Imagination inflation is a fact, not an artifact: A reply to Pezdek and Eddy

MARYANNE GARRY, STEFANIE J. SHARMAN, KIMBERLEY A. WADE,
MAREE J. HUNT, and PETER J. SMITH
Victoria University of Wellington, Wellington, New Zealand

Pezdek and Eddy (2001) claim to prove that imagination inflation is a spurious effect caused by regression to the mean (RTM). They make four predictions about what patterns of data would demonstrate a genuine effect for imagination versus those that would be explainable by RTM. We review each of those predictions, and demonstrate significant problems with them. We conclude that imagination inflation is a genuine effect, and that Pezdek and Eddy's work has contributed to the growing research showing that when people imagine fictitious events from long ago, they become more confident that those false events were genuine experiences.

A Western traveler encountering an Oriental philosopher asks him to describe the nature of the world:

"It is a great ball resting on the flat back of the world turtle."

"Ah yes, but what does the world turtle stand on?"

"On the back of a still larger turtle."

"Yes, but what does he stand on?"

"A very perceptive question. But it's no use, mister; it's turtles all the way down." (Sagan, 1979, p. 293)

Like Pezdek and Eddy (2001), we begin our reply with an instructive quote. This well-known parable, here told by Carl Sagan, illustrates the dangers of building an argument by stacking faulty thinking on top of faulty thinking. Although Pezdek and Eddy do raise some valid points, in the end their criticisms of imagination inflation are a turret of turtles.

Let us begin with their very first sentence. Contrary to Pezdek and Eddy's claims, their paper does not address "the conditions under which false autobiographical events are likely to be planted in memory" (p. 707). Imagination inflation is not about implanting memories. Pezdek and Eddy still continue to assert that it is, even though we have stressed this point with them before (Garry & Loftus, 2000). Imagination inflation is an increase in confidence that a fictional or hypothetical event that was merely imagined was actually experienced. This point is not trivial, because other imagination inflation researchers have been careful to distinguish between imagination inflation and an implanted memory. For instance, in two different papers, Heaps and Nash (1999, in press) described imag-

ination inflation variously—and accurately—as a change in subjective likelihood; a false belief about the past; or a mechanism by which "imagination can serve to raise subsequent likelihood judgments" (Heaps & Nash, in press). Mazzoni (2001) has clearly captured the difference between confidence that an event actually happened and a memory for that event. We are all confident, they note, that our births occurred as our mothers have explained them to us, although we do not actually have memories for them.

When we change people's confidence that an event might have happened to them, have we affected their memory? We think the answer is yes, because we have changed the way they think about and report their past. But is that an implanted memory? No, because the imagination inflation effect does not meet the criteria others have proposed for implanting a false memory. Take Hyman and colleagues' (Hyman & Kleinknecht, 1999; Hyman & Loftus, 1998) proposal that implanting false memories involves three conditions. First, subjects must accept that the event is plausible. Second, they must construct details about the false event that include an image plus a narrative. Finally, subjects must misattribute the source of that information, wrongly ascribing the mental information to genuine experience rather than to their own imagination. Increased confidence that the target event did happen—the definition of imagination inflation—meets the first criterion, but there is no evidence that it meets the remaining two. Of course, this is not to say that some subjects who participate in an imagination inflation experiment do not develop false memories about the events they imagine, but our aim has been to examine only the change in confidence.

The title of the original paper, "Imagination Inflation: Imagining a Childhood Event Inflates Confidence That It Occurred" (Garry, Manning, Loftus, & Sherman, 1996), makes the definition of imagination inflation clear, and so does quoting Garry and Polaschek (2000) more accurately than Pezdek and Eddy did. Pezdek and Eddy wrote:

In summarizing the results of this experiment and several others using the imagination inflation paradigm, Garry and Polaschek (2000) concluded that "a growing body of literature shows that imagining contrary-to-truth experiences can change memory" (p. 6). Furthermore, "when people think about or imagine a false event, entire false memories can be implanted. Imagination inflation can occur even when there is no overt social pressure, and when hypothetical events are imagined only briefly." (p 708–709)

While Garry and Polaschek's (2000) review paper focused on imagination inflation, it did so in the context of a larger review on imagination and memory. In that larger context, the few specific lines Pezdek and Eddy quoted from the Garry and Polaschek abstract make sense. Thus, when the Garry and Polaschek abstract foreshadows that "when people think about or imagine a false event, entire false memories can be implanted," it refers to the opening section of the Garry and Polaschek paper, which reviews what we might call the "Lost in the Mall" genre of studies: Loftus (1993), Loftus and Pickrell (1995), and Hyman and Pentland (1996). Later on in the paper, when the focus is on imagination inflation, Garry and Polaschek refer only to the "effect of imagination on memory" (p. 7) and even define imagination inflation rather tightly as the "confidence-boosting effect of imagination" (p. 7).

Now that we know what effect we are talking about, let us turn to the heart of Pezdek and Eddy's argument. Put bluntly, they assert that Garry et al.'s (1996) results might not show any confidence-inflating effect of imagination at all. Instead, Pezdek and Eddy say these "findings could be explained as simply a case of regression toward the mean" (p. 709; hereafter RTM). Pezdek and Eddy put forth a number of arguments as evidence of their claim, but we believe each argument is flawed. Below we summarize Pezdek and Eddy's experiment, review their four predictions, and examine the strength of their arguments in light of the evidence. First, however, we begin with a basic overview of RTM.

REGRESSION TO THE MEAN: THE (VERY) SHORT COURSE

Every time researchers measure some construct, they unwittingly drag some error into their measurement. Take, for instance, the case each of us might face when administering an Introductory Psychology exam. There are students for whom error depresses their score, and students for whom error inflates their score. So what is the conscientious college professor to do? Suppose you gave the exam again, to the whole class. Of course, you would give a parallel form of the exam, but it would not be perfectly correlated with the first exam. What might happen? The end result would be that some students would find that their retest scores increased, while others would find that their scores decreased. The mean, however, would stay the same, as would the standard deviation and

the shape of the distribution in general. With each retest students would change and test scores would change, as long as tests were not perfectly correlated. In short, with subsequent tests, error would be shuffled around differently. The shuffling around would tend to work so that scores farther away from the mean moved closer to the mean on a retest. Instead of referring to this shuffling around under the curve by a tedious name such as "shuffling around under the curve," we call it RTM.

Several other interesting features of RTM are outside the scope of this article, and the reader wanting more than this (very) short course could do no better than to read Campbell and Kenny's (1999) book.

Before we turn our attention to Pezdek and Eddy's specific predictions, there is an important point to note. Throughout their paper, Pezdek and Eddy seem to mistake regression to the *mean* of the items with regression to the *midpoint* of the scale, and this confusion underpins much of their thinking.

