Cultivating Expertise in Informal Reasoning

Tim van Gelder and Melanie Bissett, University of Melbourne
Geoff Cumming, La Trobe University

Abstract People generally develop some degree of competence in general informal reasoning and argument skills, but how do they go beyond this to attain higher expertise? Ericsson has proposed that high-level expertise in a variety of domains is cultivated through a specific type of practice, referred to as "deliberate practice." Applying this framework yields the empirical hypothesis that high-level expertise in informal reasoning is the outcome of extensive, deliberate practice. This paper reports results from two studies evaluating the hypothesis. University student participants completed 12 weeks of deliberate practice in informal reasoning. Quantity of practice was recorded by computer, and additionally assessed via self-report. The hypothesis was supported: Students in both studies showed a large improvement, and practice, as measured by computer, was related to amount of improvement in informal reasoning. These findings support adopting a deliberate practice approach when attempting to teach or learn expertise in informal reasoning.

The general skills of informal reasoning and argumentation, which in the following we simply call informal reasoning, are often ill understood and poorly deployed, even among those in the upper tiers of our educational systems (Graf, 2003). In her important book, *The Skills of Argument*, Deanna Kuhn reported on an extensive study of a wide range of people. She found that, for each of the major subskills of informal argumentation, around half of her subjects did not successfully exhibit that subskill (Kuhn, 1991). For example, while participants readily held opinions on controversial matters, when asked to give evidence in support of their opinions, in over half of the cases their responses did not constitute genuine evidence (let alone good evidence). Other studies have come up with similarly bleak results (Means & Voss, 1996; Perkins, 1985; Perkins, Allen, & Hafner, 1983).

Of course, almost all people do have some informal reasoning ability. They can follow, and often produce, basic inferences such as "you can't get on the bus, because you don't have a ticket." They have fragmented abilities in the range of argument skills investigated by Kuhn (1991). These abilities are deployed in many everyday circumstances. Informal reasoning appears at quite an early age (Stein & Miller, 1993) and continues to develop through secondary and tertiary education (Felton & Kuhn, 2001; Pascarella & Terenzini, 1991). A few people manage to become highly proficient.

The problem, then, seems to be that the natural development of informal reasoning often peters out while skills are still incomplete. Ordinarily, through standard processes of maturation, socialization, and education, people achieve a certain level of expertise, broadly comparable to that of their peers and adequate for most of their everyday purposes. In this paper, we call this "competence," allowing that competent people may have settled at widely varying levels of ability. The key point is that people rarely advance beyond that competence to genuine mastery of a coherent set of skills. Extraordinary efforts, pressures or opportunities might yield some improvement, but for most people this unfinished competence is a more or less stable state.

This poses a challenge for educators and trainers, especially those working at the higher ends of our educational systems: How can they help students progress beyond ordinary competence to achieve some level of mastery? To address instructional challenges of this kind, we need good theories of the psychological terrain. How, in general, are informal reasoning skills acquired? What is their developmental trajectory? What are the cognitive processes and mechanisms involved? What contexts and activities best promote growth in such skills? How is excellence achieved?

Given the pervasiveness and importance of informal reasoning, we might expect psychologists to know a considerable amount about these topics. However, there is a paucity of research on general informal reasoning and argumentation (Voss & van Dyke, 2001) and very little on how such skills are acquired. A smattering of studies address relevant issues, but there is no widely accepted, overarching account with explanatory depth or predictive power.

In this paper, we investigate one aspect of this topic: How are high levels of informal reasoning skill achieved? This question is quite different from the developmental problem of how people become competent in the first place. We wish to discover how people can climb above the plateau of competence they would have reached anyway through maturation, socialization, and standard education.

Our approach is to apply the dominant contemporary framework for understanding the acquisition of high-level expertise to the case of informal reasoning. That framework suggests the empirical hypothesis that advanced informal reasoning ability is achieved through large amounts of practice of a special sort. We report two studies that evaluated this hypothesis by comparing amount of skill gain with amount of practice of the appropriate kind among subjects who are already competent at informal reasoning.

