memory representation, on a count of memory traces of individual occurrences, or on both (see Hintzman, 1978, p. 375; Hintzman & Block, 1971; Howell, 1973; Tversky & Kahneman, 1973). We can think of the memory representations of two occurrences as containing both information that is central to defining these two as the *same* event and more peripheral (e.g., contextual) information (Johnson et al., 1977). Accuracy, then, would refer to overlap in this core information. Repetitions might increase the strength or availability of this core or defining information and thus influence frequency estimates. Variability in occurrences, assuming some minimal overlap in core information, should increase the general discriminability of separate traces and increase frequency estimates if the subject attempts to count occurrences. In addition, the minimal overlap required for an occurrence to contribute to the availability of the core and/or to be counted may vary with the nature of the event; thus, the respective roles of accuracy and variability would vary.

For example, more complex or more concrete events might require more overlap or accuracy in repetitions compared to more abstract or symbolic events. It seems reasonable to assume that words are important largely in their symbolic function; perhaps perceptual processing of them involves the activation of a fairly wide range of meaning. On the other hand, pictures are relatively more important in specifying a particular representation and perhaps result in a less generalized core representation in memory. Therefore, any repetition involving some deviation from the original memory would have more chance of overlapping with some of the core features of the original experience in the case of words than in the case of pictures. Thus, other things being equal, we might expect, on the grounds of accuracy or overlap, that the IFE effect would be greater when the external events are words as opposed to pictures. While comparisons across the present experiments must be tentative, since the concepts used were not the same in the two studies, the results suggest that imaginations are more easily confused with words than with pictures. (Compare the slopes in the left half of Figure 5 with those in Figure 2.) This would also account for why accuracy apparently outweighed variability in Experiment 1, in which pictures were used.

Subjects too may use different information and different decision strategies in different situations. For example, when the retention interval is short and the availability of separate traces (e.g., contextual information) is high, a subject might be more likely to "count" traces. In this case, variability in the representations would produce larger estimates. After a delay, however, individual traces might be less discriminable, and the subject might then respond by giving higher frequency estimates to items that elicit a single strong representation or elicit some representation quickly. In this case, consistency should lead to higher estimates. This type of change in decision strategy would be consistent with the results from the two retention intervals in Experiment 2.

In general, the outcomes of these experiments show that qualitative characteristics of thoughts influence their impact on estimates of the frequency of external events.

Reference Note

1. Johnson, M. K. *Constructive aspects of memory: Historical antecedents.* Paper presented at the annual meeting of the American Psychological Association, Chicago, September 1975.

References

Allen, G. A., Mahler, W. A., & Estes, W. K. Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 463–470.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, *80* (3, Pt. 2).

Ghatala, E. S., Levin, J. R., & Wilder, L. Apparent frequency of words and pictures as a function of pronunciation and imagery. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 85–90.

Hintzman, D. L. Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, *80*, 139–145.

Hintzman, D. L. *The psychology of learning and memory.* San Francisco: Freeman, 1978.

Hintzman, D. L., & Block, R. A. Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 1971, *88*, 297–306.

Hogan, R. M., & Kintsch, W. Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 562–567.

Howell, W. C. Representation of frequency in memory. *Psychological Bulletin*, 1973, *80*, 44–53.

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., Taylor, T. H., & Raye, C. L. Fact and fantasy: The effects of internally generated events on the apparent frequency of externally generated events. *Memory & Cognition*, 1977, *5*, 116–122.

Perky, C. W. An experimental study of imagination. *American Journal of Psychology*, 1910, *21*, 422–452.

Proctor, R. W., & Ambler, B. A. Effects of rehearsal strategy on memory for spacing and frequency. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *5*, 640–647

Pylyshyn, Z. W. What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 1973, *80*, 1–23.

Shaughnessy, J. J., & Underwood, B. J. The retention of frequency information for categorized lists. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 99–107.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4*, 210–221

Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 1973, *5*, 207–232.

Underwood, B. J., Zimmerman, J., & Freund, J. S. Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology*, 1971, *87*, 149–162.

Received November 20, 1978 ■