PEZDEK AND EDDY'S PREDICTIONS

Pezdek and Eddy hypothesize that if imagination inflation is a real effect, we should expect to see the following pattern of results.

1. The residuals between actual posttest scores and RTM-adjusted (predicted) posttest scores should be greater than zero.

2. There should be a test occasion × condition interaction.

3. All mean likelihood ratings should increase from the pre- to posttest, not just those initially rated as 1–4.

4. There should be a time × imagination × age interaction.

Below we examine the logic behind each prediction, the data on which Pezdek and Eddy rely, and the various conclusions that can be responsibly drawn.

1A. The Prediction: Actual Versus RTM-Adjusted Residuals Should Be Greater Than Zero

Pezdek and Eddy argue that if imagination inflation is a real effect, posttest confidence for imagined events will be above and beyond the increase in confidence produced by RTM alone. To analyze their data according to this hypothesis, they did the following analysis.

First, they found the correlation between pretest and posttest scores for not imagined (control) target events, and used the correlation to estimate the effects of RTM. In other words, they constructed a regression line to predict the change in posttest confidence scores that are attributable only to RTM. Second, Pezdek and Eddy used this regression line to predict the posttest scores for the imagined events, again assuming that only RTM was responsible for changing these scores. Finally, they compared the predicted imagined posttest scores with the obtained imagined posttest scores to produce residual scores. Where the residuals are positive, there is undoubtedly an

effect over and above RTM. However, what does it mean when the residuals are not different from zero? Pezdek and Eddy argue that it means imagination has no effect, and that imagination inflation is nothing more than RTM disguised as imagination inflation. But they are wrong.

1B. The Turtle: Their Analysis Does Not Work

Contrary to Pezdek and Eddy's claim, their residualized change scores analysis is not a strong statistical basis for drawing any conclusion. There are two broad problems with the attempt to predict—and then correct for—RTM using residualized change scores. The first is a general problem: Several researchers have noted that the method is fraught with difficulties (Campbell & Kenny, 1999; Rogosa, Brandt, & Zimowski, 1982; Rogosa & Willet, 1985). In the end, Campbell and Kenny recommend that if the purpose of the research is to examine causes of change, then "it is inadvisable ever to use residualized change scores or estimated true scores" (p. 99). The second problem is specific to imagination inflation: Pezdek and Eddy's use of residualized change scores just does not work. Because this article is about imagination inflation in particular, let us concentrate on the second problem. We shall see that their formula actually *fails* to detect imagination inflation in some circumstances. Below, we illustrate this problematic situation.

Hypothetical situation: Failing to detect imagination inflation. Suppose we give a sample of people the same test on two different occasions. If there is no intervening treatment, no measurement error, and no other behavioral change among our subjects, then each person's pretest score and posttest score should be the same. In other words, pretest and posttest scores would be perfectly correlated ($r = 1$), as in Figure 1A. However, a more realistic scenario is that our measurements will include some error, and/or some of our subjects' behavior will change over time. Therefore, we actually expect a shift away from the perfect correlation line and toward the mean. Figure 1A shows how the predicted y line shifts toward the mean. The amount of regression can be estimated by looking at the distance between the perfect correlation line and this predicted line (Campbell & Kenny, 1999).

Suppose we now assume, in this hypothetical example, that imagination inflation has a consistent and genuine effect on confidence. Indeed, such an effect is Pezdek and Eddy's fundamental thesis, that imagination should increase all scores by the same amount, not just low scores. Such a constant effect would produce a parallel line above the predicted y line.[1] Let us consider the first of Pezdek and Eddy's recommended analyses, that a test occasion × imagination condition interaction is evidence of imagination inflation. In our hypothetical example, there will be a significant interaction. Now let us consider the second—and what Pezdek and Eddy claim is their most sensitive technique—the residualized change scores analysis. What will this analysis find? It will never be able to find the effect, because it is unable to detect imagination inflation when the shape of the distributions does not change from pretest to posttest. The effect becomes clear if we replot Figure 1A and use standardized scores to produce Figure 1B. What will happen to the plot? The regression and imagination inflation lines will, in fact, collapse on top of one another and both go through the y-intercept at zero. In short, no matter how large the effect size, the technique touted by Pezdek and Eddy as very sensitive to RTM is *insensitive* to imagination inflation if imagination inflation is a homogenous effect.

In sum, Pezdek and Eddy claim that their residual analysis "provides the strongest statistical basis for concluding that regression to the mean—not imagination inflation" (p. 715)—caused their results. However, we have demonstrated that when the shape of the distribution does not change from pretest to posttest, the analysis will not actually detect an effect.

2A. The Prediction: There Should Be a Test Occasion and Condition Interaction

According to Pezdek and Eddy, the mean change in confidence ratings for imagined events should be greater than the mean change for not imagined events, from pretest to posttest. A lack of interaction would signify no effect over and above RTM.

