Essays on Consumer Learning and New Product Adoption

By

Darron Merrill Billeter

Committee

Co-Chair: Ajay Kalra

Co-Chair: George Loewenstein

Linda Argote

Joachim Vosgerau

Tepper School of Business

Carnegie Mellon University

April 2008

Table of Contents

Table of Contents (Continued)

1 Introduction

Many activities require the acquisition of new skills. If individuals become discouraged early during skill development, they are likely to quit, potentially depriving them of what could have otherwise been a positive experience. This research examines individuals' predictions of their own learning curves during the early stages of skill acquisition, and finds a systematic bias: Early overconfidence at the point of conceptual understanding of a novel task gives way to underconfidence immediately following initial experience with the task, which is evident in both predictions of long term learning and in short term performance forecasts. This sudden reduction in confidence is referred to as the 'all-thumbs effect'. I argue that it results from the failure to recognize how quickly automaticity develops in the acquisition of a new skill.

In the first essay, I document the all-thumbs effect, examine its generality across several tasks, and demonstrate its consequences for product devaluation and abandonment during the pivotal early skill acquisition stage. For products that require skills to use, such as computers, cell phones and sports equipment, consumers' purchase and usage decisions often depend on their prediction of the speed with which they will acquire the relevant skills. I find that after initially trying to use a new skill based product, consumers become discouraged in their own rate of learn and devalue or discard the product. I find that an important consequence of the all-thumbs effect is that people give up on products and activities that would have become beneficial if only they persisted.

Secondly, I investigate whether information about the all-thumbs effect can help learners mitigate the effect. In one experiment, participants are informed about the all-thumbs effect and then allowed to revise their learning curve predictions. Similar to previous overconfidence debiasing attempts (Lichtenstein et. al, 1982), I find that people ignore the information provided to them and persist in all-thumbs behavior.

In the second essay, I focus on investigating the mechanism underlying the all-thumbs effect. I argue that the all-thumbs effect occurs due to a failure to appreciate how rapidly a task is automated. Consistent with the proposed theory, I find that the all thumbs effect is exacerbated when the task is less automated. Additionally, I find greater underprediction for the initial phases of learning.

Another important research area that I investigate is the generalizability of the all-thumbs effect. I find that the all-thumbs effect is immune to prior experience and prediction elicitation modes, and find support for the bias in both motor and cognitive skill acquisition.

This dissertation documents an important bias in skill acquisition and new product adoption. I document important effects of the bias and investigate the underlying causes. In nine experiments and seven distinct learning tasks, I find that the all-thumbs effect is robust to many different classifications of skill learning and debiasing techniques.

2 All-Thumbs: Underpredicting Learning Curves Following Initial Experience with a Product

Since the seminal investigation of the diffusion of the hybrid corn seed innovation in Iowa (Ryan and Gross 1943), the issue of how people adopt new products has motivated research from a variety of perspectives. Factors governing the adoption of new products have been examined in marketing (Rogers 1976), economics (Katz and Shapiro 1986; Tirole 1988), strategy (Leonard-Barton 1992), information systems (Venkatesh and Davis 2000) and health care (Budman, Portnoy, and Villapiano 2003). The primary focus of this research has been on firm strategies and product characteristics that facilitate product adoption. Among the major findings is that two of the most important factors influencing the adoption and consumer's adoption intention (Davis, Bagozzi, and Warshaw 1989) of a new product are ease of use and perceived usefulness (Davis 1989).

Bagozzi, Davis, and Warshaw (1992) find that intentions to try a new technology are best forecast by consumer attitudes towards the process of learning, and their expected reactions to success and failure. Many consumer products that can increase consumer well-being are discarded by new users due to challenging initial product learning experiences. One category of goods for which initial product experience is particularly important is the domain of skill based products (Burson 2007; Murray and Haubl 2007), which require consumers to acquire skills to fully realize and appreciate the product's benefits. Examples include sporting goods (e.g. skis and sailboards), do it yourself products (e.g. home improvement, furniture) and electronic devices and appliances (e.g. computers, cameras, and bread-makers). For such products, besides perceived product benefits, adoption and ultimate usage is likely to be based in part on consumers' perception of their own abilities to master usage of the product.

There is some evidence that initial learning often serves as a barrier to new product adoption. A recent doctoral dissertation reports that consumers spend an average of only 20 min. trying to operate new electronics items before they give up, and that 50% of products returned to electronic stores that consumers claim to be defective are actually fully functional (den Ouden et al. 2006). There is also considerable anecdotal evidence that consumers either discard new products or do not fully utilize them. A survey conducted in the UK examined consumer usage of newly purchased kitchen items. The findings revealed that between 60-72% of consumers purchasing yogurt makers, plastic bag sealing devices, juicers and coffee machines ultimately fail to make use of their acquisitions (www.esure.com/news). Another survey of five hundred people found that 22% of respondents did not learn how to use a high technology gift they had received in the past year (http://techdigest.tv).

In this paper, we investigate how consumers form perceptions of their own ability to acquire skills. We contrast their perceptions with their actual ability to learn and find that consumers make a systematic error that we term the 'all-thumbs' effect. The all-thumbs effect refers to consumers' underconfidence in their own speed of mastery during the early stages of experience with a product. More generally, in our studies we observe the general patterns that, prior to any hands-on experience with a task, consumers are overconfident about both their initial mastery of a task and their speed of learning. This overconfidence quickly transmutes to underconfidence about their own speed of learning when consumers begin the skill acquisition process–the all-thumbs effect. We also find that consumers continue to erroneously underpredict their abilities during the pivotal early stages of learning, though calibration eventually improves with experience.

We demonstrate the generality of the all-thumbs effect with studies that span tasks involving visuo-spatial and fine motor skills, and also that the effect is robust across different measures of performance and learning. We also investigate a behavioral consequence of the effect, finding that the all-thumbs effect dynamically impacts product valuations, with high initial product valuations declining after product trial. We conclude with a discussion of the managerial implications of the all-thumbs effect, suggesting that consumer learning curves should be considered in planning promotional and advertising campaigns for skill based products.

2.1.1 Self-Assessments of Future Performance

The extensive literature on self-predictions of performance reveals that, although accurate assessments of performance allow consumers to make better purchasing decisions (Alba and Hutchinson 2000), consumers do not generally predict their own future performance accurately (Morwitz 1997). Mabe and West (1982) conduct a meta-analysis on the relationship between self perceptions of knowledge and actual performance and found that the correlation ranged from a high of .47 for athletics (motor skills) to a low of .17 for interpersonal skills.

A well established finding is that people have a strong optimistic bias for many activities (Dunning, Heath, and Suls 2004; Epley and Dunning 2006; McGraw, Mellers, and Ritov 2004), including using product features (Zhao, Meyer, and Han 2007). People tend to overate themselves both when it comes to predicting their own absolute performance (e.g., the planning fallacy, Buehler, Griffin, and Ross 1994) and predicting relative performance (e.g. the better-than-average effect, Kruger and Mueller 2002). Exceptions to the general finding of overconfidence have been found on difficult tasks, where worse than average ratings are found (Kruger 1999).

The measurement methods for investigating over(under)confidence have been classified as overestimation and overplacement (Moore and Small forthcoming). Over(under)estimation occurs when people predict that they will do better(worse) on the task than their actual performance. Over(under)placement occurs when a person predicts that they will do better(worse) than others on a task. Moore and Healy (2007) find that overestimation and overplacement are negatively correlated. Their conclusion is that on easy tasks people underpredict how well they will do and overestimate how well they will do relative to others. Because information about the self is so critical in determining both overestimation and overplacement, we focus on self-predictions of overestimation. This is a critical mental exercise in developing skills and determining whether a product will be beneficial.

2.1.2 Skill Acquisition

A robust finding in the literature on skill acquisition is the power law of learning, according to which learning is initially rapid, decelerates with experience, and loosely conforms to a power function (Newell and Rosenbloom 1981). The power law is one of the most fundamental principles of learning and has been illustrated for a wide range of tasks. Most of the classifications in Fleishman's classic (1975) taxonomy of auditory, visual, and perceptual motor skill tasks have been shown to follow the power law of learning (Newell and Rosenbloom 1981).

Ackerman (1987) reviews the stages of skill acquisition and finds similarity in the two major skill acquisition frameworks. Fitts (1964) provides one of the most widely used frameworks where he defines three stages of skill acquisition as cognitive (or information gathering stage), associative (or trial stage) and autonomous (or experience phase). In the first stage, the learner is focused on gathering the facts needed to understand and perform the task. In the second, or trial phase, the learner begins to try out the movements required for the task and may begin practicing. Normally, the learner also begins receiving feedback during the second stage of learning. The third, or experience phase, is when the learner begins to achieve a high level of skill performance. In this stage, the learner's actions are fast, smooth, accurate, effortless and largely removed from the learner's awareness—very similar to the concept of automaticity. In an alternative framework, Anderson (1982) describes skill acquisition in terms of the changes in knowledge structure from stage 1 (the declarative stage), to stage 2 (knowledge compilation), and finally stage 3 (procedural knowledge). But Ackerman and others view the main insights in the two frameworks to be complementary.

As a skill is acquired, the neural mechanisms governing learning also undergo important changes. For example, Haier et al. (1992) quantified the cerebral glucose metabolic rate (GMR) in Tetris video game players and found that after four to eight weeks of daily practice, GMR in cortical surface regions decreased despite a seven-fold increase in performance. Haier's findings support the notion that there is a reorganization of active brain areas with experience. Later experiments with more sophisticated brain imaging techniques have shown that during initial novel skill learning the control network (a network of discrete regions of the brain that control goal processing, attention and decision making) is initially very active but becomes less active with experience (Chein and Schneider 2005; Schneider and Chein 2003). The control network is primarily active during controlled processing, which is slow and effortful, but flexible, and is less active during automatic processing (Hill and Schneider 2006).

When the control network is active during initial skill learning, we posit that it is very difficult for a person to imagine performing the task in a state where the task has been automated and the control network is less active. This leads learners to predict that future execution of the task will require more effort than it actually will. Predictions of future behavior are powerfully influenced by what is experienced in the present, a finding known as "projection bias" (Loewenstein, O'Donoghue, and Rabin 2002) and has been shown to influence such diverse phenomenon as thirst (Van Boven and Loewenstein 2003) and product ownership (Loewenstein and Adler 1995). When comparing consumers who are engaged in controlled versus automatic processing, an effect closely analogous to projection bias would imply that, when engaged in controlled processing, it is difficult to fully appreciate the benefits in task performance that are likely to come from automation. Conversely, once a skill has been automated, consumers are likely to find it difficult to recall the level of effort that is required for skill acquisition. For example, Hinds (1999) finds that experts make greater errors in predicting novice performance than do other novices, positing that the error results from the expert's failure to recall the initial, difficult, period of mastery. As a consequence of the failure of consumers in a controlled state to appreciate how rapidly a task is automated, they are likely to underestimate their future performance.

Consistent with previous research, we predict that in the information gathering stage, prior to experience with a product, people will exhibit overconfidence. However, after gaining initial experience with a task, we predict that people will lower predictions to the extent of underpredicting their learning. We propose that this decrease in predictions and pessimistic outlook on future performance will lead to quitting behavior and reduced valuations for products that require skill acquisition. Finally, in the third, or experience phase, we anticipate self-assessments will eventually become accurately calibrated.

We conducted four experiments to demonstrate the all-thumbs effect and establish its generalizability across several types of skill acquisition tasks. Our focus is on self-assessments, not relative assessments of novel skill based product learning. For example, when a consumer contemplates purchasing a computer software package such as Quicken or assembling furniture, they are primarily interested in how quickly they will be able to learn to use or assemble the product, and not whether they will learn to use it quicker than other consumers. Additionally, Kruger finds that people focus on themselves and neglect information about others in determining relative predictions (1999), indicating the extreme importance of self-assessments even in the context of relative performance assessments.

We differ from earlier investigations into self-assessment in that we focus not on the forecast of ability at a single point but rather at multiple points during the learning process. With skill based products, consumers must navigate through an initial learning period. During this phase, consumers continually assess the costs and benefits of product usage. Whereas the extant literature generally considers predictions for only one upcoming period, our elicitation procedures allow us to investigate both short and long term, learning curve and rate of learning assessments.

We calculate overconfidence and underconfidence by comparing actual behavior to predicted behavior, similar to the investigation of self-assessment biases conducted by Epley and Dunning (2006). We also provide economic incentives for optimal performance and accurate self-assessments.

2.2 EXPERIMENT 1

In the first experiment we investigate consumers' perceptions of their ability to acquire a skill, and test the hypothesis that, after attempting to use a product, consumers lower their performance predictions, switching abruptly from overconfidence to underconfidence. That is, we test for a brief but critical lapse, in the early stages of direct experience with a product, in what is otherwise a pervasive overconfidence on the part of consumers.

2.2.1 Method and Procedure

Forty-eight participants from a paid subject pool consisting of both students and non-students were recruited for a show-up fee of $4 and performance based payments. Subjects learned the classic mirror tracing task (Snoddy 1926). This task was selected because it requires acquisition of a new skill but is simple enough that the learning rate is rapid. The mirror tracing task requires subjects to draw a shape only using a mirror's reflection. Participants were seated at a table facing a square mirror that rested at an 85 degree angle. A box was placed between the participant and the mirror, with a cloth covering the side of the box facing the participant so that subjects could place their hand in the box while tracing. Subjects were asked to trace an unbroken line between the boundaries that were formed by two, five pointed stars, with one star placed in the interior of the other star (figure 1).

Insert figure 1 about here

Participants were first given two minutes to view a folder containing instructions on performing the mirror tracing task. The instructions included the stimuli and evaluation procedures. A correctly completed trace required that the participants not cross the inner or outer boundary of the star pattern while drawing. If the boundary was touched, it was considered an error and participants were instructed to discard the trace and to begin a new drawing immediately. Subjects were instructed to correctly trace as many stars as possible in four rounds of five minutes each.

Immediately after viewing the instructions, participants were asked to predict the number of correct traces they would be able to complete in the four rounds. These predictions will be called “before-experience” (P_{BE}) predictions. After completing the before-experience predictions, participants were given two minutes to try the task. At the end of this initial experience period, participants again made performance predictions for the four rounds. These predictions will be termed “after-experience” (P_{AE}) predictions.

After these steps, the rounds commenced. All participants in a session began concurrently and a buzzer signaled the completion of the five minute round. At the completion of each round, the researcher and the subject counted and recorded the number of correctly traced stars. This process was repeated for each of the four rounds. After each round, participants made predictions for the remaining rounds. These predictions are denoted by P_{jn} where *j* indicates the time at which the prediction is made and *n* the trial for which performance is predicted. After the four rounds ended, participants were asked to roll a die to receive payment (as described below). Finally, subjects were paid, debriefed and dismissed.

The payment incentives were designed to ensure that subjects exerted maximum effort and made accurate predictions. Subjects were instructed that they would be paid for each round randomly at random either on performance or prediction accuracy. At the end of the task, a die was rolled for each of the four rounds. If the number rolled was between one and five, payment was based on performance which was 25 cents per trace completed. If a six was rolled, payment was based on both prediction accuracy and performance. Specifically, the formula used was $.25* (number of traced completed – | number of traces predicted- number of traces completed |). The incentive scheme was designed to be easily explained, incentive compatible, and to ensure that subjects would be motivated to forecast their own performance as accurately as possible while exerting full effort on the task.

We present the results using the following notations: A_n represents actual performance where n = 1, 4 indicates the round, and before experience and after experience predictions are indicated by P_{BEn} and P_{AEEn}, where n indicates the round predicted. Predictions for later rounds are indicated by P_{jn} where *j* represents the time period when the prediction is made and *n* is the round predicted.

2.2.2 Results

First, we investigate whether learning in the mirror tracing task follows the power law. We find that the power function provides a better fit to the data than the exponential model (table 1). This confirms previous findings that skill acquisition can be modeled by the power function (Newell and Rosenbloom 1981).

Insert table 1 about here

Reduction in Predictions after Initial Experience. Next, we compare the predictions made before and after they experience the task. As can be seen in table 2, participants lower their predictions for all the rounds after acquiring initial experience. The reduction in outlook is significantly different for all rounds (all *ps*<.001).[1] Thus, moving from the information gathering stage to the trial stage of learning, we see an immediate and broad reduction in the participant's outlook.