Acquisition of Expertise: Deliberate Practice

How, in general, do people become highly skilled? Currently, the dominant theoretical approach to this question is the framework developed by Ericsson and colleagues (Ericsson & Charness, 1994; Ericsson, Krampe, & Tesche-Römer, 1993; Ericsson & Lehmann, 1996). Studying elite performance in a range of domains, they found a consistent pattern: Expertise at the highest levels results from very large amounts of practice with the following features: (a) it is consciously aimed at improvement; (b) it is done in regular periods of intense concentration and limited duration; (c) it involves exercises specially designed to enhance particular aspects of performance; (d) exercises are repeated until the desired level of performance is achieved; (e) exercises are graduated so as to build mastery of progressively more sophisticated skills; and (f) it is supervised by a specialist coach who shapes the practice regime and provides guidance and feedback. They described practice of this sort as "deliberate."

Ericsson and colleagues found that attaining the highest levels of expertise in all the domains they studied required thousands of hours of deliberate practice, with the very top levels of performance resulting from 10 years or more of intensive, structured, guided effort. Further, individual differences in the level of performance reached are related to the amount of deliberate practice undertaken, and there is little explanatory work remaining to be done by the folk notion of talent, once differences in deliberate practice are factored in (Ericsson et al., 1993).

Of course, this level of dedicated effort is quite rare. Typically, people engage in little if any deliberate practice, and their abilities reach a stable plateau well below their full potential. Ericsson and colleagues (1993) found that, after initial training, amount of experience in a domain is not related to level of expertise: Merely engaging in an activity does not improve skills beyond a certain level. The social tennis player or the medical specialist can go for years without getting any better.

We conjecture that high-level expertise in informal reasoning is acquired in much the same way as in other domains. The Ericsson framework yields what we call the deliberate practice hypothesis: Advanced informal reasoning skills are achieved just to the extent that one engages in large amounts of deliberate practice. This hypothesis implies that deliberate practice will accelerate gains in informal reasoning skills at all levels of ability; that deliberate practice will allow individuals to go beyond ordinary competence; that genuine mastery of informal reasoning will take very large amounts of deliberate practice; and that reasoning activities falling short of full deliberate practice may nevertheless enhance abilities, but only to the extent that they possess the features of deliberate practice.

Evaluating the Deliberate Practice Hypothesis

To evaluate the deliberate practice hypothesis, we must study how informal reasoning skills change in competent people undertaking large amounts of deliberate practice. This presented some formidable practical challenges.

Obtaining participants. Ideally, we would set up a rigorous experiment in which equivalent groups of competent participants engage in deliberate practice whose extent and nature is carefully controlled. However, in a complex skill like informal reasoning, only a large amount of practice will make any appreciable difference to expertise. Given familiar resource constraints, a standard experimental design is out of the question. We cannot require, or pay, potential participants such as psychology undergraduates to engage in demanding practice for hundreds of hours.

The only feasible alternative is a quasi-experiment in which we study people who are engaging in deliberate practice in some context independent of our research. One convenient context is an undergraduate introductory reasoning subject. Such subjects are taught at almost every college and university, under titles such as "Introduction to Logic" or "Critical Thinking," and they typically promote themselves with the promise of enhanced reasoning abilities. They involve large numbers of students – potential research participants – investing a certain amount of targeted effort towards improving their reasoning skills. This amount of effort is insufficient for top-level expertise, but may be enough to generate a detectable effect of the kind we expect to find; according to Anderson, "Significant changes in human potential take at least 100 hours to achieve, approximately the time investment required for a semester course" (Anderson, 2002).

Our approach, then, was to study how deliberate practice affects informal reasoning skills among students enrolled in a one-semester introductory reasoning subject. We can gain insight into the relationship between amount of practice and amount of gain by exploiting the fact that such students vary quite widely in the amount of effort they invest. Carefully measuring their practice activities can provide a basis for estimating the correlation between practice and gain.