2B. The Turtle: An ANOVA Is Not Always the Best Way to Go

In some ways, the interaction between test occasion and imagination condition is intuitively appealing. Indeed, the difference between the imagined and not imagined change scores is what is evaluated in the interaction term: When it is different from zero, there is an interaction. However, the extent to which the analysis of variance (ANOVA) reveals an interaction depends not only on whether there is an interaction, but on other factors as well. It matters, for example, to what extent the data meet the ANOVA assumptions. When some assumptions are violated, the ANOVA will not reveal an effect that is there, but when other assumptions are violated, it might signal an effect that is not really there. For example, if the data are not normally distributed, the ANOVA will be less able to fit a model to the data, and it may not be sensitive to an effect should one exist. In imagination inflation research, target events are deliberately constructed so that the majority of subjects rate them low at pretest; thus, the data are skewed. Outliers pose an additional problem, as does a change in variance between pre- and posttest and across event(s). If the variance changes dramatically, the ANOVA copes by fitting a common, interpolated variance. If, for instance, one set of measurements has a large variance and the other has a small variance, the ANOVA assumes a moderate variance overall. This moderate assumption causes the ANOVA to assume a lower variance than is true in one of the subsets, which might produce false significance for effects in these parts of the data. In addition, it will assume a higher variance than
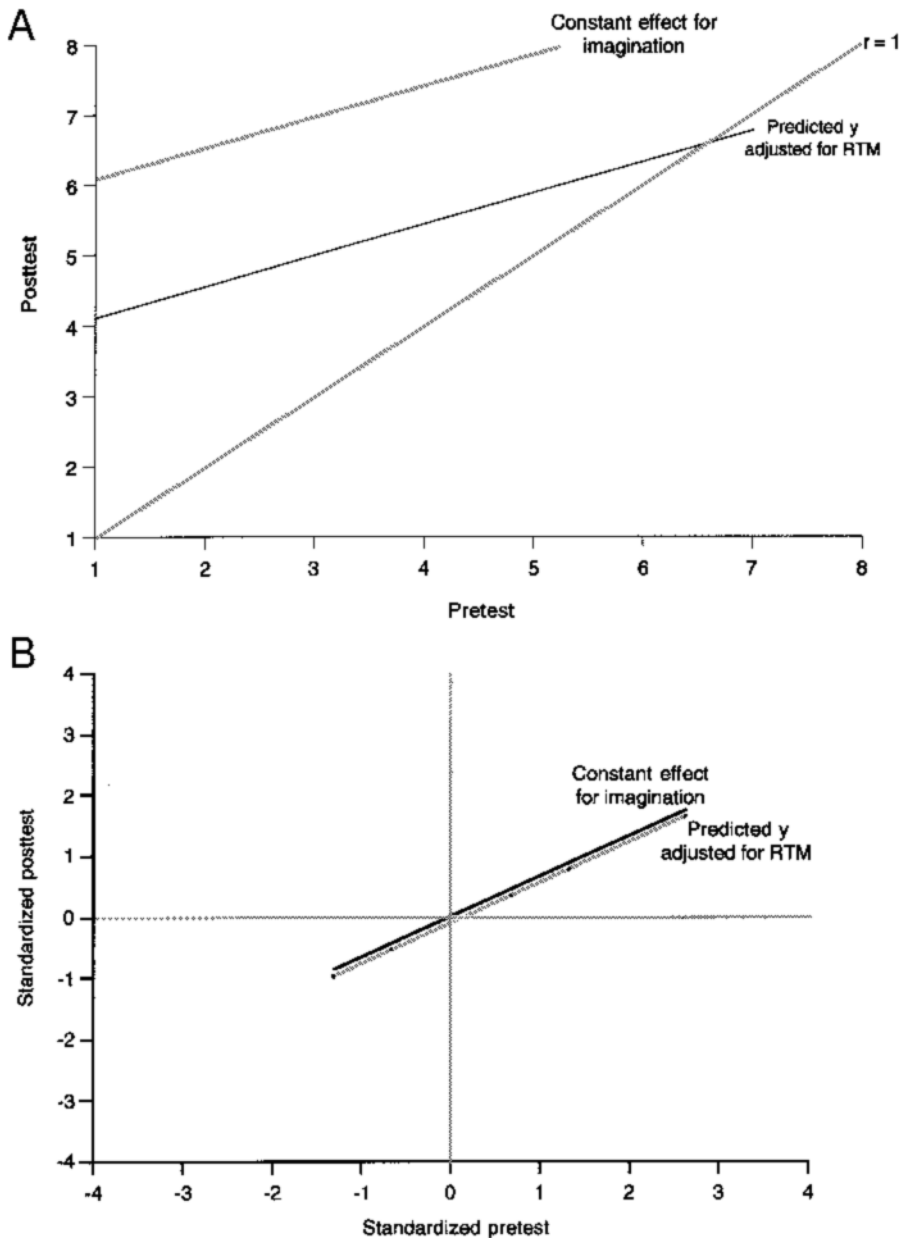
Figure 1. (A) Perfect correlation line, imagination inflation line, and regression line. (B) Figure 1A replotted with standardized scores.

is true in the other subset, which might fail to detect genuine effects found in the other parts of the data. Take, for example, Pezdek and Eddy's own Table 2, where their standard deviations vary by about a factor of 10. In short, evaluating the interaction term in an ANOVA might be a good choice in some situations, but not in others.

One solution to the violations of the ANOVA assumptions is the "proportion increase" analysis that Garry et al. (1996) used (reproduced below as Figure 2). It eliminates the problems caused by changing variance; across events, the standard deviations in Garry et al. varied by

a factor of nearly 5; thus, their choice of dependent measure (treating events as subjects) made the heterogeneity of variance across events irrelevant. As a reminder, Figure 2 shows the percentage of subjects inflating for each event, classified according to whether that event was or was not imagined. A paired t test treating events as cases showed that events were more likely to show inflation when subjects imagined them than when they did not imagine them [$t(7) = 5.48$, $p < .001$]. The effect size (J. Cohen, 1988) was 1.94.[2] Garry et al. relied mainly on a categorical dependent measure—the percentage of sub-
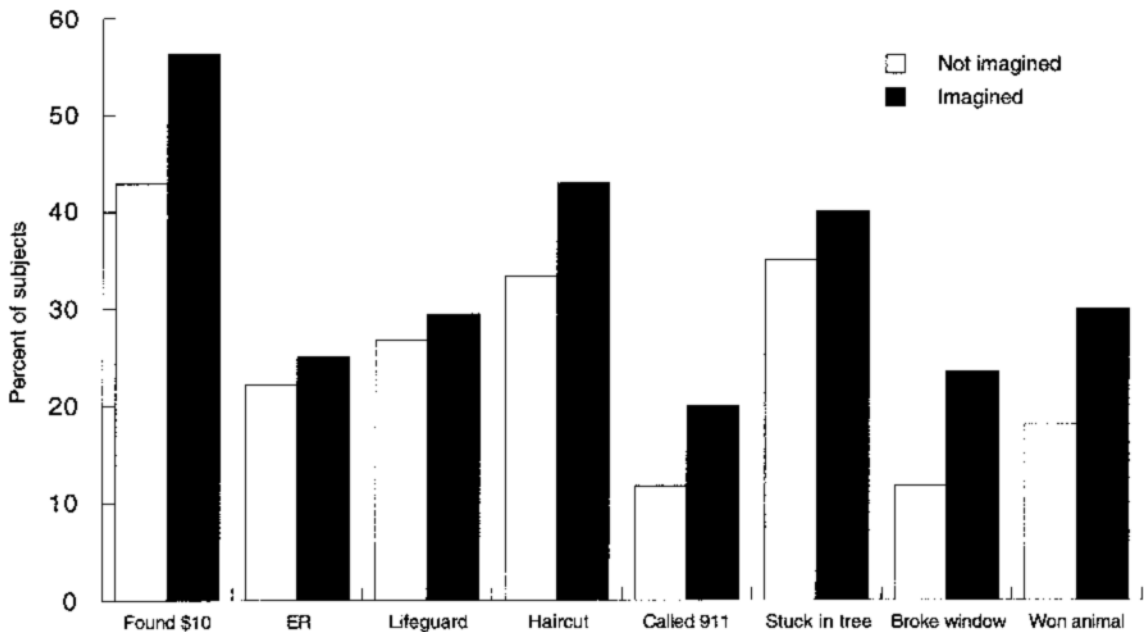
Figure 2. Percentage of subjects who increased on each item (adapted from Garry et al., 1996).

jects who inflated on each event—and then compared that categorical change using a parametric *t* test.