Insert table 2 about here

Current Predictions. We now examine current predictions of performance in a round made immediately preceding the round predicted. Before trying the task, participants significantly overpredicted performance ($M(P_{BE1}\text{-}A_1) = 2.73$, SD = 7.94, $t(47) = 2.38$, $p < .05$, *Median* = 2.50, $Z(47) = 2.60$, $p < .01$). After initial experience, the overconfidence turns to underconfidence with participants significantly underpredicting performance ($M(P_{AE1}\text{-}A_1) = -1.21$, SD = 3.65, $t(47) = 2.29$, $p < .05$, *Median* = 0.00, $Z(47) = -2.10$, $p < .05$).

Subjects continue to underpredict their performance before round two ($M(P_{22}\text{-}A_2) = -2.21$, SD = 3.52, $t(47) = 4.35$, $p < .01$, *Median* = -2.00, $Z(47) = -3.74$, $p < .001$) as well as round

[1] Round one ($M(P_{BE1}\text{-}P_{AE1}) = 3.94$, SD = 6.12, $t(47) = 4.46$, $p < .001$ *Median* = 3.00, $Z(47) = 4.57$, $p < .001$), round 2 ($M(P_{BE2}\text{-}P_{AE2}) = 4.75$, SD = 7.07, $t(47) = 4.66$, $p < .001$, *Median* = 3.00, $Z(47) = 4.63$, $p < .001$), round 3 ($M(P_{BE3}\text{-}P_{AE3}) = 5.27$, SD = 7.54, $t(47) = 4.84$, $p < .001$, *Median* = 3.00, $Z(47) = 4.72$, $p < .001$), and for round four ($M(P_{BE4}\text{-}P_{AE4}) = 5.71$, SD = 7.66, $t(47) = 5.16$, $p < .001$, *Median* = 3.00, $Z(47) = 4.85$, $p < .001$).

three (M(P_{33}-A_3) = -1.15, SD = 2.85, $t(47) = -2.78$, $p < .01$, *Median* = -1.00, $Z(47) = -2.78$, $p <$ .01). Only at round four does calibration begin to improve to the point where the prediction error is not statistically significant (M(P_{44}-A_4) = .08, SD = 4.09, $t(47) = .14$, NS, *Median* = 0.00, $Z(47)$ = .09, NS).

Consistent with prior research findings, participants are overconfident before trying the new task. As hypothesized, participants underpredict their upcoming round performance after initial experience. The underprediction persists until the beginning of round four. Thus, we find a significant shift from overconfidence to underconfidence after participants first gain experience with the mirror tracing task (see figure 2). This underconfidence persists through the stages of the learning curve when performance improvement is particularly steep.

Insert figure 2 about here

*Predictions of Maximum Learning***.** The decision to persevere with a new task or product is contingent on predictions of the level of accomplishment likely to be achieved. We therefore examine predictions made for the final round as a proxy for peak performance. Before initial experience, participants directionally underpredict performance (M(P_{BE4}-A_4) = -3.60, SD = 12.75, $t(47) = -1.96$, $p < .06$, *Median* = -5.00, $Z(47) = -2.18$, $p < .05$). After initial experience, the magnitude of inaccuracy increases, producing significant underprediction (M(P_{AE4}-A_4) = -9.31, SD = *8.19*, $t(47) = -7.88$, $p < .001$, *Median* = -9.00, $Z(47) = -5.38$, $p < .001$). Participants continue to underpredict maximum performance before rounds two (M(P_{24}-A_4) = -7.21, SD = 5.98, $t(47) = -8.35$, $p < .001$, *Median* = -6.00, $Z(47) = -5.60$, $p < .001$) and three (M(P_{34}-A_4) = -

2.48, SD = 4.92, *t(47)* = -3.49, $p < .01$, *Median* = -3.00, *Z(47)* = -3.08, $p < .01$). By round four, predictions become more accurate ($M(P_{44}-A_4) = .08$, SD = 4.09, *t(47)* = .14, NS, *Median* = 0.00, *Z(47)* = .09, NS).

The results indicate a systematic underconfidence in predicting maximum learning. Predictions prior to experience are already pessimistic, but become even more so following initial experience. The error is greatest following initial task experience.

2.2.3 Discussion

Consistent with earlier literature, we find that learning the mirror tracing task is initially steep and then decelerates with experience. We find a bias in predictions that follows a specific sequential path: in the initial stage where hands-on experience is yet to be acquired, participants are overconfident. This pervasive overconfidence immediately dissipates and lowers to underconfidence following initial experience. Though participants correctly forecast improvement in performance, the outlook remains systematically pessimistic for both current performance and maximum learning.

2.3 EXPERIMENT 2

Experiment 2 has two main objectives. First, to address the robustness of the all-thumbs effect with respect to elicitation method, participants predict task completion times rather than their performance in a unit time period. Second, the study examines whether the all-thumbs effect can be diminished or eliminated with a fairly heavy-handed form of debiasing.

2.3.1 Method and Procedure

Eighty-two undergraduate students participated in the experiment for a combination of extra course credit and performance based payment. We selected a t-shirt folding technique that most people are unfamiliar with as the learning task. The procedure involves four steps for successful folding. Participants were taught the task using a 40 sec. instructional video that demonstrates a novel way to fold a t-shirt which is much quicker than the methods that most people normally employ. The task requires acquiring both insight as well as motor skills. The instructions from the video are summarized in figure 3.

Insert figure 3 about here

The procedures followed were similar to those in the mirror tracing experiment, with a few differences. The practice period was divided into two phases of 40 sec. First, participants watched the instructional video once without being allowed to touch the t-shirt. Next, they viewed the video again but were allowed to practice folding the t-shirt. The task required was to fold two shirts in each of five rounds. Participants were instructed in using a stopwatch. Participants began and ended each round by timing themselves. The experimenter simultaneously also timed each participant, although the participant's recorded times were always used.

Each session consisted of between two and six subjects seated facing a computer terminal. To their immediate left was a flat empty workspace where participants folded the t-shirts. Headsets were provided so that participants would have the flexibility to simultaneously fold the t-shirts and listen to the instructional video without disruption. Each work station, comprising both the computer terminal and workspace, was partitioned so that it could not be viewed by other participants.

The compensation scheme was adjusted to motivate subjects to perform maximally, but also to be incentive compatible for time predictions. As in the mirror tracing task, participants rolled a die at the end of the experiment for each round and were paid based on their performance or prediction accuracy. If the die rolled was between one and five, they were paid 1000 / (number of seconds it took to fold two shirts) in cents. If the die rolled was a six, they were paid based on their prediction accuracy. Participants were paid 1000 / (number of seconds it took to fold two shirts – | predicted number of seconds it took to fold two shirts - number of seconds it took to fold two shirts|).

The experiment is a between subjects, single factor design with two conditions: control and debias. In the debias condition, we informed participants of the prediction errors hypothesized due to the all-thumbs effect. After watching the instructional video, participants in the debias condition were told, "When we conducted this study in the past, we have consistently found two things. First, we found that before people practice folding the t-shirt, they predict that they will do much better than they actually do. Second, once they start folding the t- shirts, they predict that they will do worse than they actually do.*"* As a test that they had understood the information (i.e., as a kind of manipulation check), after reading this information, participants were asked to circle the correct answer to the following two questions: 1) "Before they start practicing, people predict that they will do better / worse than they actually do," and 2) "After they start folding the t-shirts, people predict they will do better / worse than they actually do." Two subjects failed to answer both questions correctly and were dropped from the analysis.

2.3.2 Results

We begin with the dependent measures used in experiment 1. We first discuss the results within each condition and later compare the differences. The results for the control and debias condition are reported in tables 3a and 3b respectively.

Insert table 3a & 3b about here

We investigate whether learning to fold a t-shirt follows the power law. As in experiment 1, we find that the power function provides a better fit, in each condition, to the data than the exponential model (table 3a and 3b). This finding confirms that learning in the t-shirt folding task can be characterized by the power law.

Insert table 4a & 4b about here

Reduction in Predictions after Initial Experience. Consistent with experiment 1, in both the control and debias condition, predictions are significantly lowered after initial experience for all five rounds (all *p*s<.05)[2].

---

[2] ($M(P_{BE1}$-$P_{AE1})_{control}$ = -8.23, SD = 24.10, $t(38)$ = -2.13, p < .05, $M(P_{BE1}$-$P_{AE1})_{debias}$ = -9.15, SD = 26.12, $t(40)$ = -2.24, p < .05), round two ($M(P_{BE2}$-$P_{AT2})_{control}$ = -8.33, $t(38)$ = -2.39, p < .05, $M(P_{BE2}$-$P_{AE2})_{debias}$ = -9.10, SD = 18.47,

Current Predictions. Consistent with experiment 1, participants are overconfident before trying the task ($M(P_{BE1}\text{-}A_1)_{control}$ = 91.19, SD = 199.70, $t(38)$ = 2.85, p < .01, $M(P_{BE1}\text{-}A_1)_{debias}$ = 94.22, SD = 166.12, $t(40)$ = 3.63, p < .001). And, as in the prior experiment, subjects are underconfident before round two($M(P_{22}\text{-}A_2)_{control}$ = -11.12, SD = 29.12, $t(38)$ = -2.39, p < .05, $M(P_{22}\text{-}A_2)_{debias}$ = -19.74, SD = 64.33, $t(40)$ = -1.96, p < .05) and round three($M(P_{33}\text{-}A_3)_{control}$ = -5.27, SD = 9.81, $t(38)$ = -3.35, p < .01, $M(P_{33}\text{-}A_3)_{debias}$ = -8.69, SD = 30.34, $t(40)$ = -1.83, p < .05). Participants become accurate in predicting task learning before round four ($M(P_{44}\text{-}A_4)_{control}$ = .50, SD = 6.51, $t(38)$ = .48, *NS*, $M(P_{44}\text{-}A_4)_{debias}$ = -3.88, SD = 15.40, $t(40)$ = -1.62, p < .10), and this continues in round five ($M(P_{55}\text{-}A_5)_{control}$ = -.94, SD=5.67, $t(38)$ = -1.04, NS, $M(P_{55}\text{-}A_5)_{debias}$ = .38, SD = 5.87, $t(40)$ = .42, NS). Unlike experiment 1, however, we find that after the initial trial, participants remain overconfident ($M(P_{AEAE1}\text{-}A_1)_{control}$ = 82.96, SD = 195.80, $t(38)$ = 2.65, p < .01, $M(P_{AEAE1}\text{-}A_1)_{debias}$ = 85.07, SD = 20.35, $t(40)$ = 3.34, p < .01). This was unexpected. In the previous experiment, participants understood the task and had obtained significant experience. In the t-shirt folding task, many of the participants appeared to be stuck in the information gathering stage, attempting to understand what would be required of them. Many of the participants only attempted the task once in the t-shirt folding task (practice period = 40 sec., M_{Round1} = 149 sec.), perhaps enough experience for them to realize that they were overly optimistic, but perhaps not enough to realize the steepness of the learning curve. Interestingly enough, the effect still occurred in both conditions-but later (after round 1) than we had anticipated.

$t(40)$ = -3.15, p < .01), round three ($M(P_{BE3}\text{-}P_{AT3})_{control}$ = -7.33, SD = 19.50, $t(38)$ = -2.35, p < .05, $M(P_{BE3}\text{-}P_{AE3})_{debias}$ = -12.37, SD = 38.28, $t(40)$ = -2.07, p < .05), round four ($M(P_{BE4}\text{-}P_{AE4})_{control}$ = -8.41, SD = 17.23, $t(38)$ = -3.05, p < .01, $M(P_{BE4}\text{-}P_{AE4})_{debias}$ = -11.07, SD = 31.15, $t(40)$ = -2.28, p < .05), and round five($M(P_{BE5}\text{-}P_{AE5})_{control}$ = -9.31, SD = 16.58, $t(38)$ = -3.51, p < .01, $M(P_{BE5}\text{-}P_{AE5})_{debias}$ = -9.63, SD = 28.50, $t(40)$ = -2.16, p < .05).

Predictions of Maximum Learning. The results for maximum learning replicate those from the mirror tracing task. Before experiencing the task, participants underpredict their peak performance but they become more pessimistic following initial experience. This significant underprediction in peak performance continues until round four. [3]

Comparison between Control and Debias Conditions. Comparing the two conditions, we find no significant differences in the reduction of predictions after initial experience or in the predictions of current or next period learning. Thus, for all rounds of all the dependant measures, the debiasing intervention did not significantly improve participants' prediction accuracy.

2.3.3 Discussion

The results indicate the all-thumbs effect also robust across different response modes of predicting performance. The debiasing approach used appears to have no discernable effect on the distinct pattern of the all-thumbs effect. It is, of course possible that other debiasing interventions might have more of an effect. For example, participants could be provided with average performances for each round. We did not do so as pretests indicated high variances in performances.

---

[3] Participants begin by underpredicting peak performance before experience ($M(P_{BE5}\text{-}A_5)_{control}$ = -4.71, SD = 19.19, $t(38)$ = -1.53, NS, $M(P_{BE5}\text{-}A_5)_{debias}$ = -7.27, SD = 20.35, $t(40)$ = -2.29, $p < .05$) which was followed by increased inaccuracy. They continue to underpredict after initial experience ($M(P_{AE5}\text{-}A_5)_{control}$ = -14.02, SD = 27.05, $t(38)$ = -3.24, $p < .01$, $M(P_{AE5}\text{-}A_5)_{debias}$ = -16.91, SD = 31.12, $t(40)$ = -3.48, $p < .001$), before round two ($M(P_{25}\text{-}A_5)_{control}$ = -11.71, SD = 18.35, $t(38)$ = -3.99, $p < .001$, $M(P_{25}\text{-}A_5)_{debias}$ = -24.96, SD = 45.77, $t(40)$ = -3.49, $p < .01$) and before round three ($M(P_{35}\text{- }A_5)_{control}$ = -1.35, SD = 5.48, $t(38)$ = -3.25, $p < .01$, $M(P_{35}\text{- }A_5)_{debias}$ = -10.13, SD = 33.07, $t(40)$ = -1.96, $p < .05$). Peak performance predictions begin to be accurate before round four ($M(P_{45}\text{- }A_5)_{control}$ = -1.35, SD = 5.48, $t(38)$ = -1.55, NS, $M(P_{45}\text{- }A_5)_{debias}$ = -3.23, $t(40)$ = -1.55, $p < .10$), and continue before round five ($M(P_{55}\text{-}A_5)_{control}$ = -.94, SD = 5.67, $t(38)$ = -1.04, NS, $M(P_{55}\text{-}A_5)_{debias}$ = .38, SD = 5.87, $t(40)$ = .42, NS).

2.4 EXPERIMENT 3

The object of experiment 3 was to test if the findings from the earlier experiments were the result of demand artifacts created by eliciting predictions before trial. Potentially, the underconfidence results could be due to a pendulum shift begun with the elicitation of the earlier overly optimistic, inaccurate initial predictions (although anchoring seems intuitively more likely). Experiment 3 was designed to eliminate this alternative explanation. Second, we test for the existence of the all-thumbs effect on a commercially available product.

2.4.1 Method and Procedure

Seventy-one students participated in the experiment for extra credit and performance based payment. The task selected was typing using the Dvorak format keyboard. The Dvorak keyboard layout claims faster typing and less finger movement than the standard QWERTY keyboard. While these claims are controversial (Liebowitz and Margolis 1991), we selected the Dvorak keyboard due to its lack of familiarity and its commercial availability. Note that mastery of the Dvorak keyboard really requires two forms of learning, or one form of learning (of the new keyboard) and one form of unlearning (of the QWERTY keyboard).

Participants were randomly assigned to one of two conditions (single versus multiple predictions) in a single factor, between subjects design. In the multiple predictions condition, as in earlier experiments, predictions were elicited before initial experience and after round one. In the single prediction condition, participants made performance predictions only after completing round one of the task.

