Reality Monitoring in the Context of Introductory Physics Demonstrations

Jonathan D. H. Gaffney

Nov. 30, 2006

Submitted in Partial Fulfillment of the Requirements of Cognitive Processes

Course Number Psy 508 at North Carolina State University

Physics demonstrations have long been used in many classrooms as a tool to both enhance student interest and help teach science. The instructive value of these demonstrations as perceived by instructors is nicely represented by this quote, attributed to a high school physics teacher:

> I think that average and good students should, a week later, be able to close their eyes and mentally run through what they saw happen and related it to a question asked and then come up with the answer by seeing in their mind what was demonstrated. (Roth, McRobbie, Lucas, & Boutonné, 1997, p. 517)

However, research shows that demonstrations fail to facilitate student understanding of specific physics concepts (Halloun & Hestenes, 1985; Roth et al., 1997). The root of the failure is difficult to identify because of the various processes involved in learning from an observation.

One viable approach is to blame the lack of active engagement that is inherent: students traditionally sit back and watch these demonstrations as spectators. When asked to predict the outcome of the experiment, students' ability to both remember the correct outcome of an experiment and explain it on a semester-end exam improved significantly. When the students discussed the experiment and their predictions with peers, they achieve another significant gain in both areas (Crouch, Fagen, Callan, & Mazur, 2004).

Another, perhaps even more fundamental, possibility is that students' memories for the demonstrations are not veridical. Rather, students may be imagining potential variations on the demonstrations, later falsely attributing the imaginations to external reality. In the study by Crouch et al. (2004), students who saw the demonstration without any active engagement were only able to recall the outcome of the experiment 70% of the time, as opposed to students who had never seen the demonstration, who could correctly indicate the outcome 61% of the time. This small gain in performance was not statistically significant and indicates that simply watching the demonstration does little to make a lasting impression on students' memory. The imagination of the students who had never seen the demonstration was nearly as accurate as the memory of those who had actually seen it.

The processes by which an individual distinguishes between events that are internal to the person (as in the case of imagination) or external to that person are collectively known as *reality monitoring* (Johnson & Raye, 1981). In a more general sense, the processes by which we encode and later distinguish the source of information is referred to as *source monitoring* (Johnson, Hastroudi, & Lindsay, 1993). According to the source monitoring framework, memories that originate in perception usually contain more perceptual detail such as color, texture, and sound; more contextual detail such as the time and place of the encounter with the information; and more semantic detail. Memories that originate in imagination or reflection typically have more information that refers to cognitive processes that were done at the memory's inception (Johnson, 1997). These cognitive processes are a product of the effort that was required of the brain to create the memory.

The differences between perceptual and cognitive information are capitalized upon by the judgments we make to determine the origin of the memory. Johnson et al. (1993) provide a paradigm for considering these judgments. Upon recalling an event, we rely on answers to two different kinds of questions whose purpose is to distinguish between internally and externally generated memories. The quicker checks that are more affiliated with familiarity and perceptual detail are called *heuristic* checks; such checks make rapid determinations as to whether, say, a person's face is familiar or a word appeared on a list. For example, when a single choice jumps out at us on a multiple-choice test question, that item has passed the heuristic check for familiarity. An event passes the heuristic check and is deemed to be external from us when the amount of perceptual detail affiliated with it is above some threshold that is determined by the risks and rewards associated with false alarms and misses.

Events that fail to cleanly pass the heuristic check are subject to the slower, more explicit, *systematic* check. Systematic questions attempt to check whether the memory is reasonable given other information we have available to us. For example, a systematic check employed to determine whether Lara or Eric had told a joke yesterday would succeed in attributing the joke to Eric on the grounds that Lara was in another state yesterday. Like heuristic checks, the criteria for passing systematic checks are determined by perceived risks and rewards. If the penalty for misattribution is particularly strong, we would be careful to make sure the event passed both checks before committing to a decision.