Pezdek and Eddy seem to argue that the pattern demonstrated in Figure 2 is somehow less interesting or meaningful than if Garry et al. (1996) had shown evidence of a test occasion × condition interaction. In fact, for some research questions, the Garry et al. type of analysis is the one that answers the research question best. Take, for instance, a pharmaceutical company that believes it has developed a new hand cream designed to prevent hands from drying and cracking in the winter cold. To gather data about the hand cream's safety and efficacy, the company begins trials with human subjects. Suppose one aspect the company needs to evaluate is the nature and extent of side effects, such as the incidence and severity of allergic reaction in the population. One kind of allergic reaction it might look for consists of swelling of the hands. Thus, the company will gather wrist measurements for all its subjects before applying the hand cream to half of them, and a placebo cream to the other half. Then they will wait one hour and measure everyone's wrists again. If the hand cream groups' wrists inflate more than the control groups' do, that is evidence of an allergic reaction.

What kind of data analysis would be more meaningful in this example: the time × condition interaction, or the proportion of subjects in each condition who showed inflated wrists? The time × condition interaction would be uninformative if a sizeable minority of subjects showed an allergic reaction, because the substantial number of subjects who showed no change in wrist size would overshadow the minority of those who might have shown sizeable reactions. By contrast, looking at the data to see the proportion of subjects in each condition who experienced

some wrist inflation is much more informative—and is analogous to what Garry et al. (1996) did.

In fact, the parallels with the hand cream example are numerous. In the hand cream trials, the most common change in wrist size should be no change. Likewise, in the imagination inflation paradigm, the most common change is no change. Such a finding makes sense in light of Heaps and Nash's (1999) work. They found a predisposition to hypnotic suggestion and dissociation—tendencies held by a minority of the population (Waller & Ross, 1997)—predicted imagination inflation, a finding supported by the work of Paddock et al. (1998). In both instances, the "inflaters'" mean scores are already heavily affected by the zero scores. These zero scores mask the change whether one analyzes the data using subjects as cases or events as cases (subjects × critical items in each condition), yet in both instances the minority "inflaters" are the more interesting group to study. Finally, there is another obvious and important parallel here. Clearly, some subjects in the hand cream trial who receive only a placebo will still show inflated wrists, because their hands swell as the day goes on, or because of some expectancy effect, and so forth. But the key question is this: When people experience an allergic reaction to the hand cream, by how much do their wrists inflate in comparison with those of controls? Similarly, if we ask the analogous question of the Garry et al. (1996) data, we find that 42/152 imagined events inflated, with a mean inflation of 3.00 (1.98), and 38/152 not imagined events inflated with a mean inflation of 2.50 (1.35).

The problem, as we see it, is to find the right combination of design, measures, and analytical technique to detect the influence of imagination. On that count, the pa-
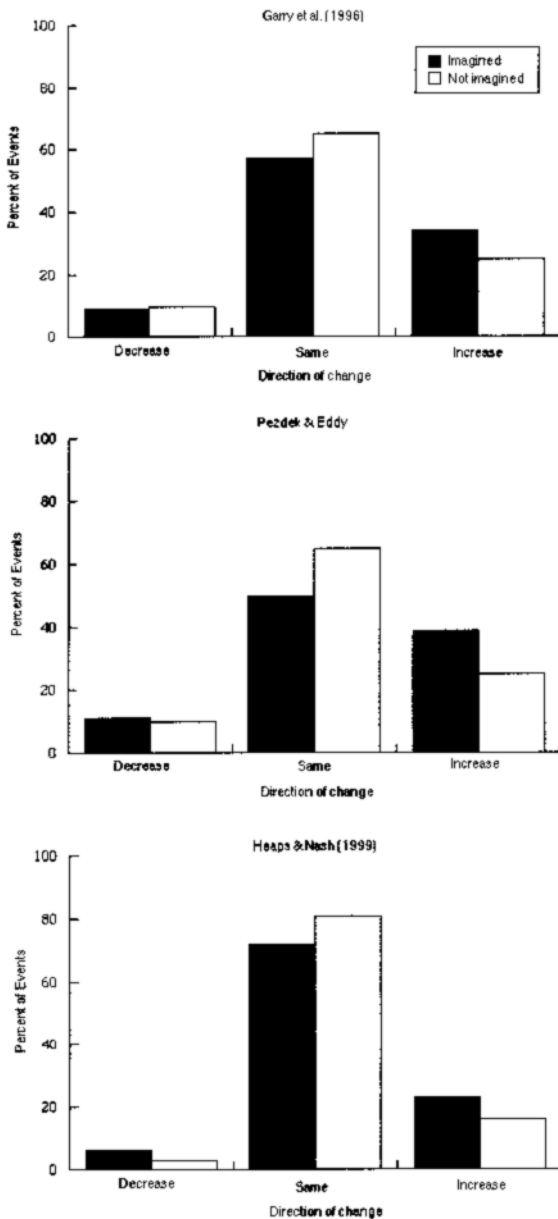
Figure 3. Direction of change for subjects who initially responded 1–4, across three studies.

pers by Garry et al. (1996), Heaps and Nash (1999), and Pezdek and Eddy all fall short, and future research on imagination inflation should look at ways to increase the size of the effect, modify the method, and come up with better ways of analyzing the data.

3A. The Prediction: All Mean Likelihood Ratings Should Increase From the Pretest to Posttest

If imagination inflation is a real effect, all mean confidence ratings for imagined events, including those rated 5–8, should inflate from pretest to posttest.

3B. The Turtle: There Is No Basis for the Prediction

To address this third prediction, Pezdek and Eddy analyze their data in several different ways. First, for each critical event, they split the data into two groups, based on pretest responses. Events that were initially rated 1–4 were put into one group, and events initially rated 5–8 were put into the other. For the sake of simplicity, let us refer to these two groups as the *low* and *high* pretest groups, respectively. Then, for each event in each group, Pezdek and Eddy calculated the percentage of subjects whose confidence increased, decreased, or stayed the same from the pretest to posttest Life Events Inventory (LEI). Within each group, they collapsed these change scores across all subjects and all events and classified them according to whether the event was imagined or not imagined. This analysis is what Garry et al. (1996) did, but they confined their analysis to low events, explaining that the purpose of their study was to examine what happens when people imagine an unlikely event, not recall a likely one.

Pezdek and Eddy's analysis of their data, and the comparable analysis of the Garry et al. (1996) data can be summarized into three broad findings.