The same procedures were followed as in previous experiments. First, participants were given two minutes to view the Dvorak keyboard layout and a copy of the words that were required to be typed. Then, participants predicted the number of words they would type (if they were assigned to the two predictions condition) for both rounds. Next, participants were given two minutes to practice with the keyboard. The Dvorak keyboard was attached to a computer monitor. The words to be typed were displayed on the screen. At the end of each round, participants were instructed that the screen would display the gross words per minute, the number of errors typed and the net words per minute. Finally, as in earlier experiments, the actual task commenced. In the one prediction condition, predictions were elicited only after round one. Participants were paid based on their performance ($.03 * # of words typed) or prediction accuracy ($.03*(absolute value|(# of words typed)-(# of words predicted)|).

2.4.2 Results

Insert table 5 about here

As can be seen in table 5, the results from the previous experiments replicate. In the condition where participants make multiple predictions they are initially overconfident ($M(P_{BE1}-A_1)$ = 14.97, SD = 15.30, $t(34)$ = 5.79, p < .001). After round one, in both conditions, they significantly underpredict their performance ($M(P_{22}-A_2)_{single}$ = - 2.08, SD = 6.60, $t(35)$ = - 1.90, p < .05, $M(P_{22}-A_2)_{multiple}$ = - 2.03, SD = 6.62, $t(34)$ = - 1.81, p < .05). The results replicate the over and underconfidence of the all-thumbs effect found in tasks where the skill acquired did not require relearning.

We next compare the predictions made after round one in both conditions, and conclude that there is no significant difference between the round two predictions (*F(1,69)* = 1.08, *NS*) or the round two prediction variance (*Levine Statistic(1,69)* = .81, *NS*). We also conclude that there is no significant difference between the round two prediction error (*p22-a2*, *F(1, 69)* = .001, *NS*) or the prediction error variance (*Levine Statistic(1,69)* = .24, *NS*). Thus, we conclude that eliciting the initial, optimistic before trial predictions do not impact the after round one predictions.

2.4.3 Discussion

This experiment eliminates the alternative explanation that demand artifacts lead to performance underpredictions. The tendency to underpredict perseveres even in the absence of optimistic predictions elicitation prior to experiencing the task. Thus, the all-thumbs error is not a vestige of measurement effects or an over-correction of early overly optimistic inaccurate predictions. The results also establish the presence of the all-thumbs effect in tasks that involves a consumer product.

2.5 EXPERIMENT 4

Prior research suggests that consumers' valuations of products are dynamic and increase as experience with the product grows (Loewenstein and Strahilevitz 1998). In categories that require consumers acquire skills to better exploit product benefits, product valuations should be related to self-predictions of future performances. In skill based products, the all-thumbs effect would imply that at the beginning of the learning process, increased experience leads to a decrease in valuation. The objective of this experiment is to investigate whether initial experience with skill based products leads to decline in product valuations.

2.5.1 Method and Procedure

Thirty three students participated in the experiment for extra credit. We again selected typing with the Dvorak keyboard as the learning task. The same procedures as in experiment 3 were followed except that there was only one round in the task. As the results have replicated the all-thumbs effect in this task, instead of eliciting performance predictions, participants were asked to provide their valuation of the Dvorak keyboard. Participants were asked to value the keyboard before initial experience and after round one. The valuations were obtained using a variation of the Becker, DeGroot and Marshak (1964) procedure. Participants were instructed to state the amount of money that the Dvorak keyboard was worth to them. They were told that after the experiment, 10% of the participants were to be selected for inclusion in an actual drawing for the keyboards. For each participant selected, a random number between $0 and $40 was drawn. If the number drawn was less than their valuation for the keyboard, then they were given the keyboard. If the number drawn was more than their valuation of the keyboard, then they received the number drawn in cash.

2.5.2 Results

Before initial experience, participants valued the Dvorak keyboard at \$8.30. After round one, the keyboards were valued at \$6.70. This reduction in valuation after gaining experience with the Dvorak keyboard was significant (M = \$1.73, SD = 4.11, $t(32) = 2.41$, $p < .05$).

2.5.3 Discussion

The results show that for products which require acquisition of skills, the pessimism resulting from the all-thumbs effect can translate into reduced valuation for the product. We find that valuations for a skill based product initially decrease, rather than increase with experience. This decrease in valuation following initial product experience has important managerial implications. Limited product trials of skill based products could be detrimental to product sales. Marketing managers must carefully design the initial customer experience for potential skill based product customers. This experiment indicates that targeting skill based product sales before customers try the task, while they are optimistic, or perhaps after they have passed through the difficult all-thumbs phase, would be most successful.

2.6 CONCLUSION

For adoption to occur, consumers must not only purchase the products but fully utilize all the features and benefits. To do so, in skill based products, consumers must surmount the learning stage of the skill acquisition process. We propose an explanation for why this learning phase can appear formidable and therefore lead to quitting. The all-thumbs effect refers to the sudden drop in confidence (typically from an initial phase of overconfidence) when consumers begin the skill acquisition process with hands-on experience. The underconfidence lingers during the learning phase. Although we eventually observed calibration with experience, in many cases we suspect that such calibration is likely to come too late to prevent initial vexation and attendant behavioral effects. For example, 25% of first time snowboarders do not take a lesson purely due to their optimistic expectation that snowboarding will be easy. After trying, 85% of all snowboarders quit and never become long term participants in the sport (NSAA Report, 2003).

We have demonstrated the generalizability of the effect across three tasks, in different types of skill acquisitions, for skills that were new and those that had to be relearned. Crucially, the all-thumbs error has the behavioral consequence of lowering product valuations.

Because the initial stage of new product learning is so turbulent with both overconfidence mixed with underconfidence and product valuations are still being formed, we expect marketing initiatives to be particularly fruitful during this stage of skill product adoption. Hoch and Deighton's (1989) conceptual model of experience learning suggests that when customers are learning new, unfamiliar products they are particularly susceptible to management interventions. The impact of promotions and incentives to help people persist through the learning curve is an obvious avenue for future research. For example, promotional schemes targeted at first-time users must be designed so that consumers have sufficient incentives to endure the initial learning phase. Promotions such as “first-lesson-free” are likely to be less effective than promotions that help first time users to achieve a level of expertise where they are no longer pessimistic about their initial learning.

Because of the consumer behavior implications of all-thumbs, marketing managers of skill based products should seek out ways to 'hold consumers hands' during the initial stages of product experience or, at the least, to communicate the message to consumers that "If, at first, you don't succeed….try, try again".

TABLE 1. COMPARISON OF POWER VS. EXPONENTIAL FIT

	Model	A	b	Std Error	r
Power Fit	$Y=ax^b$	6.00	.77	.33	.99
Exponential Fit	$Y=ae^{bx}$	5.21	.31	1.39	.97

TABLE 2. RESULTS OF EXPERIMENT 1 (MIRROR TRACING TASK)

	Round 1	Round 2	Round 3	Round 4
Traces Completed	5.81(5.95)*	10.27(8.41)	14.35(10.24)	17.25(10.74)
Before Experience Predictions	8.54(7.53)	10.58(9.10)	12.23(10.00)	13.65(11.04)
After Experience Predictions	4.60(4.41)	5.83(5.15)	6.96(6.09)	7.94(6.83)
Before Round 2 Predictions		8.06(7.37)	9.25(8.45)	10.04(8.81)
Before Round 3 Predictions			13.21(10.37)	14.77(11.72)
Before Round 4 Predictions				17.33(11.89)

*Standard deviations are in parenthesis

TABLE 3A. RESULTS OF EXPERIMENT 2 (T-SHIRT FOLDING) CONTROL CONDITION

	Round 1	Round 2	Round 3	Round 4	Round 5
Actual Time	144.60(195.68)	43.16(35.60)	30.52(26.87)	27.11(15.36)	22.70(12.16)
Before Experience Prediction	53.41(30.59)	44.67(25.78)	36.69(21.81)	31.62(18.99)	27.41(16.39)
After Experience Prediction	61.64(42.43)	53.00(36.66)	44.03(28.91)	40.03(26.52)	36.72(24.96)
Before Round 2 Prediction		54.28(36.69)	43.85(26.63)	37.79(21.88)	34.41(19.98)
Before Round 3 Prediction			35.79(28.29)	28.33(13.18)	25.25(11.87)
Before Round 4 Prediction				26.62(14.71)	24.05(13.90)
Before Round 5 Prediction					23.64(12.90)

*Standard deviation's are in parenthesis

TABLE 3B. RESULTS OF EXPERIMENT 2 (T-SHIRT FOLDING) DEBIAS CONDITION

	Round 1	Round 2	Round 3	Round 4	Round 5
Actual Time	152.90(172.40)	47.90(49.90)	32.40(27.46)	24.87(13.93)	22.09(11.90)
Before Experience Prediction	58.68(57.12)	50.61(54.87)	38.39(24.91)	33.66(22.59)	29.37(17.59)
After Experience Prediction	67.83(72.88)	59.71(61.84)	50.76(54.44)	44.73(4.13)	39.00(34.53)
Before Round 2 Prediction		67.63(76.49)	58.37(70.81)	49.12(52.97)	47.05(52.73)
Before Round 3 Prediction			41.10(48.58)	35.27(42.62)	32.22(40.17)
Before Round 4 Prediction				28.76(23.48)	25.32(18.86)
Before Round 5 Prediction					21.71(12.08)

*Standard deviations are in parenthesis

TABLE 4A. COMPARISON OF POWER VS. EXPONENTIAL FIT (CONTROL CONDITION)

	Model	A	B	Std. Error	r
Power Fit	$Y=ax^b$	142.50	-1.41	8.67	.99
Exponential Fit	$Y=ae^{bx}$	317.41	-.82	17.39	.96

TABLE 4B. COMPARISON OF POWER VS. EXPONENTIAL FIT (DEBIAS CONDITION)

	Model	A	B	Std. Error	r
Power Fit	$Y=ax^b$	151.30	-1.43	6.73	.99
Exponential Fit	$Y=ae^{bx}$	340.81	-.83	15.89	.97

TABLE 5. RESULTS OF EXPERIMENT 3 (KEYBOARD TASK)

	Single Prediction		Multiple Predictions	
	Round 1	**Round 2**	**Round 1**	**Round 2**
Actual Words Typed	14.56(5.71)	20.00(4.88)	17.14 (7.54)	21.60(6.65)
Before Experience Predictions			32.11 (16.33)	40.80(20.94)
After Round 1 Predictions		17.92 (6.03)		19.57 (7.31)

*Standard deviations are in parenthesis

FIGURE 1. MIRROR TRACING TASK STIMULI

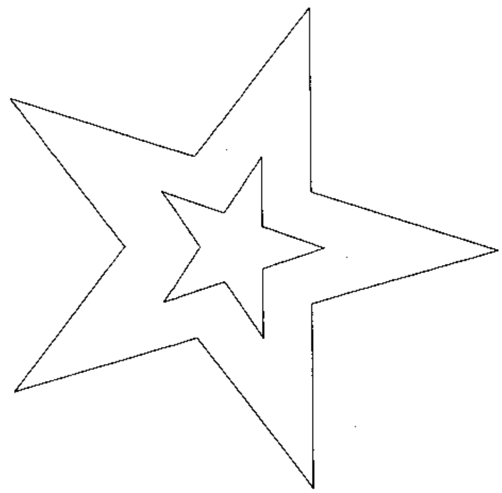


FIGURE 2. MIRROR TRACING EXPERIMENT

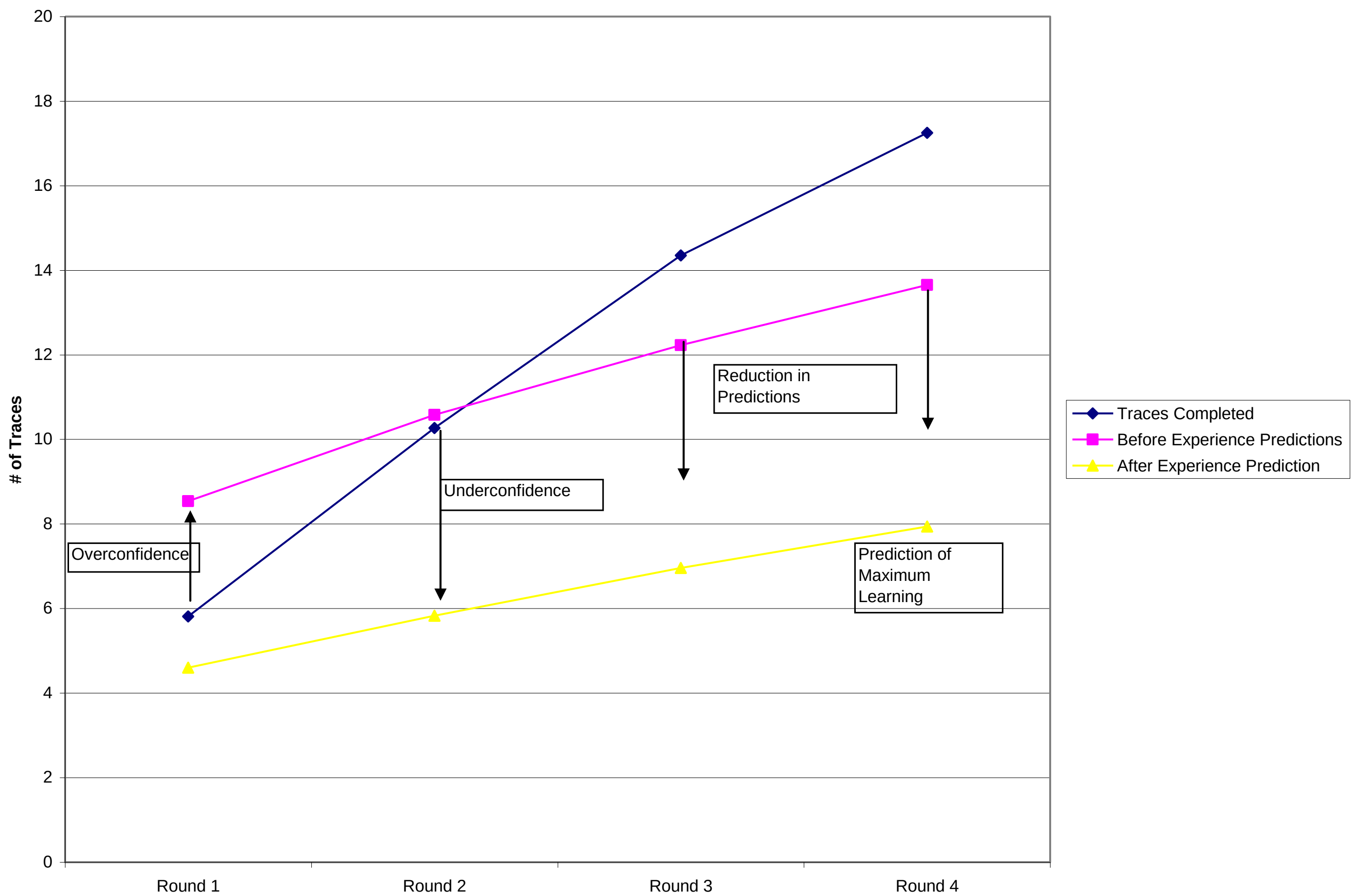
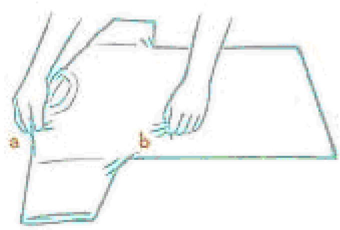


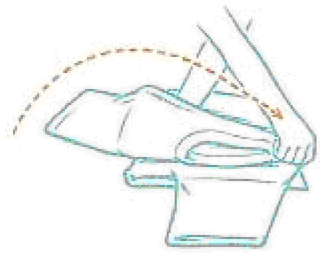
FIGURE 3. OVERVIEW OF THE T-SHIRT FOLDING TASK

T-Shirt Folding

Place the t-shirt on a flat surface in front of you, with the sleeves on the right hand side.

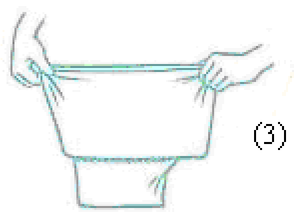


(1)

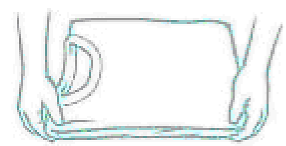


(2)

(1) Reach towards the sleeve that is farthest away from you and place your right arm on the shoulder, halfway between the beginning of the arm of the sleeve and the neck of the shirt
(2) place your left hand halfway down the shirt, parallel with your other hand and pinch the shirt with your fingers.