Together, heuristic checks and systematic checks allow us to make reasonable judgments about the origin of memories. Furthermore, they can be used to determine whether a memory originated from a perceived event or from an imagined one. In the case of reality monitoring, they tend to focus on the cognitive operations that are characteristically associated with imagination (Johnson et al., 1993). When it is effortful to imagine an event, we tend to link that imagination with the specific work that we did while encoding it. Therefore, our ability to reliably use reality monitoring depends on the strength of these cognitive links. In a study performed by Finke et al. (1988), participants were shown both solid and incomplete figures. They were told to imagine completing the partial figures by reflecting the object either across the horizontal or vertical axis. Because it was easier to create a complete image by flipping a partial figure across a vertical axis, the participants attached fewer cognitive operations to that image compared with one obtained by flipping an incomplete figure across a horizontal axis. As a consequence, participants were able to more accurately differentiate memories of a perceived whole figure from an imagined complete figure in the horizontal trials as opposed to vertical trials (Johnson, 1997). In a related study (Johnson, Raye, Foley, & Foley, 1981), source monitoring errors were prevalent when experimenter-provided cues precisely determined the word generated by the participant.

The vividness of a memory also plays a crucial role in our assessment of its origin. If the vividness of a perceived event decreases or the vividness of an imagined event increases (say, by continually thinking about it), the confusion as to the origin of the event will be increased. A study (Hashtroudi, Johnson, & Chrosniak, 1990) found that while ratings of visual clarity of real events were greater than those of imagined events initially, both lost clarity over time unless they were actually thought about. Therefore, an effect of selectively drawing imagined events from memory is that they would eventually rival memories of real events in terms of vividness and clarity. Further complicating matters, rehearsal in bringing the memory from long term to working memory may actually reduce the cognitive associations with the imagination memories, weakening the reality monitoring checks (Johnson, 1997).

Indeed, there has been evidence for a superiority effect for self-generated materials. For example, studies that tested participants' memory of generated words indicate that recall and recognition scores are better than memory of words presented to them (Slamecka & Graf, 1978). This "generation effect" is due to the actual process of generation, not to overt cues such as hearing one's own voice or seeing one's own handwriting (Johnson et al., 1981). Further experiments with frequency judgments of perceived and imagined items have shown that when subjects imagine words or pictures repeatedly, they report that they were presented with the item more frequently than they actually were (Johnson, Raye, Wang, & Taylor, 1979). Taken together, these studies indicate that we are more likely to remember events if we self-generate them, either in our imagination or in our re-representation of those events (Johnson & Raye, 1981). As previously noted, however, these self-generated memories tend to become more vivid and less cognitively attached, leading to more potential for source memory confusion.

Additionally, whenever a new event mistakenly becomes identified as an old event, there is a bias that is sometimes called the "it-had-to-be-you effect" (Johnson et al., 1993). In other words, when we recognize an event or a piece of information that should be new to us, we are quick to judge that that event was initially external in origin. For example, in experiment 3 of Johnson et al.'s study of participants' source memory (1981), participants heard cues and either generated words (labeled as S) or were told words (labeled as E) to associate with the cues. Later, they were told to indicate whether a word shown to them was one that they had previously seen (either S or E). If they indicated that such a word was old, they had to discriminate whether it was S or E. Johnson et al. found that when participants falsely indicated that a word was old, they indicated preferentially that the word was an E-type word.

This experiment intends to quantitatively show two consequences of reality monitoring failure. First, when students misidentify portions of a demonstration as having been seen or imagined before, they attribute it to a part of the real demonstration rather than to imagination. Second, when students correctly identify potions of a demonstration as having been seen or imagined before, they show reality monitoring confusions if they imagine extensions of that demonstration that are similar to the aspects of the demonstration that they had been shown.

METHODS

Participants

Ninety-nine undergraduates at a large university responding to advertisements in flyers and newspapers would be chosen to participate in this study. To participate, each of the participants will be required to have completed one semester of introductory physics in mechanics. The participants will be compensated for their time.

Equipment

For this study, a studio-style classroom will be used. In this classroom, ninety-nine Macintosh iBook computers are kept on eleven round tables of a seven-foot diameter. Each of the nine students at a table would have access to one of these machines. Students will be given individual earphones.