Finding 1: For low pretest events, ratings increased for both imagined and not imagined events. Garry et al. (1996) found that for low events, confidence ratings inflated when they were imagined and when they were not imagined; 34% of imagined events showed increased confidence, and 25% of not imagined events showed increased confidence. Pezdek and Eddy found a similar pattern in their data, as did Heaps and Nash (1999); although Heaps and Nash did not present their findings this way in their paper, we have reanalyzed their data along similar lines, shown in Figure 3.

Pezdek and Eddy say that "regression toward the mean might explain the upward shift in scores that occurred in both conditions" (p. 709). Of course, they are absolutely correct on this point. RTM is almost certainly moving low scores upward in both conditions. However, they go on to say that "if the results simply reflect regression toward the mean, then (1) likelihood ratings for events initially rated 1–4 would be more likely to increase and likelihood ratings for events initially rated 5–8 would be more likely to decrease from Time 1 to Time 2" (p. 709). This prediction is an example of Pezdek and Eddy's tendency to confuse the event mean with the scale midpoint. In fact, RTM would move confidence to less than 4—because the critical events have overall means of less than 4—and an analysis purporting to detect RTM effects needs to take as its reference the mean, not 4.

Finally, the alert reader will note that all three experiments had both a control and a treatment group. Thus, the difference between the upward shifts in the two groups constitutes an effect. We are heartened to see that Pezdek and Eddy's findings contribute to the growing list of those who have replicated imagination inflation
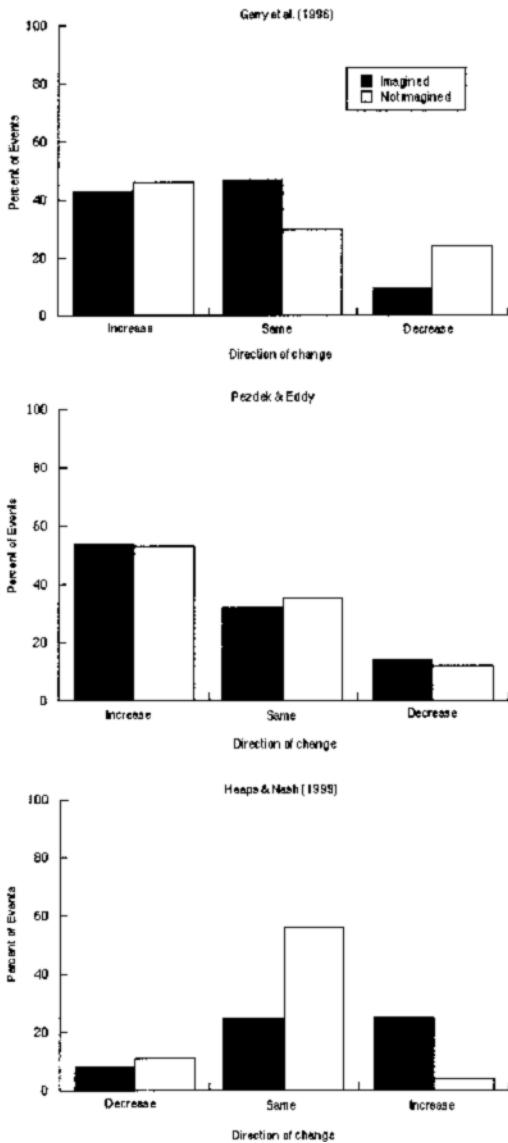
Figure 4. Direction of change for subjects who initially responded 5–8, across three studies.

(Heaps & Nash, 1999; Paddock et al.,1998; Paddock et al., 1999).

Finding 2: For high pretest events, ratings were more likely to decrease than do anything else. Consider now the top and middle panels of Figure 4. Both Garry et al. (1996) and Pezdek and Eddy found that when high pretest events were imagined, confidence tended to decrease, or stay the same. How are we to make sense of these results? Pezdek and Eddy make sense of them by saying "these findings are exactly what would be predicted by regression toward the mean" (p. 709). It is true that RTM predicts that high scores should move closer to the mean on retesting (as should low scores). However, if we now add Heaps and Nash's (1999) reanalyzed data

into the equation—on the bottom of Figure 4—the picture gets cloudier, because in their experiment, confidence rarely decreased.

Let us return to the behavior of high pretest events and ask another question. If we assume that imagination inflates confidence that the low, "probably didn't happen" events did happen, what should be the effect of imagination on confidence for the high, "probably did happen" end of the scale? We might make a number of predictions. On the one hand, Read and Lindsay (2000) have shown that when people think about genuine events, they remember more about them over time. On the basis of their research, we might expect that if high pretest confidence events indicate recall of genuine experiences, imagining these events will cause subjects to become even more confident about them than about high pretest events that are not imagined. On the other hand, it might well be the case that imagining counterfactual details about genuine experiences ultimately makes subjects less confident about what is real and what is imagined. The issue of how imagining "probably did happen" events affects memory for those events is a question worthy of study in its own right, and it is unlikely to be answered by post hoc analysis and speculation. As matters stand now, we have no idea what might be the effect of imagining events that subjects believe were probably genuine experiences. Pezdek and Eddy's assertion that if imagination inflation is a real effect, it should affect confidence about events no matter what their pretest ratings, is untenable given how little we know about the effects of imagination on memory.

A concrete example should make this point clearer. An antidepressant might work systematically to elevate the mood of depressed people, but might have a range of effects on the already well adjusted. If the drug works by somehow sucking up a bit of a biochemical that depressed people have in excess, most well-adjusted people will not have that excess biochemical, and the drug will not affect them. If, however, the drug works by bringing everyone who takes it up to some level of general good mood, perhaps some well-adjusted people might become even happier, whereas the chronically happy might already be at ceiling. The point is, we have no idea which, if any, of these mechanisms drives imagination inflation.

Of course, looking only at the decreasing scores in Figure 4 gives an incomplete picture. We should return to Figure 3 and see what happened to low pretest events. When they changed, were they more likely to increase than do anything else? The answer of course is "no." Garry et al. (1996), Pezdek and Eddy, and Heaps and Nash's (1999) data all show that the most common change for low pretest scores is no change. Here Pezdek and Eddy want it both ways: They assert that imagination inflation is really RTM dressed up as an effect, and as evidence, they show that the most common change for high pretest events is a decrease. Yet they do not expect the most common change for low pretest events to be an increase. Moreover, although Figure 3 shows a consistent pattern

across low pretest items for Garry et al., Heaps and Nash, and Pezdek and Eddy, Figure 4 shows no consistent pattern across high pretest items. We might conclude, therefore, that imagination operates more consistently on low-confidence events to produce imagination inflation than it does on high-confidence events. This is not to say that imagination has no effect on some high-confidence events; it is merely to say that the data across the three studies are inconclusive.