(3)



(4)

(3) There is an imaginary line connecting your left hand, right hand and the hem of the shirt. While maintaining your hold on the two points of the shirt, cross your right hand over you left hand so that the shoulder of the shirt meets the bottom hem (on the imaginary line) and grab the bottom of the hem. Uncross your arms without letting go of the shirt.
(4) Place the loose sleeve on the table and neatly arrange the shirt so it is presentable.

REFERENCES

Ackerman, Phillip L. (1987), "Individual Differences in Skill Learning: An Integration of Psychometric and Information Processing Perspectives," *Psychological Bulletin*, 102 (1), 3-27.

Alba, Joseph W. and J. Wesley Hutchinson (2000), "Knowledge Calibration: What Consumers Know and What They Think They Know," *Journal of Consumer Research*, 27 (September), 123-56.

Anderson, John R. (1982), "Acquisition of Cognitive Skill," *Psychological Review*, 89 (4), 369-406.

Bagozzi, Richard P., Fred D. Davis, and Paul R. Warshaw (1992), "Development and Test of a Theory of Technological Learning and Usage," *Human Relations*, 45 (July), 659-87.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marshak (1964), "Measuring Utility by a Single Response Sequential Method," *Behavioral Science*, 9 (3), 226–32.

Budman, Simon H., David Portnoy, and Albert J. Villapiano (2003), "How to Get Technological Innovation Used in Behavioral Health Care: Build It and They Still Might Not Come," *Psychotherapy: Theory, Research, Practice, Training*, 40 (Spring/Summer), 45-54.

Buehler, Roger, Dale Griffin and Michael Ross (1994), "Exploring the Planning Fallacy: Why People Underestimate Their Task Completion Times," *Journal of Personality and Social Psychology*, 67 (33), 366-81.

Burson, Katherine A. (2007), "Consumer-Product Skill Matching: The Effects of Difficulty on Relative Self-Assessment and Choice," *Journal of Consumer Research*, 34 (June), 104-10.

Chein, Jason M. and Walter Schneider (2005), “Neuroimaging Studies of Practice-Related Change: fMRI and Meta-Analytic Evidence of a Domain-General Control Network for Learning,” *Cognitive Brain Research*, 25 (3), 607-23.

Davis, Fred D. (1989), “Perceived Usefulness, Perceived Ease of Use, And User Acceptance of Information Technology,” *MIS Quarterly*, 13 (September), 319-40.

________, Richard P. Bagozzi, and Paul R. Warshaw (1989), “User Acceptance of Computer Technology: A Comparison of Two Theoretical Models,” *Management Science*, 35 (August), 982-1003.

Dunning, David, Chip Heath and Jerry M. Suls (2004), “Flawed Self-Assessment: Implications for Health, Education, and the Workplace,” *Psychological Sciences in the Public Interest*, 5 (3), 69-106.

den Ouden, Elke, Lu Yuan, Peter J.M. Sonnemans, and Aarnout C. Brombacher (2006), “Quality and Reliability Problems from a Consumer's Perspective: An Increasing Problem Overlooked by Businesses?” *Quality and Reliability Engineering International*, 22 (July), 821-38.

Epley, Nicholas and David Dunning (2006), “The Mixed Blessings of Self-Knowledge in Behavioral Prediction: Enhanced Discrimination but Exacerbated Bias,” *Personality and Social Psychology Bulletin*, 32 (5), 641-55.

E-sure Insurance and ICM (2006), "Annual Unused Gadget Poll 2006: New Entrants Fuel Growth of UK's 80m White Gadgets," http://www.esure.com/news_item.jsp?nId=18968.

Fleishman, Edwin A. (1975), “Toward a Taxonomy of Human Performance,” *American Psychologist*, 30 (12), 1127- 1149.

Fitts, Paul M. (1964), “Perceptual-motor Skill Learning,” in *Categories of Human Learning*, ed. Arthur W. Melton, London: Academic Press, 243-85.

Haier, Richard J., Benjamin V. Siegel Jr., Andrew MacLachlan, Eric Soderling, Stephen Lottenberg, and Monte S. Buchsbaum (1993), “Regional Glucose Metabolic Changes After Learning a Complex Visuospatial/motor task: A Positron Emission Tomographic Study,” *Brain Research* 570 (January), 134-43.

Hill, Nicole M. and Walter Schneider (2006), “Brain Changes in the Development of Expertise: Neurological Evidence on Skill-based Adaptations,” in *Cambridge Handbook of Expertise and Expert Performance*, ed. K. Anders Ericsson, Neal Charness, Paul J. Feltovich, and Robert R. Hoffman, New York: Cambridge University, 653-82.

Hinds, Pamela (1999), “The Curse of Expertise: The Effects of Expertise and Debiasing Methods on Predictions of Novice Performance,” *Journal of Experimental Psychology: Applied*, 5(2), 205-21.

Hoch, Stephen J. and John Deighton (1989), “Managing What Consumers Learn from Experience,” *Journal of Marketing*, 53 (April), 1-20.

Katz, Michael L. and Carl Shapiro (1986), "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy*, 94 (August), 822-41.

Kruger, Justin (1999), "Lake Wobegon be gone! The 'Below-average-effect' and the Egocentric Nature of Comparative Ability Judgments," *Journal of Personality and Social Psychology*, 77 (2), 321-32.

________and Reinhold A. Mueller (2002), “Unskilled, Unaware, or Both? The Better-than-average Heuristic and Statistical Regression Predict Errors in Estimates of Own Performance,” *Journal of Personality and Social Psychology*, 82 (2), 189-92.

Leonard-Barton, Dorothy (1992), *Core Capabilities and Core Rigidities: A Paradox in Managing New Product Development*, New York: J. Wiley.

Liebowitz, Stan J. and Stephen E. Margolis (2002), “The Fable of the Keys,” in *The Economics of Qwerty: History, Theory, Policy: Essays by Stan J. Liebowitz and Stephen E. Margolis*, ed. Peter Lewin, New York: University Press, 27-51.

Loewenstein, George F. and Daniel Adler (1995), “A Bias in the Prediction of Tastes,” *Economic Journal*, 105 (431), 929-37.

________, Ted O’Donoghue, and Matthew Rabin (2003), “Projection Bias in Predicting Future Utility,” *Quarterly Journal of Economics*, 118, 1209-1248.

________and Michal Strahilevitz (1998), “The Effect of Ownership History on the Valuation of Objects,” *Journal of Consumer Research*, 25 (December), 276-89.

Mabe, Paul A. and Stephen G. West (1982), “Validity of Self-Evaluation of Ability: A Review and Meta-Analysis,” *Journal of Applied Psychology*, 67 (3), 280-96.

McGraw, A. Peter, Barbara A. Mellers, and Ilana Ritov (2004), “The Affective Costs of Overconfidence,” *Journal of Behavioral Decision Making*, 17 (4), 281-95.

Moore, Don and Paul J. Healy (2007), "Bayesian Overconfidence," Working Paper, Carnegie Mellon University Center for Behavioral Decision Research, Pittsburgh, PA, 15213.

________and Deborah Small (forthcoming), "When it's Rational for the Majority to Believe That They are Better Than Average. In Joachim I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn M. Dawes*. Mahwah, NJ: Erlbaum.

Morwitz, Vicki G. (1997), “Why Consumers Don’t Always Accurately Predict Their Own Future Behavior,” *Marketing Letters*, 8 (January), 57-70.

Murray, Kyle B. and Gerald G. Haubl (2007), "Explaining Cognitive Lock-In: The Role of Skill-Based Habits of Use in Consumer Choice," *Journal of Consumer Research*, 34 (June), 77-88.

National Ski Areas Association and RRC Associates (2003), "Model for Growth: Beginner Conversion Research," Boulder, CO, 80301.

Newell, Allen and Paul S. Rosenbloom (1981), "Mechanisms of Skill Acquisition and the Law of Practice," in *Cognitive Skills and Their Acquisition*, ed. John R. Anderson, Hillsdale, NJ: Erlbaum, 1-55.

Rogers, Everett M. (1976), "New Product Adoption and Diffusion," *Journal of Consumer Research*, 2 (March), 290-301.

Ryan, Bryce and Neal C. Gross (1943), "The Diffusion of Hybrid Seed Corn In Two Iowa Communities," *Rural Sociology*, 8 (January), 15-24.

Schneider, Walter and Jason M. Chein (2003), "Controlled and Automatic Processing: Behavior, Theory, and Biological Mechanisms," *Cognitive Science*, 27 (3), 525-59.

Snoddy, George (1926), "Learning and Stability," *Journal of Applied Psychology*, *10* (February), 1-6.

Techdigest.tv (2007), "1.25 Billion Wasted on Gadgets Due to Fear and Ignorance," http://techdigest.tv/2007/05/125bn_wasted_on.html.

Tirole, Jean (1988), *Theory of Industrial Organization*, Cambridge: MIT Press.

Van Boven, Leaf and George Loewenstein (2003), "Social Projection of Transient Drive States," Personality and Social Psychology Bulletin, 29, 1158-1168.

Venkatesh, Viswanath and Fred D. Davis (2000), "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies," *Management Science*, 46 (February), 186-204.

Zhao, Shenghui, Robert J. Meyer, and Jin Han (2007), "A Tale of Two Judgments: Biases in Prior Valuation and Subsequent Utilization of Novel Technological Product Attributes," Working paper, The Wharton School of Business, University of Pennsylvania, Philadelphia, PA, 19104.

3. Predictions of Learning Curves

Over a century of research has investigated the functional form of, and psychological mechanisms underlying, skill acquisition (Thorndike & Woodworth, 1901). This research generally finds that skill acquisition can be approximated by a power function (Newell & Rosenbloom, 1981), with learning that is initially rapid, but which decelerates over time.[1] Learning curves of this type have been estimated for perceptual motor-tasks such as juggling (Swift, 1905), mirror tracing (e.g., Snoddy, 1926) and learning to use a typewriter (Swift, 1904), and for tasks that involve visual processing (Kolers, 1975), memory (Martin & Fernberger, 1929) and human-computer interaction (Card, Moran & Newell, 1983; Johnson, Bellman & Lohse, 2003).

Research into novel skill learning (reviewed in Hill & Schneider, 2006) has uncovered some of the processes that underlie most skill acquisition. This research finds that controlled processes, which are slow and effortful, but flexible, direct behavior in the initial stages of skill acquisition. Automaticity gradually develops with experience and the repeated pairing of a stimulus with a response (Shriffrin & Dumais, 1981; Shiffrin & Schneider, 1977). Automation of a task is normally accompanied by increases in speed, decreases in error, and decreases in the subjective feeling of mental effort (Logan, 1988). Additional processing models include a third phase, where learners utilize a mix of controlled to automatic processing (Shiffrin & Schneider, 1977), and is often referred to as the knowledge compilation (Anderson, 1981), or associative learning (Ackerman, 1988), phase of skill acquisition.

Although the research on learning curves has been exceptionally productive, it has largely failed to address a problem of particular importance for decision making. Decisions about whether to pursue, a particular task will not be based on objective learning curves, but on individuals' *predictions* of their own learning curves. For example, although learning curves for snowboarding and windsurfing are steep (most people can reach a level of acceptable proficiency rapidly), people often give up in early stages because initial learning is often frustrating and painful. Whether people end up deriving long-term pleasure from these activities, and more generally from activities that require mastery, therefore, is likely to depend on initial predictions of the rapidity of learning.[2]

There is extensive literature on self-predictions of performance (e.g., Buehler & Griffin, 2003; Koehler & Poon, 2006) across a wide variety of skill domains, including cognitive (Dunning et al., 1990), assembly (Byram, 1997), and perceptual-motor skills (Hinds, 1999). This literature has revealed a pervasive optimistic bias both when it comes to existing skills (Jagacinski, Isaac & Burke, 1977) and acquisition of new skills (Keren, 1987), a phenomenon that Buehler, Griffin and Ross (1994) document and dub the "planning fallacy" (see also Newby-Clark et al., 2000). This optimistic bias also impacts judgments of relative performance. For example, 70% of all high school students rate themselves as above average in leadership ability while only 2% rate themselves below average (Dunning et al, 2002) leading to a stream of research on "The better than average effect."

One exception to the general finding of overconfidence in performance predictions comes from memory research on judgments of learning. In tasks involving memorizing word pair associations, and then judging how many of the word pairs they will recall, Koriat and coauthors (2002) found that participants initially over-predict, then under-predict their own memory, which, the authors speculate, results from the failure to appreciate increased memory storage and retrieval fluency as word pair associations are memorized. Several other explanations have been proposed for the findings including the anchoring and adjustment mechanism (Scheck and

Nelson, 2005), and a mnemonic debiasing account (Koriat et al, 2006) while other explanations such as retrieval fluency (Serra and Dunlosky, 2005) or task difficulty (Koriat et al, 2002) have been eliminated.

In this article, we examine learning curve predictions in a task that requires acquisition of a novel skill. In contrast to prior literature, our focus is on predictions of learning curves and perceptions of improvement rather than performance at a single time point, and our specific interest is in the impact of early experience on such predictions. Prior to such experience, and consistent with earlier research, we expect that people will exhibit overconfidence (c.f., Hinds, 1999). However, after gaining initial experience with a task, we predict that people will exhibit underconfidence as a result of their failure to appreciate the speed and effectiveness of the shift from controlled to automatic processing.

Skill acquisition is initially directed by controlled processes, followed by a gradual shift to automatic processes with experience (Anderson, 1981). When the process is automatic, it is likely that people cannot recall the process when it is was controlled. There is indirect evidence that the initial skill acquisition process is improperly judged. Comparing expert and novices' predictions of the time required for novices to learn new cell phone functions, Hinds (1999) finds that novices are more accurate than experts in predicting the performance of first-time users.

We postulate that during initial experience with a task, when the neural processing is controlled, people find it difficult to imagine the process becoming automated. Because initial learning is so slow and effortful, novices likely envision that the rate of skill acquisition will be commensurately slow. We therefore expect that immediately after acquiring early experience on a difficult task, people will adjust their prediction of their own learning curve downwards. Depending on its magnitude, this downward adjustment could be sufficient to result in a shift from overconfidence to underconfidence. Only when sufficient automation has occurred would we expect predictions of subsequent learning to become more accurate. We term the predicted sudden drop in optimism about future learning following initial experience the 'All-Thumbs Effect'.

We present five experiments that document, and examine specific features of, the All-Thumbs effect. Experiment 1 tests for the occurrence of the All-Thumbs effect in the acquisition of a new motor skill. We examine the immediate adjustments made in predictions when the first phase of cognitive understanding of a new skill moves to the second phase of initial experience, and also examine ongoing predictions of future performance during the process of skill development. We delineate forecasts as either long-term predictions of learning rates and peak performances or as short-term performance projections. Experiment 2 investigates whether the All-Thumbs effect is due to the failure to appreciate how quickly automaticity is achieved. Experiment 3 examines whether the All-Thumbs effect is mitigated by experience – whether people who experience it on one task, and then realize that their initial projections of progress were too pessimistic, gain any insight into the effect that they carry over into subsequent tasks. Experiment 4 addresses the question of whether the All-Thumbs effect is robust to alternative prediction response modes. Experiment 5 examines whether the All-Thumbs effect potentially due to a demand artifact and also explores the generalizability of the effect.

3.1 Experiment 1

3.1.1 Procedure

Fifty-three undergraduate students at Carnegie Mellon University participated in the experiment for extra course credit and performance-based payments. The task they performed, tying a double bowline knot, was selected to be novel, conceptually simple, and easy to learn.

Participants were instructed that, after learning about the task, they would be required to tie as many knots as possible in 5 trials of five minutes each. Additionally, they were told that they would be asked to predict the number of knots that they could successfully tie in each trial and that they would be compensated based on both their speed at tying knots and the accuracy of their predictions.

Between 1 and 4 participants were run in a single session, seated so that they could not see each other. First, participants were given two minutes to view an animated instructional program on a computer that demonstrated the double bowline knot. The animation detailed the sequential steps required to tie the knot. The program ran continually and automatically restarted after each demonstration. After viewing the instructions for two minutes, participants predicted the number of knots they could tie in each of the five trials. These predictions will be referred to as “after-instructions” (P_{AI}) predictions.