Implementing a deliberate practice training regime. Another challenge is ensuring that the students engage in genuine deliberate practice. Unfortunately, conventional introductory reasoning subjects approximate deliberate practice only poorly. Cultivating expertise is but one of many pedagogical objectives. The strategy for improving informal reasoning skills is typically didactic, focusing primarily on conveying knowledge and understanding, making the traditional but ill-founded assumption that expertise or "knowledge how" will flow automatically from theory or "knowledge that" (Dreyfus & Dreyfus, 1986). These subjects do involve practice, but it falls short of proper deliberate practice in at least two major regards: Practice is not sufficient to secure mastery at each stage (hence the spread of grades on any given assignment), and the "coaching" (direct attention from the instructor) is very thinly spread.

Thus, in order to conduct our studies, we had to develop a new way of teaching introductory reasoning. The "Reason!" approach was designed from the outset to maximize informal reasoning skill gains, and incorporates key features of deliberate practice. The approach centres on sets of exercises specially designed to enhance particular aspects of informal reasoning. These exercises progress from simple to complex, and are sufficiently numerous to promote mastery through repetition at each stage.

An important aspect of deliberate practice is direct supervision by an expert coach providing tailored guidance and feedback, such as that provided for athletes and musicians aspiring to top-level performance. In systems of mass education, including standard colleges and universities, this kind of coaching is far too expensive. If students are to engage in anything resembling deliberate practice, some substitute must

War is declared by one state upon another state. Al Qaeda isn't a state, therefore the U.S. hasn't declared war on it.

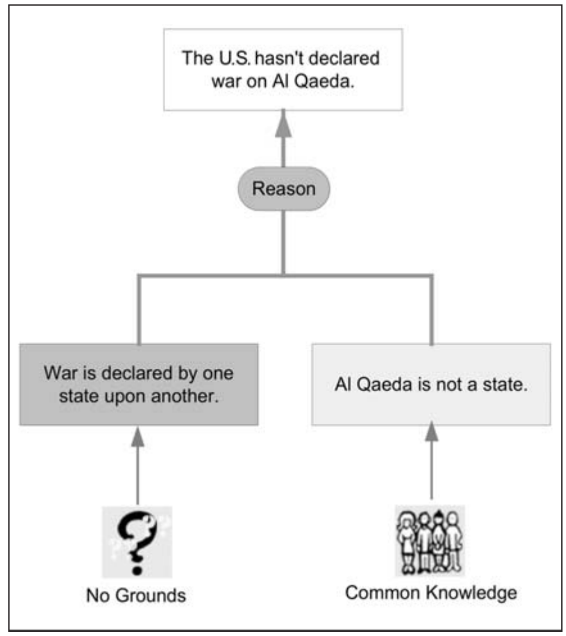

Figure 1. A sample exercise in the Reason! approach. The exercise provides a piece of everyday informal reasoning, such as the text above (a posting to the Internet). The student must produce an argument map displaying the structure of the reasoning and his/her evaluation of it, as illustrated on the right. Structure is displayed using boxes, lines, and position in space; evaluative aspects are displayed using icons and colours. Note that this exercise is one of the very simplest in the Reason! approach.

be found. To this end, we designed a software package, Reason!Able, to function as a "practice environment" (van Gelder, 2001; van Gelder & Bulka, 2000). Reason!Able compensates for the lack of extensive personal coaching in two ways. First, it provides scaffolding that heavily constrains students' reasoning activities, and within this framework, it provides some context-sensitive guidance. Second, through the use of argument mapping, the software exploits the available human intelligence (that of the instructor, and the students themselves) to provide more, and more effective, feedback than is usually provided in conventional instruction.

Any reasoning activity must work with some kind of representation of propositions and their logical relationships. In Reason!Able, reasoning is represented using argument maps. These are graphical presentations of the structure of reasoning or argumentation; typically, they are "box and arrow" diagrams (van Gelder, 2002). Almost all exercises in the Reason! approach involve constructing and modifying argument maps. Figure 1 provides an example.

Argument maps have been around for a century or more (Buckingham Shum, 2002) but are not often used, despite their considerable advantages over standard prose as a medium for presenting reasoning. This is in part because producing argument maps, beyond the simplest cases, is a slow and laborious business. Reason!Able helps students rapidly and easily construct, modify, and distribute (e.g., print or e-mail) argument maps. Using the software, they can focus on the reasoning activities, unencumbered by the tedious labour of producing diagrams by hand.