Design

All participants will be shown a video that consists of a series of 10 clips, complete with sound, on his or her computer screen. Each video clip is a sequence of two parts of a mechanics demonstration that has at least four possible parts. Immediately following those two parts, the student will receive instruction to pause the video and imagine a possible third part. In half of the clips, the participant will be told to imagine a demonstration very similar to the parts already shown. In the other half, the participant will be told to imagine a different demonstration. This will be varied between participants so that half of the participants imagine a similar part to each demonstration. After generating this imaginary event, the participant will be told to record the vividness of his imagination on a scale from 1 to 7, akin to studies of this kind (Johnson et al., 1979). This process will be followed for each of the 10 video clips.

After a period of one week passes, the participants will be gathered in the studio classroom once again for a recognition test. They will be shown a series of 10 videos clips that consist of three parts each. After seeing each part, students will be instructed to indicate whether the part of the demonstration they had just watched was “E,” meaning that the part had been presented to them by the experimenter on video in the earlier session, “S,” meaning that the part was self-generated in the earlier session, or “N,” meaning that the part was completely new. The letter codes (E, S, N) and procedure are similar to the Johnson, et al. study (1981) investigating similar effects in word lists.

There are two independent variables. The first is a variable between participants: imagined demonstration parts were either similar or different to presented parts. The second is a variable within each participant: video clips were either experimenter-presented or self-generated. Accuracy on the recognition test is reflected in the dependent variables.

Stimuli

Ten distinct mechanics demonstrations, each containing at least four parts, will be videotaped as separate parts. The demonstrations will cover the range of topics within a typical introductory course. For example, one demonstration is of an instructor sitting on a chair with a bicycle wheel (Roth et al., 1997). When the instructor turns his hands one way, he turns counterclockwise. When he turns them the other way, he spins clockwise. Two similar parts involve the bicycle wheel spinning the other direction. A dissimilar part would be the instructor throwing the wheel forward and rolling backward in the chair. This part is different not only due to surface features but also deep physical meaning; it demonstrates conservation of linear momentum rather than conservation of angular momentum. Participants will be shown two of the four similar parts and will be told to imagine either a similar or a dissimilar part.

For the recognition task, three of the (at least) four distinct parts of the demonstration will be combined in random order to make each of the video clips. Each of the parts to be shown in the recognition test is similar. That is, while participants may have been told to imagine a dissimilar part, they will never be shown a dissimilar part for recognition purposes. These 10 clips will also be randomly ordered for each of the participants.

Procedure

All participants will be tested together. However, they will be isolated from each other in the sense that each will have his or her own video screen and audio signal. Instructions will be handed out before the participants begin their videos, and these instructions will contain no warning about possible memory tests. The "cover task" is that the participants are to report vividness in imagination. After each of the 10 clips, participants will be told to imagine either a similar or different demonstration part. The computer display will pause for 30 seconds, after which a prompt will appear for the student to rank the vividness of his or her imagination on a 7-point Likert scale. Each participant will be given a 5-minute break halfway through.

During the recognition test, a prompt will appear on the screen after each part of the video. Participants will be given five seconds to indicate the part's origin as E, S, or N. Again, participants will be given a 5-minute break halfway through to combat fatigue.

RESULTS

In the recognition task, a miss is defined as calling either an E or an S item "new." False positives occur when the participant identifies a new clip as either E or S. For an item to be classified as a hit, the participant merely needs to correctly make the determination as to whether it is "new" or not. Correct determination of E or S is dealt with separately as source monitoring.

One important result to be determined lies within the items classified as false positives. We expect a significantly higher number of false positives to be classified as "E" as opposed to "S." The motivation for this is clear from the "had-to-be-you effect:" when participants incorrectly indentify a new item, they have been shown to preferentially claim that the experimenter originally presented it (Johnson et al. 1981).