Next, during the 'initial experience period', participants were given two minutes of experience tying the knot with a 10 inch untied piece of rope with the instructional program still showing on the video screen in front of them. At the end of the two minutes, participants again made performance predictions for each of the five trials, which we refer to as “after-experience” (P_{AE}) predictions.

After these steps, the trials commenced. Each participant was given a box of 10 inch ropes. All participants in a session began simultaneously, and an alarm indicated when the five minutes were complete. At the completion of each trial, both the participant and the researcher counted the number of correctly tied knots. After recording their knot production, participants again predicted their performances for the remaining trials. This process was repeated for each of the five trials.

The compensation scheme was structured to motivate maximum effort and accurate predictions. Payment was based either on the level of output (number of knots tied correctly) or on the level of output *and* prediction accuracy. At the end of the task, a 6-sided die was rolled for each of the 5 trials. If the number rolled was between 1 and 5, payment was based on performance and was equal to 5 cents per knot completed. If a 6 was rolled, participant payment was based on both performance and prediction accuracy. Specifically, the formula used was $.05* (number of knots tied – | number of knots predicted- number of knots tied|). This payment scheme ensures that participants will be motivated to perform as well as possible but also to guess their own performance as accurately as possible.

Note that our measure of over- or under-confidence is based on an objective measure of accuracy rather than relative (Burson & Klayman, 2007) or probabilitstic measures (Lichtenstein et al, 1982). Since assessments of self are more fundamental, we use absolute measures used in planning fallacy literature rather than relative measures. Probablistic measures were not considered due to methodological concerns (Gigerenzer et al., 1991; Juslin et al., 2000).

3.1.2 Results

We use the following notation in describing our results: A_n denotes actual performance where n = 1,5 indicates the trial, and the after-instructions and after-experience predictions are denoted by P_{AIn} and P_{AEn} where n indicates the trial for which the prediction is made. Later (post-experience) predictions are denoted by P_{jn} where j indicates the time at which the prediction is made and n the trial for which performance is predicted. Thus, P_{24} denotes prediction made before Trial 2 for performance in Trial 4.

The results for the after-instruction predictions, after-experience predictions and actual knots tied in Experiment 1 are summarized in Figure 1. We begin by conducting an ANOVA with the repeated factors trail (1-5) and prediction mode (after instruction predictions, after practice predictions and actual performance) as independent variables. Within this omnibus ANOVA, we use contrasts to investigate the All Thumbs effect and compare predictions to actual knot tying performance.

After receiving instructions, but before any experience, it can be seen that participants were generally overconfident about their initial performance but relatively well calibrated thereafter (because their initial overconfidence was compensated for by a tendency to underestimate their own speed of learning). (P_{AI1} - A_1= 1.75, $F(1,52)=2.99$, $p=.04$,). Thirty-six percent of the participants were underconfident (U), 19% were accurate and 45% were overconfident (O).

Our main prediction, however, is that participants' forecasts of their long-term rate of learning will be adjusted downwards after the new task is initially experienced. We examine the significance of this shift by comparing the predicted slope between periods 1 and 5 after instructions (P_{AI}) and after experience (P_{AE}). After watching the instructional video (after-instructions), the predicted rate of learning (P_{AI5} - P_{AI1}/4) is 2.52. After initial experience, the predicted rate of learning (P_{AE5} - P_{AE1}/4) is 1.79. The predicted learning slope is significantly lower after experience ($F(1,52)=11.49$, $p=.001$). After initial experience, 60% of the participants decrease their learning slope prediction after practice while only 19% increase their estimate.

After experience, participants were, on average, underconfident about their future improvement in performance. While the after-experience slope prediction ((P_{AE5}- P_{AE1})/4) is 1.79 knots per period, the actual rate of learning ((A_5-A_1)/4) of 2.79 knots per period is significantly higher ($F(1,52)=9.43$, $p=.002$, U = 75% , O = 19%).

Similar to the downward adjustments made in the predictions of slope, there is a downward shift in predictions of peak (i.e., end) performance. The peak performance predicted is significantly reduced following experience. (P_{AI5}- P_{AE5} = 5.57; $F(1,52)$=18.77, p<.0001). Seventy-four percent of participants lowered their prediction after experience, whereas only 17% raised their prediction.

We now identify the stage in the learning process at which the prediction errors are most egregious. To do so, at each stage, we calculate the learning slope predictions for the next two trials. We focus on the after experience predictions. In the steep, initial part of the learning curve, participants predict an improvement of 4.08 knots (SE=.64) between trials 1 and 3 but average 7.70 (SE=.59) significantly underpredicting their improvement (P_{AE3} - P_{AE1})-(A_3-A_1)=-, $F(1,52)$=16.77, p=.000). Between trial 2 and trial 4 the predicted and actual improvements are 3.14 knots and 5.04 knots respectively. This difference is also significant (P_{AE4} - P_{AE2})-(A_4-A_2)= -1.86, $F(1,52)$=6.19, p=.01). By trial 3, participants are better calibrated predicting an improvement of 3.47 knots (SE=.54) while improving by 3.07 knots (SE=.46). The difference is no longer significant ((P_{AE5} - P_{AE3})-(A_5-A_3)= -.40, $F(1,52)$=.40, p=ns). The findings indicate that underpredictions occurs most during the initial, steep portion of the learning curve. People eventually realize that practice makes perfect, but not that most of the perfection comes in the initial learning trials[iii].

Intermediate Predictions

The measures discussed earlier were predictions for the future made at the outset, either immediately after learning about the task, or immediately after first experiencing it. What happens when participants gain experience with the task, and hence feedback about the accuracy of their earlier predictions? After receiving feedback from trial 1 performance, underconfidence remains. Before trial 2, participants significantly underpredict performance (*F(1,52)*=9.09, p=.002, *p*<.01). Underconfidence continues to persist in predictions made before trial 3(*F(1,52)*=11.27, *p*=.0005). Participants are still largely underconfident before trial 4(*F(1,52)*=3.18, p=.04, *p*<.05) and calibrate their predictions prior to round 5 (*F(1,52)*=.007, p=.ns).

3.1.3. Discussion

The results reveal a systematic pattern of bias in the prediction of learning curves. In the first phase, after the task is described to participants, there is marginal overconfidence. In the next phase, however, when participants first begin attempting the new task, there is an immediate lowering of expectations. They are less confident, and indeed underconfident, about their long term rate of learning and about the ultimate level of performance that they are likely to achieve. Moreover, contrary to the normal finding of improved calibration with experience, there is a pervasive pessimism in predictions that is not fully overcome until the last trial. In the next experiment, we further examine the impact of experience and feedback on the All-Thumbs effect.

3.2 Experiment 2: All Thumbs and Automaticity

We have theorized that the all thumbs effect is due to a failure to appreciate how rapidly a task is automated. A key characteristic of automaticity is the improved ability to complete dual tasks simultaneously (Bargh, 1991, Moors & DeHeuver, 2006, Shiffrin and Dumais, 1981). If erroneous predictions of how quickly automaticity develops account for the all thumbs effect, inaccuracies in predictions should be more pronounced for tasks that are more controlled such as those which require multitasking as compared to single tasking assignments which are more automatic. In this experiment we examine whether misconstruing how quickly automaticity is achieved explains the all thumbs effect.

3.2.1 Procedure

Fifty-two students participated in the experiment for extra course credit and performance based pay. The experimental design was a single factor between participants design with two levels (single task and dual task). In the single task condition, participants learned Snoddy's (1926) classic mirror tracing task, which was chosen as it is a visual-spatial motor skill. The procedure was similar to experiment 1 except that there were four trials of five minutes and the payment was 25 cents for each correctly completed trace.

In the dual task condition, participants were asked to perform mirror tracing along with a counting task. Participants, at the beginning of the experiment, were given an individualized four digit number on a sheet of paper kept in front on them. After each trial began, a buzzer was sounded every twenty seconds and participants were asked to subtract 7 from the number assigned. This continued until the end of each trial. To ensure that effort was devoted to both tasks, the payment mechanism for the counting task was aligned with the mirror tracing task. Participants were instructed that they would loose 5% of their mirror tracing earnings for each miscalculated number in the round.

The experimental procedure began with a description of both tasks to all participants. Participants were informed that they would be randomly assigned to either of the two conditions. Next participants were asked to make performance predictions first for the single task which was followed by eliciting predictions for the dual task. Then, all participants practiced the dual task for two minutes. After practice, predictions were obtained for the single and dual task in that order. Once the after practice predictions were obtained, participants were assigned to either the single or dual task condition.

3.2.2. Results

Single Task

We begin by examining the single task condition. The results are provided in table 2a. First, we examine predictions made by participants after receiving instructions on the mirror tracing task. As in experiment 1, participants are initially overconfident ($P_{AI1(Single\ Task)} - A_{1(Single\ Task)}$=1.17, *SE*=.64, *F(1,22)*=7.57, *p*=.006*,* (U)=20%, (O)=68%).

Next, we investigate predictions made after experiencing the task. We find that participants significantly underpredict their slope improvement (*F(1,22)*=56.54, *p*=.000). As before, participants also significantly reduce their slope predictions after initial experience with the task (*F(1,22)*=18.85, *p*=.000). Additionally, participants reduce their peak performance predictions (*F(1,22)* =8.93, *p*=.0035) after experiencing the task.

Intermediate Predictions

We next investigate underconfidence after participants receive feedback on their performance. As in experiment 1, we find that participants continue to significantly underpredict their performance prior to trial 2 (*F(1,22)* =49.02, *p*<.000), trial 3 (*F(1,22)* =15.05, *p*=.0005) and trail 4 ($P_{44} – A_4$ = -1.50, SE=.73, *F(1,24)* =27.40, *p*=.000).

Dual Task

We now discuss the dual task condition results. As earlier, we begin by looking at predictions after instructions. Participants are directionally overconfident in predicting the performance of the dual task ($P_{AI1(Multitask)} - A_{1(Multitask)}$= .73, *SE*=.97, *F(1,25)*=.56, *p*=.ns (U)=34%, (O)=50%).

As before, after initial experience, participants significantly underpredict their improvement ($F(1,25)$=84.48, p=.000). Participants also significantly reduce their slope predictions after initial experience with the task ($F(1,25)$ = 17.38, p=.000). Predictions of peak performance also significantly reduce (($F(1,25)$ =4.27, p=.000) after the practice period.

Intermediate Predictions

The intermediate predictions also replicate. Participants continue to underpredict their performance after receiving feedback before trial 2 (F(1,25)=71.56, p=.000), trial 3 (F(1,25)=17.82, p=.000) and trail 4 (F(1,25)=4.27, p=.025).

Single vs. Dual Task

The failure to appreciate the development of automaticity can be gauged by first comparing the performances in the single and dual task conditions. At the starting point, trial 1, performances are marginally lower in the single task condition (F(1,46)=1.83, p=.086). At trial 4, the finishing point, there is no significant difference in performance (F(1,46)=.96, ns). Comparing the predictions for the single and the dual tasks shows that's that the incipient predictions however were much lower for the job requiring multitasking. After instructions, participants significantly downgraded their trial four predictions from 10.25 for the single task to 7.38 for the multi task ($F(1,46)$=32.53, p=.000). After initial experience, predictions for trial 4 performances were revised from 5.35 for the single task to 3.75 for the multitask ($F(1,46)$=43.94, p=.000).

While the lowering of predictions are not entirely surprising, the important result is that the slope predictions for multitasking were significantly lower than for single tasking after instructions (F(1,46)=16.42, p=.000) and also were lower after initial experience (F(1,46)=5.68, p=.01). The actual learning rates in the single and multitask conditions though are the same

($F(1,46)=.14$, p=ns) and performance rates converge by trial 4. Though actual learning rates for both tasks are similar, participants are much worse in predicting learning rates for dual task skill acquisition. The high discrepancy between predictions and actual learning rates for more controlled tasks suggests that the failure to recognize the speed at which automaticity develops may account for the all thumbs effect.

3.2.3. Discussion

Consistent with our proposed mechanism, we find that in tasks that are more controlled as they require multitasking, people are less sanguine. The experiment shows that the actual learning rates in the single and dual tasks were similar. The increased pessimism particularly of the slope predictions in the dual task suggests that people misconstrue the rapidity at which tasks are automated. This experiment also replicates the all thumbs effect in visual spatial skill acquisition.

3.3 Experiment 3: All-Thumbs and Previous Experience

Like learning curves, there is a long tradition of research on knowledge transfer – on the improvement in performance on one task as a result of learning a different task (Thorndike & Woodworth, 1901) or learning a task in a different context (Lewandowsky et al, 2002) at both the individual (Gray & Orasanu, 1987; Gregan-Paxton & Roedder John, 1997) and organizational levels (Argote et al., 1990). At the organizational level, knowledge transfer is seen as a source of competitive advantage (Argote & Ingram, 2000), and is facilitated by group members' perceived social identity (Kane et al., 2005). At the individual level, most studies find that motor skill transfer is generally small or even negligible, although exceptions have been found between tasks that are extremely similar (Schmidt, 1988).

Although research on knowledge transfer at the individual level is extensive, the focus has generally been on the impact of prior learning on subsequent actual performance rather than on predictions of performance. The first objective of Experiment 2, therefore, is to test if the sudden drop in confidence that we have referred to as the All-Thumbs effect recurs when an initial experience of all thumbs, followed by corrective feedback, is succeeded by another skill acquisition and prediction task. The second purpose of Experiment 2 was to examine the robustness of the All- Thumbs effect by attempting to replicate it with a task different from knot tying. Third, we investigate prediction carry over effects from one task to the next.

3.3.1. Procedure

One hundred and twelve undergraduate students participated in Experiment 3. The methodology was similar to Experiment 1 except that there were two stages in each of which participants acquired a new skill. The first task was manipulated between participants while in the second task, all participants learned the double bowline knot -- the same knot used in Experiment 1. Unlike Experiment 1, there were only three trials in each stage.

In the first stage, participants were randomly assigned to one of three initial task conditions in which they learned to tie either a relatively difficult knot (carrick bend) an easier knot (zeppelin) or the mirror tracing task used in Experiment 2.

3.3.2. Results of Task 1

The predictions and performance results are summarized in Figure 2 and presented in detail in Table 2. The figure shows three slopes: the actual slopes, the slope predicted after instructions and the slopes predicted after experience. The results for mirror tracing and easy knot conditions replicate experiment 1 for all dependent measures. For the mirror tracing task, participants were initially (after instruction but before experience) directionally overconfident, albeit not significantly so ($P_{AI1} – A_{1=}$ 2.06, *F(1,35*)=1.60, *p*=.11, U=42%, O=47%) for the first trial. More importantly, and consistent with the All Thumbs prediction, participants significantly reduced their estimates of learning curves immediately following initial experience (((P_{AI3} - P_{AI1})-(P_{AE3}- P_{AE1}))/2=1.35, *F(1,35)=12.24*, *p*=.0005, D=72%, I=11%), as well as their predictions of peak performance ((P_{AI3})-(P_{AE3})=7.56, *F(1,35)*=19.96, *p*=.000, D=78%, I=17%). This drop in expectations led participants to significantly underpredict their learning slope (((P_{AE3}- P_{AE1})-(A_3-A_1))/2= -3.47, *F*(1,35)=41.85, *p*=.000, U=83%, O=8%), an effect that persisted into the second trial (P_{22} – A_2=-3.64, *F(1,35)* =49.63, *p*=.000, U=75%, O=8%). A similar pattern of results was obtained with the easy (Carrick) knot.[3]

The results for the hard knot task are consistent with one component of the All-Thumbs effect in that participants begin overconfident about their own learning curve ($P_{AI1} – A_{1=}$ 3.89, (*F(1,37*)=2.86, *p=.04*, U=50%, O=42%) and, more importantly, there is a sudden downward adjustment in predictions made following initial experience (((P_{AI3} - P_{AI1})-(P_{AE3}- P_{AE1}))/2=1.98, *F(1,37)=4.43*, *p=.026*, D=53%, I=21%, and (P_{AI3})-(P_{AE3})=6.92,*F(1,37)=4.60*, *p*=.02, D=58%, I=29%). However initial overconfidence is so extreme that, despite the very large decrease in confidence following initial experience, participants did not end up significantly underpredicting their learning slopes (((P_{AE3}- P_{AE1})-(A_3-A_1))/2= -.19, *F(1,37*)=.13, *p*=.ns, U=58%, O=34%). In retrospect, the extreme initial confidence is not surprising, given that substantial research shows that people are more overconfident on harder tasks than easier ones (Lichenstein et al., 1982). Overall, these results suggest that an important aspect of All-Thumbs Effect, the reduction in predictions following experience, is quite robust.