Reason!-style argument maps enhance the quantity and quality of feedback by making the thinking involved in any reasoning activity very explicit. Students can then generate feedback for themselves by comparing their diagrams with model diagrams, and instructors can provide targeted feedback by pointing to relevant features of the students' diagrams.

Although the Reason! approach embodies many key features of deliberate practice, students' actual reasoning activities may fall short of full-blooded deliberate practice in a number of ways. Our claim is that when using the Reason! approach, the students as a group approximate deliberate practice more closely than in standard instruction, and about as closely as is possible given the severe resource constraints typical of undergraduate education.

Measuring gains. Measuring gains in informal reasoning skill requires assessing each participant's expertise at various points. Ideally, we would make repeated measurements during the course of the training regime. However, practice effects undermine the validity of such data, and multiple equivalent forms of the same test are not available. Therefore, we used two testing sessions, pre and post, using the two available forms of an independently developed and objective (rather than self-report) test of informal reasoning.

Measuring practice. Our research design left us little control over how much practice each participant was doing. Students were largely free to vary the amount of effort they invested. Therefore, we had to measure the amount of practice they were actually doing as accurately as possible.

The most obvious approach is to simply ask the students themselves. We did this using an end-of-semester questionnaire, on which students provided their own subjective, retrospective estimates of the extent of their practice. However, such estimates can be invalid. One of the biggest challenges was obtaining an objective record of the amount of practice students were actually doing.

The Reason!Able software provided an avenue for meeting this challenge. Since almost all practice in the Reason! approach is supposed to be done with the software, we could use the software to measure how much practice each student was doing. The software was adapted so that every single action (e.g., mouse click) was recorded and uploaded to a server. This objective data gave us an independent estimate of the amount of practice students were doing, as well as an objective check on the reliability of the self-report data.

Predictions

To summarize, our approach to evaluating the deliberate practice hypothesis was to employ a quasi-experimental design using students enrolled in a one-semester introductory reasoning subject. We developed a pedagogical approach based as far as possible on deliberate practice, using argument mapping supported by custom-built software. We measured gains by pre- and post-testing using an objective, independently established test of reasoning skills and measured practice via questionnaires and software data logging.

In this context, the deliberate practice hypothesis predicts that participants' informal reasoning skills will improve substantially over the semester, and that the more deliberate practice they do, the more they will improve. We conducted two studies aimed at testing these predictions.

Method

Participants

All students enrolled in "Critical Thinking: The Art of Reasoning," a first-year undergraduate subject taught in the Department of Philosophy at the University of Melbourne in 2002 ($N = 146$) and 2003 ($N = 146$) voluntarily consented to participate. Of these, 117 completed the pre and post measures of informal reasoning skills in 2002, and 115 in 2003 (51% male, M age = 20.3 years). Of students who completed the measures of skill, 69 of the 2002 cohort, and 55 of the 2003 cohort indicated that their computer-recorded measures of practice were valid (because they did not give their password to another person, did not spend time away from the computer while logged in, etc). Of students who completed the measures of skill, 84 of the 2002 cohort, and 66 of the 2003 cohort completed the self-report measures. However, the self-report measure of the number of hours per week spent practicing was only administered to the 2003 cohort, so the number of participants who completed all measures was 51.

The subject was taught in the first semester of college (12 weeks: March-June), and almost half the students enrolled were undertaking an Arts degree (45%) with the remainder enrolled in a wide variety of degrees. The students were of high academic ability compared with the general population.

Materials

Learning materials. Deliberate practice was undertaken using the Reason!Able software plus a set of online learning materials. These comprised four graduated modules containing practice and homework exercises, 24 supplementary "lessons" covering key concepts and procedures, and six tutorials on argument mapping. Each module contained a specification of the concepts and skills to be mastered in the module, a set of practice exercises, and a set of homework exercises. Almost all exercises were completed using the Reason!Able software.