Finding 3: Items tend to move toward the mean at posttest. Pezdek and Eddy examined their data according to the following rationale: "if regression toward the mean is operative in the imagination inflation paradigm, it should be evidenced with the nontarget events, as well as with the target events" (p. 711). On this point, we have no dispute with Pezdek and Eddy. On the contrary, we believe that of course the data show evidence of RTM. Indeed, interpretation of any pretest–posttest design must take into account RTM. The best way to take RTM into account is to use a control group, or control items, that "can be used to estimate the effect of regression to the mean" so that "any difference observed between the two groups can be attributed to the treatment" (Kotz & Johnson, 1982, p. 707). This interpretation also accounts for the results Pezdek and Eddy found when they looked at change in confidence ratings for events rated 1–2, 3–4, 5–6, and 7–8 at pretest. They found that extreme scores showed more change than did scores already closer to the mean. Again, we are unsurprised by this pattern of data, and we agree with Pezdek and Eddy's claim that these results support "the significant role of regression toward the mean in accounting for the results of the imagination inflation paradigm" (p. 712), because regression to the mean plays a role in the overall results. However, their results do not account for the imagination inflation *effect*.

Finding 4: Only the low pretest group showed inflation. We were also heartened to see that Pezdek and Eddy replicated another of Garry et al.'s (1996) findings. That is, when low events were treated as cases, there was more inflation (the percentage of positive change) for imagined events than for not imagined events. It is interesting to note that even though Pezdek and Eddy used fewer critical items—and thus had fewer degrees of freedom in their *t* test—they still showed imagination inflation. Pezdek and Eddy then did the same analysis for high pretest events and found no evidence of imagination inflation. We agree with part of their conclusion, that "these results suggest a role of imagination inflation over and above the effect of regression toward the mean" (p. 712). However, they temper their conclusion by saying that it does not square with their magnitude of change analysis, part of which appears in their Table 2.

Pezdek and Eddy's Table 2 presents the magnitude of change in confidence ratings for low events. They conclude that because the increases are small, and because "large" changes were more likely to occur for not imagined events than for imagined events, the changes are

"not a product of the subjects' imagining the target events" (p. 713). We disagree with this conclusion. Yes, the changes are small, but they are consistent. A paired *t* test treating these four events as cases showed more inflation for imagined events ($M = 1.20$, $SD = 0.60$, $SEM = 0.30$) than for not imagined events ($M = 0.76$, $SD = 0.55$, $SEM = 0.27$). This difference was significant with a one-tailed test [$t(3) = 2.56, p = .04$]. In fact, in J. Cohen's (1988) terms, the effect is not small; it is an impressive $d = 1.26$.

4A. The Prediction: There Should Be an Age and Imagination Interaction

On the basis of G. Cohen and Faulkner's (1989) finding that older adults perform more poorly than younger adults at distinguishing between externally and internally generated information (watched vs. imagined events), Pezdek and Eddy claim that if imagination inflation is a real effect, there should be an age × imagination interaction when one compares the effects of imagination on college aged and on older adults. However, if RTM is behind imagination inflation, there will be no interaction, because RTM will affect both age groups equally.

4B. The Turtle: Pezdek and Eddy Failed to Consider Other Mechanisms

Although Pezdek and Eddy describe their prediction in a couple of sentences, we think it is worth more than a brief overview. There is nothing wrong with this prediction per se, but in focusing on the source-monitoring explanation (Johnson, Hashtroudi, & Lindsay, 1993), Pezdek and Eddy have failed to take into account other equally likely explanations, such as a familiarity-based explanation. Imagination inflation can also occur if the familiarity of an event is misattributed to the actual occurrence of the event rather than the imagining of the event.

Although the source-monitoring account suggests that the poor perceptual qualities of an imagined event are easily confusable with a long-ago faded memory of a real event, and that source judgments become less accurate with age (Schacter, Koutstaal, Johnson, Gross, & Angell, 1997), a familiarity account of imagination inflation makes no such prediction. Familiarity does not change with age. With the use of Jacoby's (1991) process dissociation procedure, it has been shown in many studies that the effects of familiarity (an automatic process) are not influenced by age, whereas the effects of the search component (a controlled process) are (Jennings & Jacoby, 1993; Schmitter-Edgecombe, 1999; Titov & Knight, 1997). Therefore, if familiarity is the main contributor to imagination inflation, we would expect to see no age × imagination interaction, because both younger and older adults are able to use familiarity judgments equally well.

Pezdek and Eddy's failure to consider other mechanisms for imagination inflation is not the only possible reason why they found no age × imagination interaction. Along with her colleagues, Pezdek herself has shown that

event plausibility is a crucial factor in changing what people believe about their past (Pezdek, Finger, & Hodge, 1997; Pezdek & Hodge, 1999). Furthermore, Anderson (1983) and Sherman, Cialdini, Schwartzman, and Reynolds's (1985) research on imagination and future likelihood judgements suggest that imagination is more likely to affect judgments about easy-to-imagine events than about hard-to-imagine ones; therefore, it might be reasonable to assume that implausible events are also difficult to imagine. Thus, we should expect imagination inflation to be influenced by the plausibility of an event.

Pezdek and Eddy failed to take plausibility into account when they included "Found a $10 bill in a parking lot" as one of only two critical events imagined by half of the subjects. Their college subjects (mean age, 20.9 years) probably believed that finding a $10 bill in the parking lot in 1988 (when these subjects were approximately 10 years old) was quite plausible. However, consider the situation for the older adults (mean age, 75.7 years). Finding a $10 bill in 1933 (when the older subjects were approximately 10) would be the equivalent of finding nearly $120 today according to the Bureau of Labor Statistics' Consumer Price Index (2000). Finding such a significant sum of money is not very plausible at all, particularly during the Great Depression. A second plausibility problem caused by using the "$10 bill" event is the frequency with which subjects might have been in a parking lot. When the college students were 10 years old, there were approximately 137 million cars in the United States (Bureau of the Census, 1988). However, when the older adults were 10, there were only about 1 million cars, and—presumably—not as great a need for somewhere to park them (Bureau of the Census, 1934). Thus, we may conclude that Pezdek and Eddy's older adults might have found the idea of being in a parking lot as a child to be a rather unusual event. The conjunction of these two situations—finding the equivalent of $120 and finding it in a parking lot—would be even less likely than either one alone.