We will also pay attention to the misses. While there is no clear prediction here, we might expect that if the imaginations are not sufficiently vivid, participants will misidentify more S items than E items. We will run a *t*-test here to determine whether the similar or dissimilar parts are more likely to be missed. If we see something here, that will be an interesting opportunity for further research.

In the case of hits from the recognition test where an item is properly identified as not being new, we expect to find a large source confusion effect in situations where participants had been asked to imagine a related part of a demonstration. Thus, we expect that the proportion of incorrectly identified similar video parts to be statistically significantly higher than when dissimilar parts are imagined. We will perform a *t*-test to determine whether the difference between the two groups is significant.

Furthermore, there should be large confusion in the similar imagination category. Such an effect would reveal itself as the proportion of correct identifications being near chance performance. In prior studies, confusion due to similar imaginations often showed only small effects (Lindsay et al., 1991). However, because of the complexity of this study and the delay between exposure and testing, we expect to detect a stronger effect.

DISCUSSION

When reality monitoring fails, it may lead to confabulatory false recall effects. After repeatedly thinking about a certain event, people may retrieve pieces of other actual events that may or may not be related (Schacter, Norman, & Koustaal, 1998). Then, when they are asked to recognize an event, people are likely to misattribute its source. Moreover, because PET scans have shown that the same brain regions are activated by visual imagery as by visual perception (Kosslyn et al., 1993), people may be recalling pieces of imagined events. In fact, when participants are explicitly instructed to imagine an event, the probability of false event recall was significantly increased (Hyman & Pentland, 1996). The overall implication is that when a person imagines an event, the likelihood that it will be remembered as an actual, perceived event rather than an imagination is increased (Garry, Manning, Loftus, & Sherman, 1996).

This confabulation, when combined with the "it-had-to-be-you effect," can display a false memory effect. Based on previous work (Deese, 1959), Roedinger and McDermott (1995) found that listing words that are closely related to a certain target word, called a "critical lure," causes reliable memory distortions. For example, when a participant listened to a listing of a dozen words about "sleep," he or she is very likely to recognize the critical lure "sleep" even though it was not on the list. Here, we can interpret this as the participant classifying the word "sleep" incorrectly as an old word. Then, falling prey to the "it-had-to-be-you effect," the participant associates the word with an external origin (the list) rather than an internal origin (his or her imagination). Similarly, students may confabulate part of a demonstration because other parts of the demonstration are related to some critical lure. If the student then imagines that critical lure, it may persist even if it is physically impossible.

Overall, the expected findings of this study would imply that due to source monitoring errors, students' memories of demonstrations are not veridical. While students are watching the demonstrations, they may be imagining possible extensions, which are often prompted by the instructor to provide more depth to the demonstration. However, if the student's imagination is incorrect but similar in form to what the instructor shows, it will lead to confusion later. Specifically, if the student attributes a memory of an impossible physics event (such as a rock hitting the ground before a feather when dropped simultaneously from the same height in a vacuum chamber) to a performed demonstration rather than to an imagination, the results would be obviously catastrophic, as the student would be convinced of an incorrect conception (for example, that acceleration of objects is proportional to their mass). To address this concern, one possible future work to investigate the persistence of imagined memory would be to present students with videos that have extensions that are not easily predictable and ask them to record their imaginations. After a period of instruction, where the incorrect predictions are explained and revised (but where the appropriate video clip is not shown), a recognition test similar to the one in this study is given to determine whether the imagination persists and continues to be confused with external stimuli.

Previous research has been done to show that the source monitoring effects discussed here not only occur with associated word lists but also for stories told by actors on video recordings (Lindsay, Johnson, & Kwon, 1991). This study extends those findings by adding visual action as well as conceptual depth to the video clips, which extends the ecological validity of reality monitoring. Future work along this vein would incorporate live demonstrations and even student-performed laboratory experiments. Specifically, students may be less prone to source monitoring effects when they are physically in contact with and in control of the demonstration (Lindsay et al., 1991). However, in certain cases, when imagined tasks are very similar to performed tasks, these effects may still be significant.