Measure of informal reasoning skills. Informal reasoning skill was measured using the A and B parallel forms of the California Critical Thinking Skills Test (CCTST) (Facione & Facione, 1992). The CCTST is a widely used, quantitative measure comprising 34 equally weighted multiple-choice questions with five subscales: analysis, evaluation, inference, deduction, and induction. The instrument was developed following a Delphi process to develop an expert consensus on the definition of critical thinking (Facione, 1990). The U.S.A. college student norms, as published in the test manual, are: mean, 15.89 out of 34, and standard deviation, 4.46 (Facione & Facione, 1992). Critical thinking and informal reasoning are not identical, but informal reasoning is a major component of critical thinking, and the CCTST is primarily a test of informal reasoning skills and subskills. Note that the CCTST questions differ considerably from the exercises in the Reason! approach, and so any gains cannot be easily explained away as the result of "teaching to the test."

Measures of practice. We obtained three objective and three self-report measures of practice. Almost all practice was completed using the Reason!Able software, enabling objective, computer-recorded details of the amount of practice each student completed. Periodically, the software would require the recorded activity data be uploaded to a server, which saved all data to a single file. Custom-written algorithms then processed this data to derive, for each student, the total time spent using the software, the total number of distinct practice sessions, and the total number of distinct actions (e.g., adding a "reason" box or editing a claim). Many factors could interfere with the recording and uploading of Reason!Able activity data. Only data whose recording and uploading was judged to be valid by the participant and researchers formed part of our analysis.

In addition, we used a questionnaire at the end of semester to gather participants' estimates of the amount of practice they had undertaken. Number of hours per week spent practicing was assessed using the question, "On average how many hours a week did you spend on this subject?" The percentage of extra (voluntary, nonhomework) activities completed was assessed using the question, "What proportion of the practice exercises did you do conscientiously?" The amount of effort expended on practice was measured by asking, "On a scale of 1 (not at all) to 5 (extremely), how hard did you work in this subject as compared with other subjects?"

Other measures. We collected additional information from the students to assess six potential moderators of the practice-gain relationship for the 2003 cohort only. To assess academic ability, we asked students to report their final high school score, and the maximum possible score. Perceived critical thinking ability, at the start of the semester, was assessed on a scale from 1 (extremely poor) to 5 (extremely good). Amount of exposure to critical thinking instruction, and to argument mapping prior to the semester was each assessed on a scale from 1 (not at all) to 5 (extremely). Amount of collaboration with other students, and amount of personal assistance from a tutor or lecturer during the semester were also each assessed on a scale from 1 (not at all) to 5 (extremely).

Procedure

The procedure was identical in 2002 and 2003. During the first week of semester, students completed the pretest of the CCTST. Approximately half completed Form A and the remainder Form B, based on the first letter of their surnames. In an attempt to ensure that students were equally motivated on both tests, 5% of their overall grade for the subject was determined by the higher of their two scores. Students in the 2003 cohort also completed a questionnaire assessing four of the six potential moderators of the practice-gain relationship (final high school grade, perceived critical thinking ability, amount of prior exposure to critical thinking instruction, and amount of prior exposure to argument mapping).

During the 12-week semester, students were provided two lectures and one tutorial (discussion section) per week. Students were assigned homework exercises, but were free to do as many of the practice exercises as they felt they needed in order to prepare themselves for the homework exercises and the tests. There was one lecturer for the class of around 150 students, and one tutor for each group of 16-19 students. Tutors gave written feedback on selected homework exercis-

TABLE 1
Sample Sizes, Means, and Standard Deviations, and Correlations Among the Measure of Gain and the Measures of Practice

	Measure (*n, M, SD*)	1	2	3	4	5	6	7
1	Gain on CCTST[a] (232, 3.58, 3.87)	—	.31*[b]	.27*[b]	.16[b]	.10[c]	.09[d]	.19*[d]
	Practice: Computer recorded							
2	Number of hours (124, 9.97, 6.84)		—	.92*[b]	.76*[b]	.22[e]	.20*[b]	.18[b]
3	Number of activities (124, 3,505, 2,577)			—	.75*[b]	.27[e]	.18*[b]	.19*[b]
4	Number of sessions (124, 34.19, 29.13)				—	.17[e]	.11[b]	.10[b]
	Practice: Self-report							
5	Number of hours per week[f] (66, 4.70, 2.85)					—	.28*[c]	.46*[c]
6	Percentage of extra activities (175, 48.21, 30.92)						—	.19*[d]
7	Amount of effort (out of 5) (175, 3.18, 0.85)							—