3.3.3. Results of Task 2

The results of Task 2, which was identical to the task from Experiment 1, replicate the All-Thumbs effect – the sudden drop in confidence – whether the task was preceded by the easy knot condition, $(((P_{AI3} - P_{AI1})-(P_{AE3}- P_{AE1}))/2=.39$, *F(1,36)*=3.78, *p*=.03, D=51%, I=27%) the hard knot condition, $(((P_{AI3} - P_{AI1})-(P_{AE3}- P_{AE1}))/2 =.58$, *F(1,37)*=10.19, *p*=.003, , D=55%, I=11%) or the mirror tracing condition $(((P_{AI3} - P_{AI1})-(P_{AE3}- P_{AE1}))/2=.93$, *F(35)=15.03*, *p*=.000, D=58%, I=14%). The basic result of underpredicting the rate of learning also holds, regardless of which task came before: (easy knot condition, $((P_{AE3}- P_{AE1})- (A_3-A_1))/2=-.59$, *F(1,36)=4.19, p=.024*, U=54%, O=32%; hard knot condition, $((P_{AE3}- P_{AE1})- (A_3-A_1))/2=-2.09$, *F(1,37)=14.18, p=.001* U=74%, O=26%; mirror tracing task condition, $((P_{AE3}- P_{AE1})- (A_3-A_1))/2 =-.86$, *F(1,35)=3.06, p=.044*, , U=64%, O=36%).

As in the first skill acquisition task, participants adjust their prediction downwards after practice and remained underconfident in their intermediate predictions. The major impact of experience in the first task appears to be a reduction in participants' initial, post-instruction, overconfidence in task 2; in the second task of Experiment 2, instead of observing overconfidence following instructions (but before experience), there is directional underconfidence in the after-instructions predictions for trial 1 in all three experimental conditions – i.e., regardless of which task preceded task 2 (easy knot tying condition, $P_{AI1} - A_1=-.05$, *F(1,36)=.004*, *p=.95*, U=54%, O=35%; hard knot tying condition, $P_{AI1} - A_1=-1.74$, *F(1,37)=3.38, p=.03*, U=66%, O=29%; mirror tracing condition, $P_{AI1} - A_1=-1.67$, *F(1,35)=2.45, p=.06*, U=58%, O=39%).[4] After experiencing the All-Thumbs Effect, we find that pessimism carries over and impacts the initial predictions on the second task. On the second task we observe initial pessimism instead of optimism. Apparently, the feedback they received from the previous task had the effect of diminishing their initial overconfidence on the new task, but did not impact their reduction in predictions – i.e., the All-Thumbs effect.

3.3.4. Discussion

The results of Experiment 2 suggest that the All-Thumbs effect is robust. We observed an All-Thumbs effect in all three tasks in part one of the experiment, and an All-Thumbs effect in the second part, even when it was preceded by an extremely similar task. The results also suggest that the All-Thumbs effect is largely resistant to prior experience. Despite becoming aware that their post-experience predictions were insufficiently optimistic on the first task, this did not result in improved post-experience predictions on the second. Surprisingly, however, working on the initial task did largely eliminate the initial overconfidence that was observed in the first experiment after instruction but prior to experience.

The results show that the all thumbs effect is muted in the hard tasks. Consistent with the all thumbs effect, the initial overconfidence lowers but does not change to persistent underconfidence. This result is consistent with the hard/easy effect that finds overconfidence in difficult tasks (Lichtenstein & Fischoff, 1977). We also observe better calibration in the hard task which we suspect is due to ceiling effects. For very hard tasks, the maximum achievable performance rates are low. Since the peak is low, so is the potential to make errors.

3.4 Experiment 4: Directly Measuring Changes in Performances

So far, we have demonstrated a systematic bias in predicting the rate of skill acquisition. However, the predictions elicited from participants in the first two experiments were for their absolute performance rather than for changes in performance. The objective of this experiment is to examine if the bias is robust to alternative prediction elicitation approaches. Specifically, we now measure improvement rates directly.

3.4.1. Procedure

Thirty-three students participated in the experiment for extra course credit and performance based payment. Participants learned the mirror tracing task used in experiment 2. The procedures followed were very similar to the mirror tracing task in experiment 2 with the following exceptions. First, the experiment consisted of only three rounds. Second, as earlier, after reading the instructions and after initial experience, participants predicted their round 1 performance. Participants were then asked to predict the "additional number of traces" that would successfully complete from round 1 to round 3. After round 1, predictions of improvement from round 1 to round 3 were obtained again.

The compensation structure was adjusted for the changed prediction elicitation procedures. Participants were compensated similarly to the earlier experiments for round 1. However, they were compensated on their improvement or prediction of improvement subsequently. The same procedures as in earlier experiments were followed to determine whether payment was based on performance or prediction accuracy.

3.4.2. Results

The prediction and performance results are presented in Table 5 and Figure 4 are consistent with previous findings. Participants are overconfident after instructions ($P_{AI1} - A_1$=2.94, $F(1,32)$=4.34, p=.02, U=39%, O=52%). After instructions, participants predicted an improvement of 4.73 traces between round 1 and round 3 while after experience, they lowered their improvement prediction to 2.21. In fact, the improvement between round 1 and 3 was 6.82, indicating that participants were already underconfident about their own speed of learning following instruction and became even more underconfident following experience. The reduction in expectations regarding participants speed of learning is significantly different before and after experience $((P_{AI3} - P_{AI1})-(P_{AE3}- P_{AE1}))/2=1.27$, *F(1,32)*=9.57, *p*=.002, D=85%, I=9%). (see Figure 3). Also, consistent with previous results, participants significantly underpredict their actual improvement in learning $(((P_{AE3}- P_{AE1})-(A_3-A_1))/2= -.83$, *F(1,32)*=45.18, *p*=.000, *p*<.05, U=79%, O=12%).

Prior to trial 2, the underconfidence is maintained. Subjects predict an improvement of 3.36 additional traces (SE=.42) while actually completing 6.82 additional traces (SE=.64). The underprediction is significant (*F(1,32)*=21.79, *p*=.000).

3.4.3. Discussion

The results show that the All-Thumbs Effect is robust to an alternative prediction elicitation approach. Regardless of whether predictions are made for absolute performances or made for performance improvements, there is systematic underconfidence in predicting learning rates in the acquisition of a new skill.

3.5 Experiment 5 Mirror Reading

Experiment 5 serves two objectives. First, we test for the generalizability of the effect by using a cognitive learning task. Second, a limitation of the earlier experiments is that multiple predictions were elicited from participants. This procedure lends itself to the potential problem of demand artifacts. Through repeated prediction elicitations, participants could possibly become more sensitive to the dependent measures as the experiment progresses (Sawyer, 1975) or may believe that that the experimenters require them to change their predictions. We overcome this limitation in Experiment 5 by eliciting predictions on only one occasion.

3.5.1. Procedure

Seventy-nine participants received a $10 show up fee and additional performance based pay for participating. The learning task designed was mirror reading. In mirror reading, participants are required to decipher mirror imaged words. The procedure was similar to the previous experiments. Each session consisted of three to eight participants. The experiment consisted of three trials of three minutes each. After the procedure was explained, participants were given a sheet containing the words and experienced the task for 30 seconds. The task requires that the participants decode and write the words. Participants were paid $.10 for each correctly deciphered word and as before, were also paid based on their prediction accuracy. The words used were all eight-letters presented in normal Times Roman text mirrored using Pic2Pic software. Words with frequency ratings between 10 and 35 according to the Kucera & Francis's (1967) lexicon of English usage were selected. Examples of words used are "nowadays" and "charcoal" (see figure 1). To ensure that the task difficult did not vary across trials, the set of words presented across trials were a combination of identical frequency ratings.

The experimental design was a single factor between participants with two levels (multiple predictions vs single predictions). In the multiple predictions condition, predictions were elicited after instructions and after experience. In the single predictions condition, predictions were obtained only after experience. As in experiment 4, participants predicted the number of words they would successfully decipher in trial 1 and then predicted the number of additional words they would decode in trial 3.

3.5.2. Results

Multiple Predictions Condition

We first examine whether the All Thumbs effect replicates in the acquisition of a cognitive skill. After instructions, participants actually underpredict their round 1 mirror reading performance $((P_{AI1} - A_1)_{MP}$ =-5.42, *SE*=3.63, *F(1,37)=2.23*, *p*=.08, (U)=73.7%, (O)=31.1%) and then become even more pessimistic after experience. After practicing, participants significantly underpredict their first trial performance $(P_{AP1} - A_1)_{MP}$ =-8.00, *SE* =2.32, *F* (1,37) = 11.91, *p*=.0005, (U)=78.9%, (O)=18.4%).

After instruction, participants predicted an improvement of 11.13 (*SE*=2.63) words between rounds 1 and 3, while actually improving by 8.23 (*SE*=.82) words (see table 1). The overconfidence in their slope predictions is directional $((P_{AI3} - P_{AI1}) - (A_3 - A_1)_{MP}$ =2.89, *SE*=2.73, *F*(1,37)=1.12, ns (U)=44.7% , (O)=50%). As in earlier experiments, the overconfidence dissipates after initial experience. Comparing after instructions slope predictions with after experience slope predictions reveals that participants significantly reduce their slope predictions following experience $((P_{AI3} - P_{AI1}) - (P_{AP3} - P_{AP1})$ =-4.71, *SE* = 2.52, *F(1,37)*=3.47, *p*=.035, (I)=13.2%, (D)=55.3%). Participants also significantly underpredict their performance improvement after experience $((P_{AP3} - P_{AP1}) - (A_3 - A_1)_{MP}$ =-1.82, *SE* =.83,*F* (1,37) = 4.79, *p*=.0175, (U)=60.5%, (O)=26.3%).

Single Predictions Condition

In the single prediction condition (SP), participants predict only after experience. After experience, participants are pessimistic significantly underpredicting their first trial performance $(P_{AP1} - A_1)_{SP}$ =-11.12, *SE* =1.55, *F* (1,40) = 51.51, *p*=.000, (U)=87.8, (O)=8.3%). Also consistent with the All Thumbs effect, we find that participants significantly underpredict their performance improvement $(((P_{AP3} - P_{AP1}) - (A_3 - A_1)_{SP}$ =-2.22, SE= .88, $F\,(1,40) = 6.34$, p=.008, (U)=56.1%, (O)=31.73%).

Single vs. Multiple Predictions

Finally, we examine if predictions are influenced by repeated measurement. Comparisons of the after experience predictions with actual performance in each condition reveals no significant difference between the after experience slope predictions in the single prediction and multiple prediction conditions $((P_{AP3} - P_{AP1})\text{-}(A_3 - A_1)$, $F(1,77)$=.11, ns). There is also no significant difference between the after experience point predictions across conditions $((P_{AP1} - A_1)$, $F(1,77)$=1.29, p=ns). The results indicate that initial prediction elicitations do not significantly impact subsequent learning curve predictions.

3.5.3. Discussion

The results of this experiment indicate that the all thumbs effect is also generalizable to cognitive learning tasks. Consistent with prior experiments, participants underpredict their learning curves following experience with the task. Although we do not find initial overconfidence in the point predictions of the mirror reading task, we find that participants are initially overconfident in their rate of learning. Additionally, the evidence from the experiment eliminates the possibility that the effect is caused by demand artifacts resulting from repeated elicitations of performance predictions.

3.6 Conclusion

From snow-boarding to using the internet to mastering a new statistical pack or operating a new kind of wheelchair, many products require acquisition of new skills. Many opportunities both for business and for consumers are undoubtedly missed when new goods and services are abandoned after adverse initial experiences. Thus, one recent doctoral dissertation reported that 50% of products returned to electronic stores that consumers claim to be defective are actually fully functional (den Ouden et al., 2006), and that consumers spent an average of only 20 minutes trying to operate a new product before they gave up.

Having documented the All-Thumbs effect, there are many obvious and important lines of follow-up research. First, given the social welfare consequences of the phenomenon, it could be helpful to attempt to devise ways to debias the effect, e.g., through product design changes or through the provision of information about learning curves. It might also be helpful to incorporate incentives to induce people to work through the all-thumbs phase, despite pessimistic predictions of progress, to the point where they become better calibrated about their own learning curve. Finally, both of these goals would be facilitated by a better understanding of the cognitive and neural underpinnings of the under-prediction of learning curves. There is now a large volume of research on the neural underpinnings of skill learning, but, despite the importance of predictions of such learning, no research to the best of our knowledge addressing the neural basis of pessimism with regard to such learning. For example, is the All- Thumbs effect more pronounced for tasks that have an all-or-none, insight-like, quality? More generally, we hope that an awareness of the phenomenon will lead to new tactics to remedy some of its most pernicious effects, most notably the tendency to quickly give up on tasks which, if persisted on, would yield substantial benefits.

References

Ackerman, P. (1987). Individual Differences in Skill Learning: An Integration of Psychometric and Information Processing Perspectives. *Psychological Bulletin*, 102 (1), 3-27.

Anderson, J. (1982). Acquisition of Cognitive Skill. *Psychological Review*, 89 (4), 369-406.

Argote, L., Beckman, S. L., & Epple, D. (1990). The persistence and transfer of learning in industrial settings. *Management Science*, *36*(2), 140-154.

Argote, L., & Ingram, P. (2000). Knowledge transfer: A basis for competitive advantage in firms. *Organizational Behavior and Human Decision Processes*, 82(1), 150-169.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R.S. Wyer, Jr. & T.K. Srull (Eds.), *Handbook of social cognition: Basic processes* (2nd ed., pp. 1-40). New York, NY: Guilford Press.

Buehler, R., & Griffin, D. (2003). Planning, personality, and prediction: The role of future focus in optimistic time predictions. *Organizational Behavior and Human Decision Processes, 92,* 80-90.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology, 67*, 366-381.

Burson, K., Klayman, J., (2007). Judgments of performance: The relative, the absolute and the in-between. Working Paper. University of Michigan.

Byram, S. J. (1997). Cognitive and motivational factors influencing time predictions. *Journal of Experimental Psychology: Applied*, *3*(3), 216-239.

Card, S.K., Moran, T.P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

den Ouden, E., Yuan, L., Sonnemans, P., & Brombacher, A. (2006). Quality and reliability problems from a consumer's perspective: An increasing problem overlooked by businesses? *Quality and Reliability Engineering International*, *22,* 821-838.

Dunning, D., Griffin, D., Milojkovic, J., & Ross, L. (1990). The overconfidence effect in social predictions. *Journal of Personality and Social Psychology*, *58*(4), 568-581.

Dunning, D., Meyerowitz, J.A., & Holzberg, A.D (1978). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp.324-333). New York: Cambridge University Press..

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*, 42-54.

Gray, W.D., & Orasanu, J.M. (1987). Transfer of cognitive skills. In S.M.Cormier & J.D. Hagman (Eds.), *Transfer of learning: Contemporary research and applications*. New York: Academic Press.

Gregan-Paxton, J., & Roedder John, D., (1997). Consumer learning by analogy: A model of internal knowledge transfer. *Journal of Consumer Research*, *24*(3), 266-284.

Haider, H. and Frensch, P.A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard, Delaney et al. and Palmeri. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 392-406.

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*, 185–207.

Hill, N.H., & Schneider, W. (2006). Brain changes in the development of expertise: Neurological evidence on skill-based adaptations. In K. A. Ericsson, N. Charness, P. Feltovich, & R. Hoffman (Eds.), *Cambridge Handbook of Expertise and Expert Performance* (pp. 653-682). New York: Cambridge University.

Hinds, P. (1999). The curse of expertise: The effects of expertise and debiasing methods on predictions of novice performance. *Journal of Experimental Psychology: Applied*, *5*(2), 205-221.

Jagacinski, R.J., Isaac, P.D., & Burke, M.W. (1977). Application of signal detection theory to perceptual-motor skills: Decision processes in basketball shooting. *Journal of Motor Behavior*, *9*, 225-234.