In short, we believe that the 50% of older subjects who imagined finding a $10 bill in a parking lot may well have found the experience implausible, and/or difficult to imagine. These factors might have offset any tendency of these older subjects for increased source confusion (which by itself might have caused more imagination inflation), resulting in no effect for age on imagination inflation. Therefore, Pezdek and Eddy's claim that no age × imagination interaction shows that imagination inflation is nothing more than RTM is premature.

OTHER TURTLES

Do Pezdek and Eddy offer imagination inflation researchers a way out of the methodological quagmire? No. In fact, their discussion forecasts doom for any researcher who uses a pretest–posttest design. Note their treatment of Paddock et al.'s (1998) imagination inflation research. Paddock et al. extended the Garry et al. (1996) analysis by not examining the low pretest subjects separately from the high pretest subjects. In other words, they simply compared posttest means overall (having removed subjects whose pretest responses were 8) for imagined and not imagined events. Paddock et al. found that imagined events inflated more than not imagined events. However, Pezdek and Eddy suggest that because imagination inflation studies tend to rely on pretest distributions that are loaded heavily with low confidence ratings, "analyses that combine results for all events are not sensitive to the direction of effect for events with initial high likelihood ratings" (p. 717).

Although at face value that statement is true, Pezdek and Eddy seem to use it to prop up two claims that are not true. The first is that because the overwhelming majority of pretest scores are low, RTM will cause the overwhelming majority of scores to go up. The second is that because only a tiny portion of the pretest scores are high, RTM will cause only a tiny portion of the scores to go down. The net effect is an increase in mean confidence, or a spurious imagination inflation result.

Their warning reveals a misunderstanding of RTM for several reasons. First, if there is no effect for imagination, their scenario should apply to the control condition as well. After all, RTM should not be promoted by imagination; indeed, Pezdek and Eddy's fundamental thesis is that imagination has no effect. Second, when we consider only the effect of RTM on pretest and posttest scores, what should happen to the mean? Nothing. The overall mean stays the same. Third, when we consider only the effect of RTM on pretest and posttest scores, what should happen to the shape of the distribution? Nothing. Does it matter that the pretest confidence ratings are not normally distributed? No. If it were the case that repeated administrations of a test caused a mean increase in any distribution that was heavily weighted toward lower scores, we would find that the massive increase in the incidence of depression (Weissman, Livingston-Bruce, Leaf, Florio, & Holzer, 1991) could be attributed to repeated administrations of the Beck Depression Inventory.

Moreover, their warning assumes yet again that RTM operates to cause regression to the midpoint of the LEI scale, not to the mean of items. It is important for us to note here that for theoretical reasons, imagination inflation researchers sometimes analyze data by splitting it at the LEI scale midpoint. A midpoint split, we think, corresponds to confidence that the event probably did not happen (1–4) or probably did happen (5–8). However, Pezdek and Eddy continue to disregard the theoretical basis for the split, and throughout their paper they assume that splitting data on the midpoint is the equivalent to splitting data on the mean. Of course, these are not the same thing.

In short, an analysis that combines data points from subjects, without considering separately high and low scores, can be a reasonable approach if researchers use a control versus experimental design to control for RTM and other factors, such as time effects. Even if imagination operates mostly on low pretest scores, combining

data affords the researcher a wider range of statistical techniques, with the downside probably limited to a loss of statistical power, which can usually be compensated for by boosting the sample size. Besides Paddock et al. (1998), others have taken this approach (Garry, Frame, & Loftus, 1999; Hayes, 1999). As an alternative to simply increasing sample size, researchers might consider using Heaps and Nash's (1999) findings showing which factors predispose people to imagination inflation, to screen potential subjects' involvement in an imagination inflation experiment. Such an approach would boost the effect size, permitting combined data analysis with a smaller *n*.

FINAL COMMENTS

We have illustrated several shortcomings in Pezdek and Eddy's paper. These shortcomings range from the relatively minor instances of misrepresenting what imagination inflation is and what others have said about it, to major problems with their predictions. Perhaps the most important problem in their arguments is the lack of appreciation that comparisons between control and experimental conditions are used to adjust for a variety of possible factors, including RTM. Still, we believe that Pezdek and Eddy have made a contribution to the research on the effects of imagination and memory, in much the same way as McCloskey and Zaragoza (1985) did for misinformation research (Loftus, Miller, & Burns, 1978). McCloskey and Zaragoza attempted to demonstrate that there was no evidence that misleading postevent information had an effect on memory. Although they did not claim that previous evidence of postevent memory distortion was simply RTM disguised as an effect, they did claim something similar: that the standard forced-choice memory test Loftus and colleagues used biased subjects to respond in such a way as to give the appearance of an effect. Later, Belli (1989) showed that McCloskey and Zaragoza's modified test was actually insensitive to memory impairment. In many ways, McCloskey and Zaragoza's most important contribution was to spark renewed interest in the area, the end result being that this new research advanced our knowledge of how postevent suggestions affect memory (Abeles & Morton, 1999; Belli, 1989; Frost, 2000; Lindsay, 1990; Loftus & Hoffman, 1989; Zaragoza & Lane, 1998; Zaragoza & McCloskey, 1989). What we discovered grew quickly, and by leaps and bounds. We are grateful to Pezdek and Eddy for providing researchers with a similar opportunity to discover the ways in which imagination can influence our sense of ourselves and our past. Let us hope our research advances faster than a turtle does.

REFERENCES

ABELES, P., & MORTON, J. (1999). Avoiding misinformation: Reinstating target modality. *Quarterly Journal of Experimental Psychology*, **52A**, 581-592.

ANDERSON, C. A. (1983). Imagination and expectation: The effect of imagining behavioral scripts on personal influences. *Journal of Personality & Social Psychology*, **45**, 293-305.

BELLI, R. F. (1989). Influences of misleading postevent information: Misinformation interference and acceptance. *Journal of Experimental Psychology: General*, **118**, 72-85.

BUREAU OF THE CENSUS, US DEPARTMENT OF COMMERCE (1934). *Statistical Abstracts of the United States* (57th ed.). Washington, DC: Government Printing Office.

BUREAU OF THE CENSUS, US DEPARTMENT OF COMMERCE (1988). *Statistical Abstracts of the United States* (108th ed.). Washington, DC: Government Printing Office.

BUREAU OF LABOR STATISTICS (2000). *Consumer price index homepage* [On line]. Available: http://www.stats.bls.gov.cpihome.html.

CAMPBELL, D. T., & KENNY, D. A. (1999). *A primer on regression artifacts*. New York: Guilford.

COHEN, G., & FAULKNER, D. (1989). Age differences in source forgetting: Effects on reality monitoring and on eyewitness testimony. *Psychology & Aging*, **4**, 10-17.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

FROST, P. (2000). The quality of false memory over time: Is memory for misinformation "remembered" or "known"? *Psychonomic Bulletin & Review*, **7**, 531-536.