Note. [a] The California Critical Thinking Skills Test; [b]n = 124; [c]n = 66; [d]n = 175; [e]n = 51. [f] Data were collected from the 2003 cohort only.
*p < .05 level (two-tailed).

es, as well as direct personal guidance and feedback in tutorials and office hours. Tutorials took place in a computer-equipped classroom in which students often worked in groups of three or four. Students had use of a computer laboratory with Reason!Able installed and access to online learning materials, and they could download the software to their personal computers.

At the end of the semester, the CCTST was administered as a post-test. Students who completed Form A at pretest completed Form B at post-test and vice versa. A questionnaire was also administered, containing the three self-report practice questions (number of hours per week, percentage of extra activities, and amount of effort) and for the 2003 cohort the two remaining potential moderator questions (amount of collaboration with other students and amount of personal assistance from a tutor or lecturer during the semester). In 2002, the CCTST post-test, and the questionnaire, were both given in the final week of semester. In 2003, the questionnaire was given in the final week and the CCTST post-test was given as part of the final exam 3 weeks after the end of semester.

Results

For all analyses, the 2002 and 2003 cohorts were analyzed separately so that we could assess the replicability of results. The results for each cohort were very similar, and so for economy of presentation the datasets are combined. At each stage of data analysis, we examined the data using frequency histograms and scatter plots. In no case did we find marked skew, curvilinearity, or other unusual features.

Gain in Informal Reasoning

Standardized effect sizes for gain in critical thinking were computed by subtracting the mean pretest CCTST score (18.82) from the mean post-test score (22.40) and dividing this value by the estimated population standard deviation (4.46) derived from the CCTST manual. This value was used as a single stable estimate of the population standard deviation rather than our sample standard deviations (pretest: 4.68; post-test: 4.60) to improve our confidence interval calculations. The gain in informal reasoning skill over the 12-week semester was 0.80 (CI_{95} = 0.66, 0.94), and this was significantly greater than zero [$t_{dependent}$ (231) = 14.08, p <.001, two-tailed].

Relationships Between the Measures of Practice

Table 1 presents means, standard deviations, and correlations for the measures of practice. It reveals strong, positive relationships among the computer-recorded measures of practice, and weak to moderate positive relationships among the self-report measures. The computer-recorded measures of practice were weakly related to self-report measures, including the computer-recorded and self-reported measures of amount of time spent practicing (r = 0.22) and the computer-recorded total activities completed and self-reported percentage of extra activities completed (r = 0.18). A principal components analysis did not reveal any additional patterns.

The low correlations between the objective and self-report measures suggest that students are unable to retrospectively estimate their practice levels accurately. Therefore, future studies may also wish to utilize computer-recorded measures of practice.

The Relationship Between Practice and Gain

Table 1 also presents the correlations between the measures of practice and gain in informal reasoning, and reveals positive relationships between several measures and gain. Among the objective measures, amount of time spent practicing with the software and the number of activities with the software were related to gain, to a weak-to-moderate and statistically significant extent. Among the self-report measures, amount of effort was related to gain, to a weak-to-moderate and statistically significant extent. Other correlations were very small, and were not statistically significant, although all were positive.

A standard regression analysis determined the proportion of variance in gain accounted for by the practice variables. Twenty percent of variance in gain in informal logic skill was accounted for by the six practice variables, $R(6, 43) = 0.45$. This indicates that practice does predict gain in critical thinking to some extent; however, the sample size was small because one of the variables was measured for the 2003 cohort only.

Potential Moderators of the Relationship Between Practice and Gain

We conducted hierarchical regressions to assess potential moderators of the practice-gain relationship. The moderator measures were administered to the 2003 cohort only. The following potential beginning-of-semester moderators were assessed: academic ability (tertiary entrance score), initial perceived critical thinking ability (self-reported), prior exposure to CT instruction (self-reported), and prior exposure to argument mapping (self-reported). The following potential end-of-semester moderators were assessed: amount of collaboration with other students during the semester (self-reported) and amount of personal assistance from a tutor or lecturer during the semester (self-reported). Including each term for the interaction of a moderator with each practice variable did not improve the prediction of gain by more than 8% over prediction by practice variables alone. We thus found no evidence that the variables we examined were moderating the practice-gain relationship.