Johnson, E. J., Bellman, S., & Lohse, G. L. (2003). Cognitive lock-in and the power law of practice. *Journal of Marketing, 67(*2*)*, 62-82.

Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. Psychological Review, 107, 384-396.

Kane, A., Argote, L., & Levine, J. (2005). Knowledge transfer between groups via personnel rotation: effects of social identity and knowledge quality. *Organizational Behavior and Human Decision Processes, 96(1),* 56-71.

Keren, G. (1987). Facing uncertainty in the game of bridge: A calibration study. *Organizational Behavior and Human Decision Processes*, *39,* 98-114.

Koehler, D.J., & Poon, C.S.K. (2006). Self predictions overweight strength of current intentions. *Journal of Experimental Social Psychology, 42,* 517-524.

Kolers, P. A. (1975). Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(6), 689-701.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*(2), 147-162.

Koriat, A., Ma'ayan, H., Sheffer, L., & Bjork, R.A. (2006). Exploring a mnemonic debiasing account of the underconfidence-with-practice effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 595-608.

Lewandowsky, S., Kalish, M., & Ngang, S.K. (2002). Simplified learning in complex situations: Knowledge partitioning in function learning. *Journal of Experimental Psychology: General*, *131*(2), 163-193.

Lichtenstein, S., & Fischhoff, B., (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance,* 20, 159-183.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of probabilities: The state of the art to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). New York: Cambridge University Press.

Logan, G. D. (1988). Toward an instance theory of automatization, *Psychological Review, 95*(4), 492-597.

Martin, P. R., & Fernberger, S. W. (1929). Improvement in memory span. *American Journal of Psychology, 41,* 91-94.

Menon, G., & Raghubir, P. (2003). Ease of retrieval as an automatic input in judgment: A mere accessibility framework? *Journal of Consumer Research, 30*(2), 230-243.

Moore, D., & Small, D. (2007). Error and Bias in Comparative Judgment: On being bothe better and worse than we think we are. *Journal of Personality and Social Psychology,* Forthcoming.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*(2), 297-326.

Newby-Clark, I., Ross, M., Buehler, R., Koehler, D., & Griffin, D. (2000). People focus on optimistic scenarios and disregard pessimistic scenarios while predicting task completion times, *Journal of Experimental Psychology: Applied, 6*(3), 171-182.

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Sawyer, A. (1975). Demand Artifacts in Laboratory Experiments in Consumer Research. *Journal of Consumer Research*, 1, 20-30.

Scheck, P., & Nelson, T.O. (2005). Lack of pervasiveness of the underconfidence-with-practice effect: Boundary conditions and an explanation via anchoring. *Journal of Experimental Psychology: General*, *134*(1), 124-128.

Schmidt, R.A. (1988). *Motor control and learning: A behavioral emphasis*. Champaign, IL: Human Kinetics Publishers, Inc.

Serra, M.J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1258-1266.

Shiffrin, R. M., & Dumais, S.T. (1981). The development of automatism. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 111-140). Hillsdale, NJ: Erlbaum.

Shiffrin, R.M., & Schneider, W. (1977). Controlled and automatic human information processing. *Psychological Review*, *84*, 127-190.

Snoddy, G.S. (1926). Learning and stability. *Journal of Applied Psychology*, *10*, 1-6.

Swift, E. J. (1905). Memory of a complex skillful act. *American Journal of Psychology*, *16*(1), 131-133.

Swift , E. J. (1904). The acquisition of skill in type-writing: A contribution to the psychology of learning. *Psychological Bulletin*, *1*(August), 295-305.

Thorndike, E. L., & Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions. *Psychological Review*, *8*(4), 384-395.

Footnotes

i For an alternative perspective see Haider & Frensch, 2002, Heathcote, Brown & Mewhort, 2000

ii Whether people persist will also depend critically on their ability to tolerate the immediate frustration of poor performance.

iii We would like to thank an anonymous reviewer for suggesting this analysis and providing this insight.

iv $((P_{AE3} - P_{AE1})-(A_3-A_1))/2=-1.58$, F(1,36)=10.27, p=.0015, U=68%, O=19%; ((PAI3 - PAI1)-(PAE3-PAE1))/2=.76, =*F(1,36)*=6.95, *p*=.006, D=46%, I=14%; (PAI3)-(PAE3)= 2.43,F(1,36)=4.17, p=.024, D=57%, I=27%; PAI1 – A1= 2.35, *F(1,36)*=1.49, *p*=.11, U=46%, O=43%; P2* – A*, F(1,36)=20.20, p=.000,

v In both knot conditions, participants remained underconfident in their trial 2 predictions (easy knot $P_{2*} - A_*$, F(1,36)=6.21, *p*=.008; hard knot condition $P_{2*} - A_*$, *F(1,37)=13.00, p=.001*) but participants in the mirror tracing condition quickly calibrated ($P_{2*} - A_* = .11$, *F(1,35)=.025, p=NS*) and were not underconfident.

Table 1: Experiment 1 Predictions and Performance

	Trial 1		Trial 2		Trial 3		Trial 4		Trial 5	
	Mean	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
Actual Knots Tied	10.17	.85	14.87	1.02	17.87	1.10	19.91	1.24	21.34	1.25
P_{AI}	11.92	1.40	15.08	1.62	17.30	1.73	20.15	1.91	22.00	2.08
P_{AE}	9.28	1.05	11.57	1.35	13.36	1.5	14.72	1.64	16.43	1.85
P_{2n}			13.57	1.04	15.64	1.22	16.98	1.33	18.34	1.40
P_{3n}					16.72	1.06	18.04	1.15	19.26	1.22
P_{4n}							19.08	1.20	20.43	1.26
P_{5n}									21.38	1.33
N=	53		53		53		53		53	

Table 2a: Predictions and Performance for Experiment 2--Single Task

	Trial 1		**Trial 2**		**Trial 3**		**Trial 4**	
	Mean	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
Actual	4.50	0.74	8.09	1.02	11.77	1.29	15.73	1.33
P_{AI} (Single Task)	6.27	0.86	7.73	0.95	9.00	1.03	9.91	1.12
P_{AI} (Dual Task)	4.77	0.81	5.95	0.93	6.95	0.97	7.36	0.98
P_{AE} (Single Task)	3.09	0.55	3.59	0.56	4.00	0.60	4.50	0.64
P_{AE} (Dual Task)	2.05	0.38	2.55	0.44	2.95	0.52	3.41	0.55
P2			5.36	0.83	5.86	0.84	6.36	0.83
P3					9.68	1.24	10.95	1.33
P4							13.36	1.50

Table 2b: Predictions and Performance for Experiment 2--Dual Task

	Trial 1		Trial 2		Trial 3		Trial 4	
	Mean	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
Actual	3.23	0.60	7.54	0.90	10.65	1.04	13.88	1.31
P_{AI} (Single Task)	5.88	.83	7.46	1.00	9.19	1.22	10.54	1.43
P_{AI} (Dual Task)	3.96	.86	5.27	1.02	6.42	1.21	7.38	1.38
P_{AE} (Single Task)	3.65	0.49	4.42	0.57	5.31	0.71	6.08	0.88
P_{AE} (Dual Task)	2.35	0.37	2.88	0.43	3.54	0.52	4.04	0.59
P2			4.88	0.80	5.50	0.92	6.31	0.93
P3					9.15	1.18	10.27	1.37
P4							12.38	1.38

Table 3a: Experiment 3 Predictions and Performance in Task 1

Condition		Trial 1 Mean	Trial 1 S.E.	Trial 2 Mean	Trial 2 S.E.	Trial 3 Mean	Trial 3 S.E.
Mirror Tracing	Actual	7.92	0.97	13.47	1.42	17.94	1.63
	P_{AI}	9.97	1.72	12.58	1.95	15.75	2.40
	P_{AE}	5.11	1.11	6.67	1.29	8.19	1.48
	P_{2n}			9.83	1.15	11.61	1.32
	P_{3n}					17.17	1.93
Easy Knot	Actual	9.30	0.78	14.00	0.94	16.76	1.14
	P_{AI}	11.65	1.75	15.03	1.93	17.46	2.14
	P_{AE}	10.73	1.47	13.27	1.66	15.03	1.80
	P_{2n}			12.54	0.86	14.59	.99
	P_{3n}					16.24	1.04
Hard Knot	Actual	8.87	0.73	11.92	.89	14.53	.92
	P_{AI}	12.76	2.33	17.39	2.80	22.00	3.75
	P_{AE}	9.79	1.38	12.79	1.74	15.08	1.90
	P_{2n}			11.68	.91	13.39	1.07
	P_{3n}					14.39	.98

Table 3b: Experiment 3 Predictions and Performance in Task 2 (Double Bowline Knot)

		Trial 1		Trial 2		Trial 3	
Condition		**Mean**	**S.E.**	**Mean**	**S.E.**	**Mean**	**S.E.**
Mirror Tracing	Actual	10.25	.95	14.22	1.21	17.86	1.27
	P_{AI}	8.58	.87	12.44	1.10	16.33	1.47
	P_{AE}	7.89	.88	10.78	1.04	13.78	1.33
	P_{2n}			14.33	1.18	17.47	1.45
	P_{3n}					17.67	1.37
Easy Knot	Actual	9.24	.83	12.49	.92	15.22	1.12
	P_{AI}	9.19	.92	12.35	1.04	14.78	1.15
	P_{AE}	7.89	.74	10.49	.90	12.70	1.04
	P_{2n}			11.84	.93	14.00	1.04
	P_{3n}					14.51	1.01
Hard Knot	Actual	10.55	.79	15.53	1.25	19.13	1.39
	P_{AI}	8.82	.83	12.03	1.01	14.37	1.09
	P_{AE}	8.89	.90	11.42	1.04	13.29	1.10
	P_{2n}			13.50	1.30	15.03	1.31
	P_{3n}					17.13	1.39

Table 4: Experiment 4 Predictions and Performance

	Trial 1		Improvement Between Trial 1 and Trial 3	
	Mean	**S.E.**	**Mean**	**S.E.**
Actual	4.91	0.85	6.82	0.64
$\mathbf{P_{AI}}$	7.85	1.20	4.73	1.04
$\mathbf{P_{AE}}$	3.24	.42	2.21	.33
$\mathbf{P_{2n}}$	N/A		3.36	.42

Table 5a: Experiment 5: Multiple Predictions Condition--Predictions and Performance

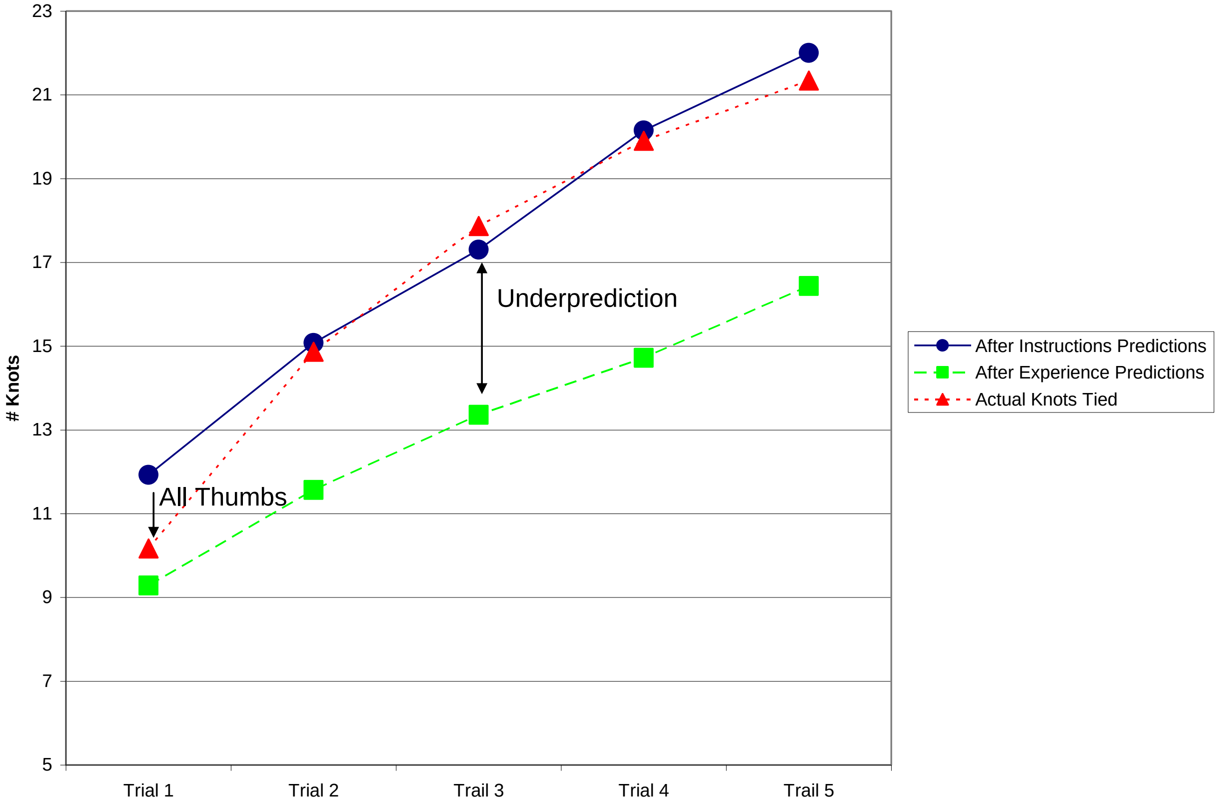
	Trial 1		**Improvement Between Trial 1 and Trial 3**	
	Mean	**S.E.**	**Mean**	**S.E.**
Actual	29.08	1.74	8.23	.82
P_{AI}	23.66	3.35	11.13	2.63
P_{AE}	21.08	2.42	6.42	.63

Table 5b: Experiment 5: Single Predictions Condition--Predictions and Performance

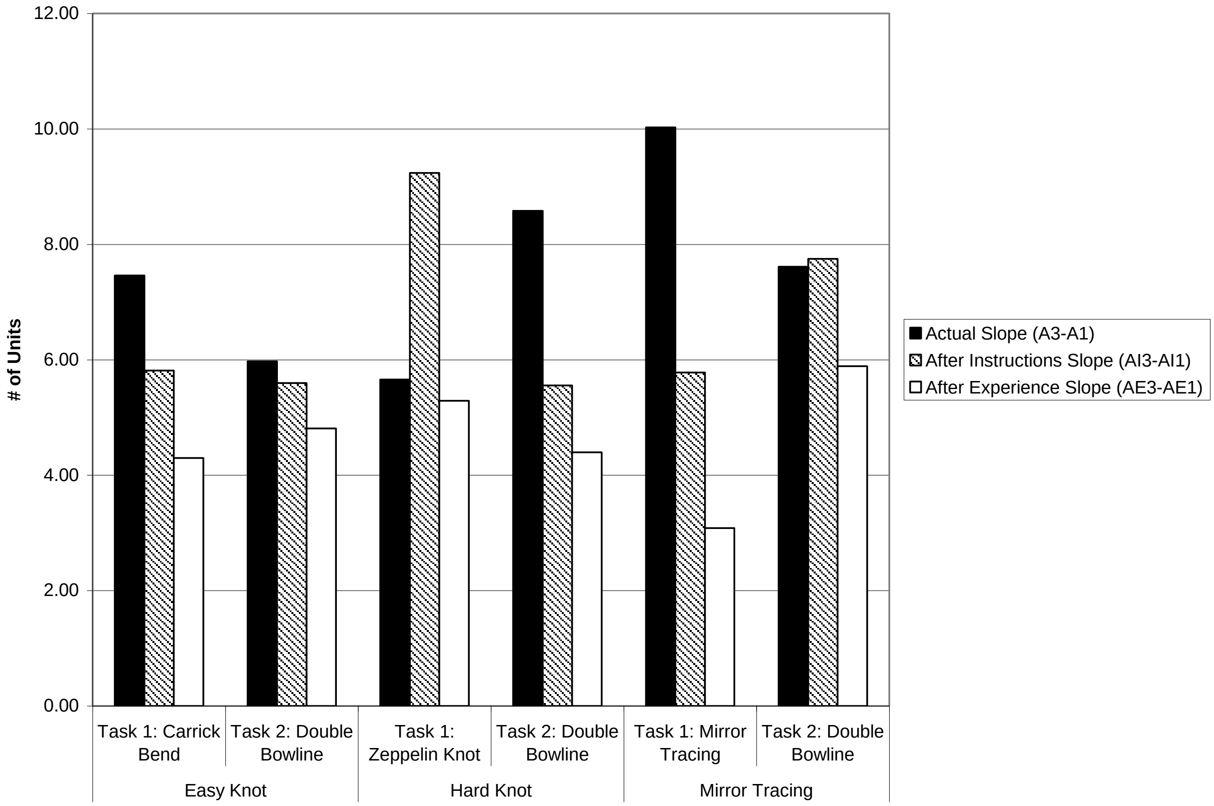
	Trial 1		**Improvement Between Trial 1 and Trial 3**	
	Mean	**S.E.**	**Mean**	**S.E.**
Actual	30.85	1.53	8.39	.66
P_{AE}	19.73	1.93	6.17	.61