Certain procedures have shown some reduction in misidentification associated with poor reality monitoring effects. For example, Johnson et al. (1981) showed that writing down observations reduced source monitoring errors. However, Roth et al. (1996) found that during demonstrations, students rarely take notes about what they observe. They expect to remember it accurately and thus only record physical or mathematical information rather than the apparatus, initial conditions, or results. In the study by Crouch et al. (2004), the significant improvement that students showed in correctly explaining the outcome of a demonstration due to predicting and/or discussing that prediction with a neighbor might be a result of improved memory as much as anything else. Students in the "predict" and "explain" modes recorded their predictions prior to the performance of the demonstration. By doing so, they may have recorded information pertinent to recognizing the precise initial set-up and trial performed. In fact, the authors of that study even noted surprise that only a marginal difference in performance on the semester-end exam existed between the "predict" and "explain" groups.

If source monitoring is a significant cause of poor performance on examinations in physics based on demonstrations, we should see only small differences between those two groups, as the improvement in source memory comes more from recording external stimuli than discussing predictions, which may be nothing more than imaginations. Discussing predictions may encourage those imaginations to remain, as they are continually recalled into working memory and made more vivid (Johnson, 1997). The complication to this claim is that a well-established method for improving understanding of conceptually deep physics material is to elicit student ideas, confront errors, and offer students an opportunity to confront those errors (McDermott, 2001). Because of this potential conflict, it should be established whether source monitoring errors are affected by eliciting student predictions. If they are, processes should be developed that do less to reinforce inaccurate imaginations.

By studying reality monitoring in an arrangement such as this, we can begin to understand why comments such as that of the physics teacher at the beginning of the paper are misguided. It is likely that students simply do not have veridical memories of classroom demonstrations. If this study supports that claim, ways to improve student memories need to be sought if we expect classroom demonstrations to be effective teaching tools.

References

Crouch, C. H., Fagen, A. P., Callan, J. P., & Mazur, E. (2004). Classroom demonstrations: Learning tools or entertainment? *American Journal of Physics, 72*(6), 835-838.

Deese, J. (1959). On the preditcion of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*, 17-22.

Garry, M., Manning, C., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review, 3*(2), 208-214.

Halloun, I. A., & Hestenes, D. (1985). Common sense concepts about motion. *American Journal of Physics, 53*, 1056-1065.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging, 5*(1), 119-126.

Hyman, I. E., & Pentland, J. (1996). The role of mental images in the creation of false childhood memories. *Journal of Memory and Language, 35*(2), 101-117.

Johnson, M. K. (1997). Source monitoring and memory distortion. *Philosophical Transactions of the Royal Society of London, 352*, 1733-1745.

Johnson, M. K., Hastroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3-28.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.

Johnson, M. K., Raye, C. L., Foley, H. J., & Foley, M. A. (1981). Cognitive operations and decision bias in reality monitoring. *The American Journal of Psychology, 94*(1), 37-64.

Johnson, M. K., Raye, C. L., Wang, A. Y., & Taylor, T. H. (1979). Fact and fantasy: The roles of accuracy and variability in confusing imaginations with perceptual experiences. *Journal of Experimental Psychology: Human Learning and Memory, 5*(3), 229-240.

Kosslyn, S. M., Alpert, N. M., Thompson, W. L., Maljkovic, V., Weise, S. B., Chabris, C. F., et al. (1993). Visual mental imagery activates topographically organized visual cortex: Pet investigations. *Journal of Cognitive Neuroscience, 5*(3), 263-287.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52*, 297-318.

McDermott, L. C. (2001). Oersted medal lecture 2001: Physics education research - the key to student learning. *American Journal of Physics, 69*, 1127-1137.

Roedinger, H. L. I., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology, 21*, 803-814.

Roth, W.-M., McRobbie, C. J., Lucas, K. B., & Boutonné, S. (1997). Why may students fail to learn from demonstrations? A social practice perspective on learning in physics. *Journal of Research in Science Teaching, 34*(5), 509-533.

Schacter, D. L., Norman, K. A., & Koustaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49*, 289-318.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 4*(6), 592.