GARRY, M., FRAME, S., & LOFTUS, E. F. (1999). Lie down and let me tell you about your childhood. In S. D. Sala (Ed.), *Mind myths: Exploring popular assumptions about the mind and brain* (pp. 113-124). Chichester, U.K.: Wiley.

GARRY, M., & LOFTUS, E. F. (2000, March). *Imagination inflation is not a statistical artifact.* Paper presented at the meeting of the American Psychology-Law Society, New Orleans.

GARRY, M., MANNING, C. G., LOFTUS, E. F., & SHERMAN, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, **3**, 208-214.

GARRY, M., & POLASCHEK, D. L. L. (2000). Imagination and memory. *Current Directions in Psychological Science*, **9**, 6-10.

HAYES, J. E. (1999). *The effect of imagined time-distance on imagination inflation for both children and adults.* Unpublished master's thesis, Victoria University of Wellington, New Zealand.

HEAPS, C., & NASH, M. (1999). Individual differences in imagination inflation. *Psychonomic Bulletin & Review*, **6**, 313-318.

HEAPS, C., & NASH, M. (in press). Comparing recollective experience in true and false autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, & Cognition*.

HYMAN, I. E., JR., & KLEINKNECHT, E. E. (1999). False childhood memories. In L. M. Williams & V. L. Banyard (Eds.), *Trauma and memory* (pp. 178-188). Thousand Oaks, CA: Sage.

HYMAN, I. E., JR., & LOFTUS, E. F. (1998). Errors in autobiographical memory. *Clinical Psychology: General*, **117**, 371-376.

HYMAN, I. E., JR., & PENTLAND, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory & Language*, **35**, 101-117.

JACOBY, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.

JENNINGS, J. M., & JACOBY, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology & Aging*, **8**, 283-293.

JOHNSON, M. K., HASHTROUDI, S., & LINDSAY, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.

KOTZ, S., & JOHNSON, N. L. (1982). *Encyclopedia of statistical sciences*. New York: Wiley.

LINDSAY, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **16**, 1077-1083.

LOFTUS, E. F. (1993). The reality of repressed memories. *American Psychologist*, **48**, 518-537.

LOFTUS, E. F., & HOFFMAN, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General*, **118**, 100-104.

LOFTUS, E. F., MILLER, D. G., & BURNS, H. J. (1978). Semantic inte-

gration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 19-31.

LOFTUS, E. F., & PICKRELL, J. E. (1995). The formation of false memories. *Psychiatric Annals*, **25**, 720-725.

MAZZONI, G. (2001, June). *Time course in the development of false beliefs and memories after imagination*. Paper presented at the meeting of the Society for Applied Research in Memory and Cognition, Kingston, ON.

MCCLOSKEY, M., & ZARAGOZA, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, **114**, 1-16.

PADDOCK, J. R., JOSEPH, A. L., CHAN, F. M., TERRANOVA, S., MANNING, C., & LOFTUS, E. F. (1998). When guided visualization procedures may backfire: Imagination inflation and predicting individual differences in suggestibility. *Applied Cognitive Psychology*, **12**, S63-S75.

PADDOCK, J. R., NOEL, M., TERRANOVA, S., EBER, H. W., MANNING, C. G., & LOFTUS, E. F. (1999). Imagination inflation and the perils of guided visualization. *Journal of Psychology*, **133**, 581-595.

PEZDEK, K., & EDDY, R. M. (2001). Imagination inflation: A statistical artifact of regression toward the mean. *Memory & Cognition*, **29**, 707-718.

PEZDEK, K., FINGER, K., & HODGE, D. (1997). Planting false childhood memories: The role of event plausibility. *Psychological Science*, **8**, 437-441.

PEZDEK, K., & HODGE, D. (1999). Planting false childhood memories in children: The role of event plausibility. *Child Development*, **70**, 887-895.

READ, J. D., & LINDSAY, D. S. (2000). "Amnesia" for summer camps and high school graduation: Memory work increases reports of prior periods of remembering less. *Journal of Traumatic Stress*, **13**, 129-147.

ROGOSA, D., BRANDT, D., & ZIMOWSKI, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, **92**, 726-748.

ROGOSA, D. R., & WILLETT, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika*, **50**, 203-228.

SAGAN, C. (1979). *Broca's brain: Reflections on the romance of science*. New York: Random House.

SCHACTER, D. L., KOUTSTAAL, W., JOHNSON, M. K., GROSS, M. S., & ANGELL, K. A. (1997). False recollection induced by photographs: A comparison of older and younger adults. *Psychology & Aging*, **12**, 203-215.

SCHMITTER-EDGECOMBE, M. (1999). Effects of divided attention and time course on automatic and controlled components of memory in older adults. *Psychology & Aging*, **14**, 331-345.

SHERMAN, S. J., CIALDINI, R. B., SCHWARTZMAN, D. F., & REYNOLDS, K. D. (1985). Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imagery. *Personality & Social Psychology Bulletin*, **11**, 118-127.

TITOV, N., & KNIGHT, R. G. (1997). Adult age differences in controlled and automatic memory processing. *Psychology & Aging*, **12**, 565-573.

WALLER, N. G., & ROSS, C. A. (1997). The prevalence and biometric structure of pathological dissociation in the general population taxometric and behavior genetic findings. *Journal of Abnormal Psychology*, **106**, 499-510.

WEISSMAN, N. M., LIVINGSTON-BRUCE, M. L., LEAF, P. J., FLORIO, L. P., & HOLZER, C. E. (1991). Affective disorders. In L. N. Robbins & D. A. Reiger (Eds.), *Psychiatric disorders in America: The epidemiologic catchment area study* (pp. 53-80). New York: Free Press.

ZARAGOZA, M. S., & LANE, S. M. (1998). Processing resources and eyewitness suggestibility. *Legal & Criminological Psychology*, **3**, 305-320.

ZARAGOZA, M. S., & MCCLOSKEY, M. (1989). Misleading postevent information and the memory impairment hypothesis: Comment on Belli and reply to Tversky and Tuchin. *Journal of Experimental Psychology: General*, **118**, 92-99.

NOTES

1. Although the scenario that we describe is obviously possible yet extreme, we would not expect a parallel line when pretest scores were above 4, because we do not know how imagination affects confidence ratings for events judged as probably having happened. However, the issue of high pretest ratings is not a problem in our scenario, because Pezdek and Eddy included only scores below the mean. For those scores, we assume that the line might be parallel.

2. Garry et al. (1996) reported effect size using a different formula, the mean difference/$SD_{\text{not imagined}}$.