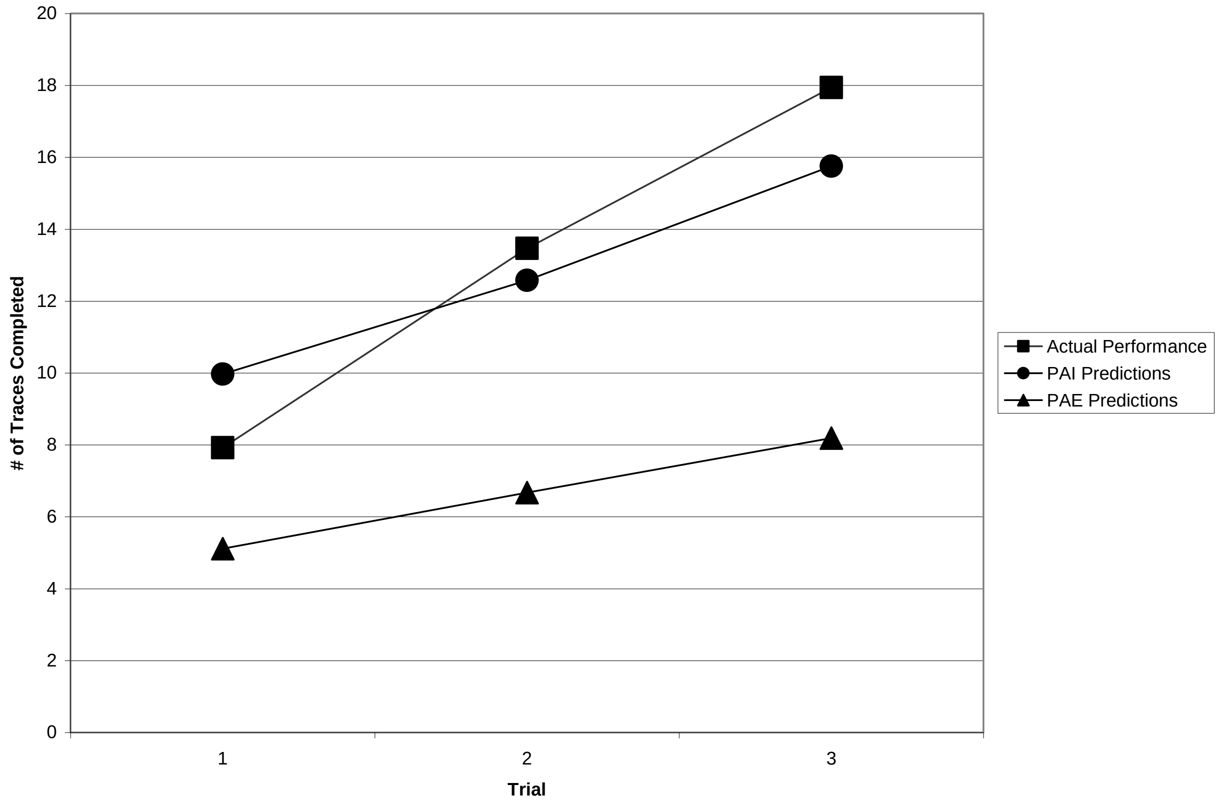
Figure Captions

Figure 1: Results of Experiment 1
Figure 2: Results of Experiment 3
Figure 3a: Results Mirror Tracing Task
Figure 3b: Results Easy Knot Task
Figure 3c: Results Hard Knot Task
Figure 4: Experiment 4 Slope Prediction Results
Figure 5: Stimuli for Experiment 5

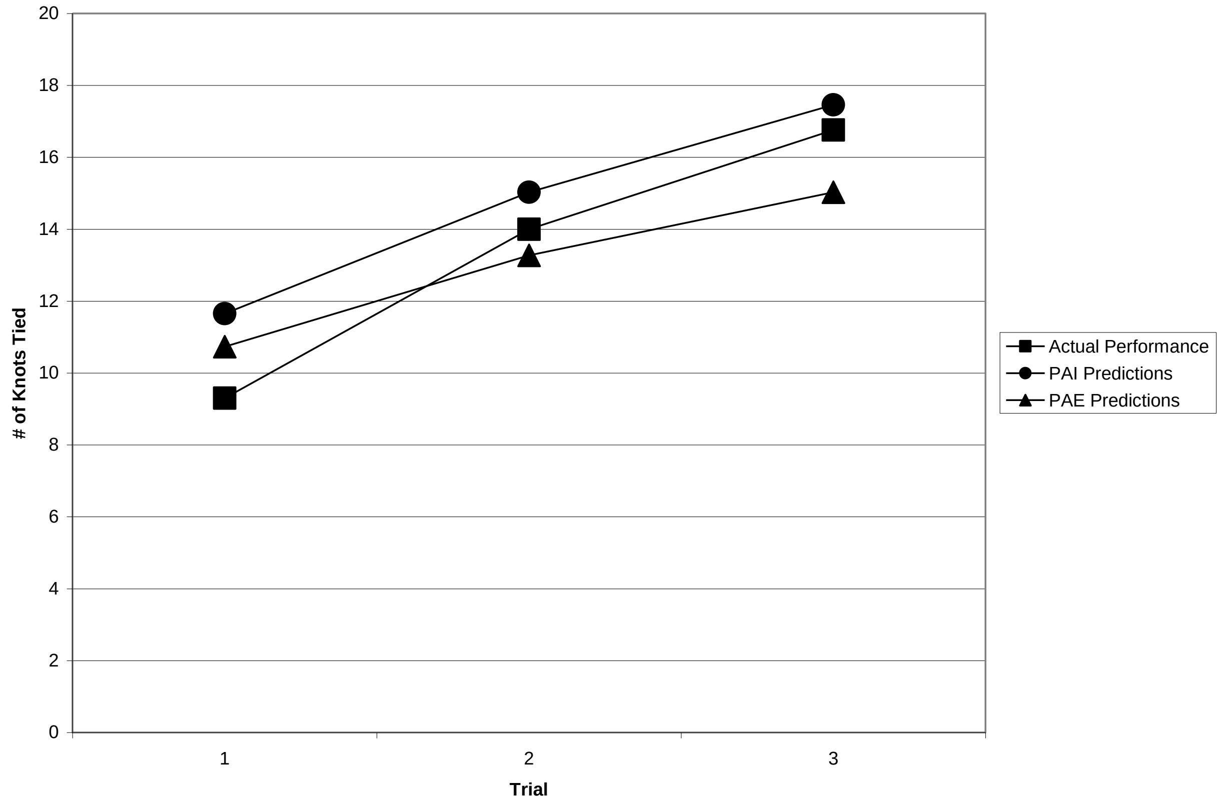


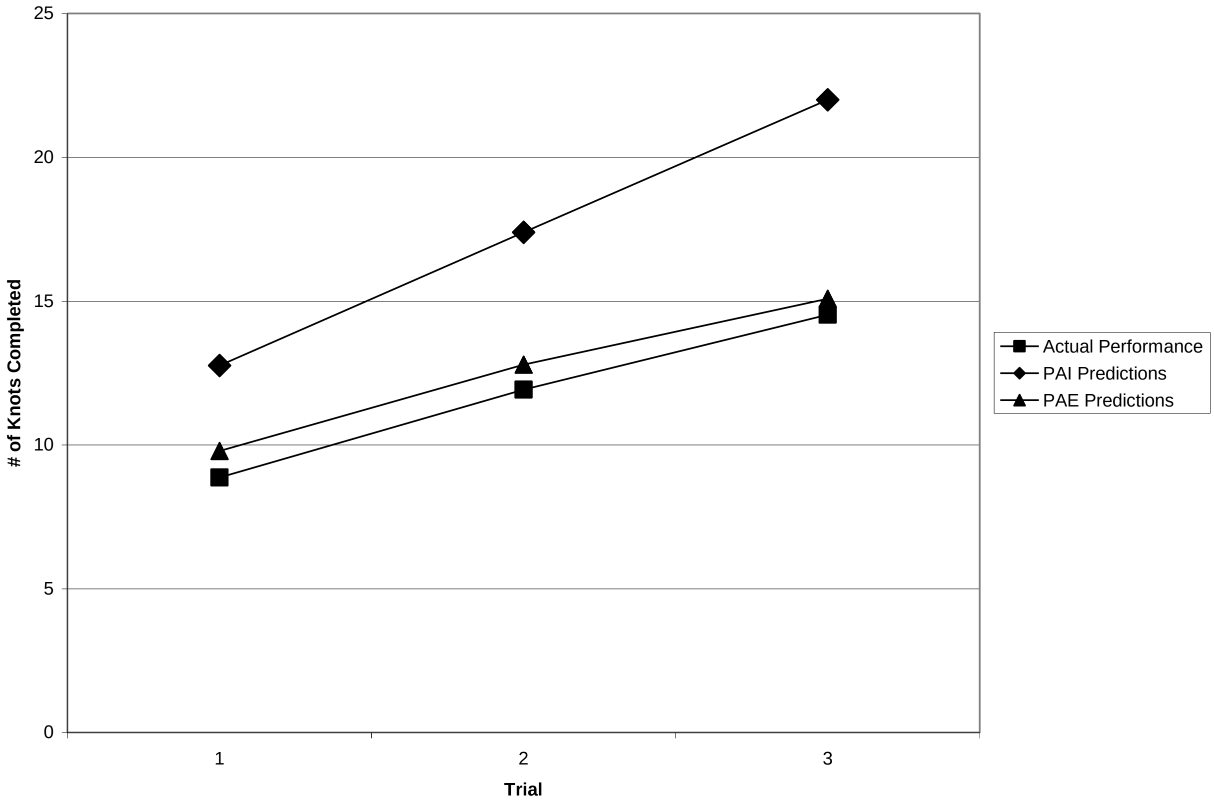
23
21
19
17
15
13
11
9
7
5
Knots
Trial 1
Trial 2
Trail 3
Trial 4
Trail 5
Underprediction
All Thumbs
After Instructions Predictions
After Experience Predictions
Actual Knots Tied


12.00
10.00
8.00
6.00
4.00
2.00
0.00
# of Units
Actual Slope (A3-A1)
After Instructions Slope (AI3-AI1)
After Experience Slope (AE3-AE1)
Task 1: Carrick Bend
Task 2: Double Bowline
Task 1: Zeppelin Knot
Task 2: Double Bowline
Task 1: Mirror Tracing
Task 2: Double Bowline
Easy Knot
Hard Knot
Mirror Tracing

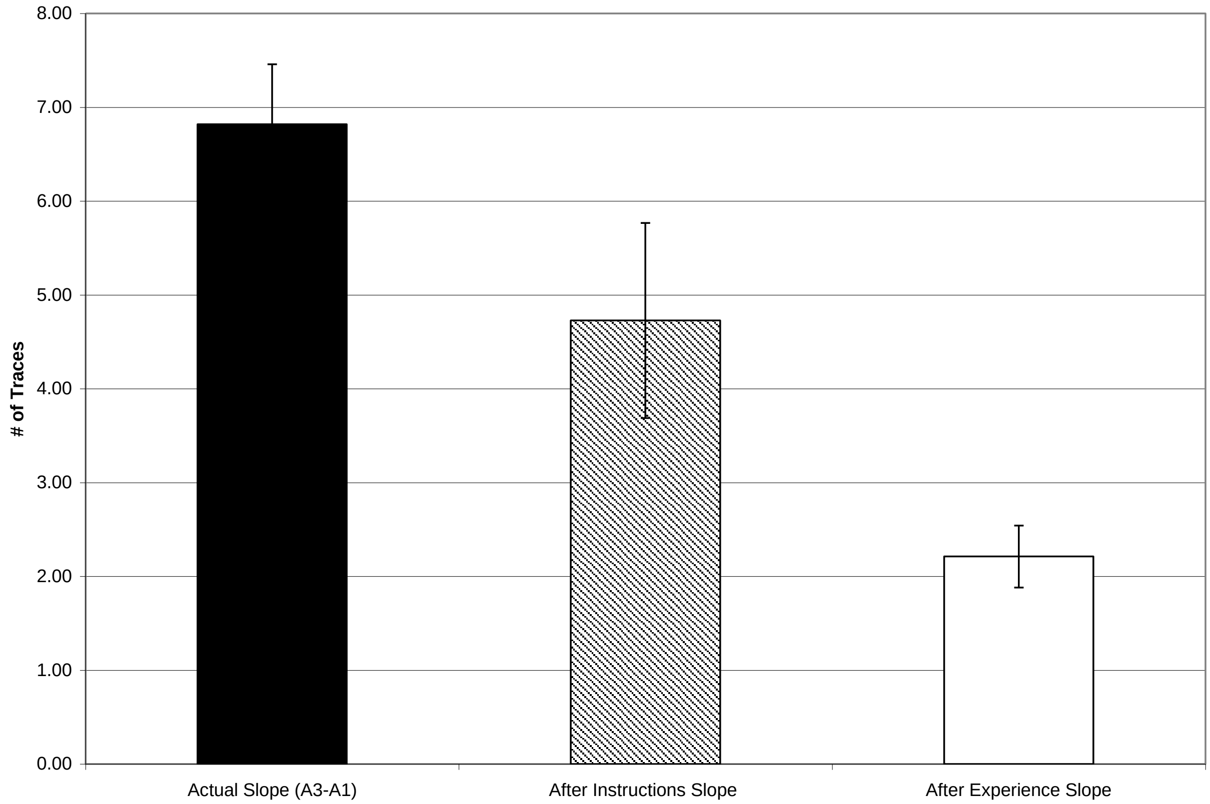


20
18
16
14
12
10
8
6
4
2
0
# of Traces Completed
1
2
3
Trial
Actual Performance
PAI Predictions
PAE Predictions


20
18
16
14
12
10
8
6
4
2
0
# of Knots Tied
1
2
3
Trial
Actual Performance
PAI Predictions
PAE Predictions



25
20
15
10
5
0
# of Knots Completed
1
2
3
Trial
Actual Performance
PAI Predictions
PAE Predictions


8.00
7.00
6.00
5.00
4.00
3.00
2.00
1.00
0.00
# of Traces
Actual Slope (A3-A1)
After Instructions Slope
After Experience Slope

PRESSING		CAROLINA		FLEXIBLE		MIDNIGHT
NARRATOR		EFFLUENT		PROCLAIM		EMPLOYEE
THREATEN		PROCEEDS		CARDINAL		CHAMPION
COMPRISE		INFANTRY		SENSIBLE		MONUMENT
RESTRICT		ALUMINUM		MANIFOLD		REVEREND
EPIDEMIC		SUNLIGHT		DRAINAGE		BROADWAY
SETTLING		MINIMIZE		PLEADING		LAUGHTER
COMMUTER		DRIVEWAY		APPLAUSE		MERCHANT
HESPERUS		PETITION		TERMINAL		CEREMONY
PLEASING		BLOCKADE		COLORADO		SITUATED
CHILDISH		MONOPOLY		HONESTLY		MURDERER
LIFETIME		FOREHEAD		SINISTER		COLONIAL
SCOTTISH		HORRIBLE		EVERYDAY		HERITAGE
AIRPLANE		CONTENTS		IMPERIAL		FRICTION
GIGANTIC		TOUCHING		ANTIBODY		INTERVAL
BLESSING		DAYLIGHT		STRAINED		OVERHEAD
MULTIPLY		PENTAGON		MONSIEUR		BATHROOM
DRILLING		GRADIENT		DIALOGUE		STIRRING