Discussion

Were the Predictions Supported?

The results support the two predictions addressed by our studies. First, participants did make substantial gains in informal reasoning skill. The average effect size, across the two cohorts, was 0.8. This is large in terms of Cohen's standard classification (Cohen, 1988). The effect corresponds to a student at the median shifting from the 50th to the 79th percentile. In *How College Affects Students*, Pascarella and Terenzini estimated that critical thinking skills grow approximately one standard deviation over an undergraduate education (Pascarella & Terenzini, 1991). More recently, taking into account research conducted in the 1990s, Pascarella has revised this estimate to about 0.5 standard deviations (Pascarella, 2003). If the true figure is somewhere in this range, participants in our study were gaining approximately as much over one semester as would normally be gained over 3-4 years of undergraduate education.

How much of the gain from pre- to post-test would have been found anyway? Some gain may result just from taking the test a second time (note, however, that we used two different forms of the test in a crossover design). More importantly, there is no doubt that students' informal reasoning skills do improve over a first-year semester, even without deliberate practice or explicit instruction. The largest study of growth of informal reasoning skills was Lehmann's (1963), which found a 0.67 standard deviation gain across all courses during the first year of college, equating to a per-semester gain of 0.34 standard deviations. A more recent series of studies of six cohorts of first-year undergraduates enrolled in a year-long combined writing and critical thinking subject found a weighted average gain of 0.55 standard deviations for the year, or per-semester gain of 0.28 standard deviations, using the CCTST (Hatcher, 1999, 2001). Based on such studies, we estimate that the students in our study would have gained approximately 0.3 standard deviations even without deliberate practice. The magnitude of the extra gain attributable to the deliberate practice is thus approximately 0.5 standard deviations.

Second, the amount of gain was positively related to amount of practice. The correlations we found were approximately 0.3 for a number of our measures of practice, in particular our objective, computer-recorded measures, and about 0.4 for the multiple regression with the full set of (related) practice variables. It is notable that very similar values were found in both years, suggesting that the relationship is quite robust.

The correlations we found between practice and gain are not as strong as we had expected. Using a quasi-experimental design, we could examine this relationship only using correlation, and only to the extent that amounts of practice did vary over students. An observed correlation could be influenced by any factor influencing both practice and gain, or moderating the postulated effect of amount of practice on amount of gain.

Recognizing that our observed correlations may be attenuated, we assessed the contribution of interactions with a number of potential moderating variables. No clear moderation could be identified. This analysis could, however, examine only some potential moderators, and given the small sample sizes and limited nature of the measures available, we cannot conclude that no attenuation occurred. Thus the relation between practice and gain may be stronger than is suggested by our observed correlations.

Another basis for thinking that the correlations in our data underestimate the actual relationship between deliberate practice and gain is that our participants' practice only approximates genuine deliberate practice. The Reason! approach was developed in order to have students engage in something as close to deliberate practice as possible under the resource constraints standard in higher education. In areas such as the amount of expert guidance and feedback, it falls short of deliberate practice as it would be experienced by elite performers in other domains. A more intensive – and expensive – training regime may well show a stronger practice-gain relationship.

Implications for Understanding Informal Reasoning

The confirmation of the predictions supports the more general deliberate practice hypothesis, the claim that high-level expertise in informal reasoning is the outcome of sufficiently large amounts of deliberate practice.

Our studies could not test the deliberate practice hypothesis in its entirety. Most obviously, they do not show that the very highest levels of informal reasoning performance – roughly corresponding to world-class competitive performance in domains such as chess, music, and sport – are obtained through deliberate practice. Although our participants had strong skills relative to the general population, none had anything like world-class skills, and the amount of practice they did during the semester falls a very long way short of the 10 years or more of dedicated training usually required to achieve such skills. Our results show that deliberate practice can make good performers better, but not that it can produce the very best.

Further, our results do not establish that deliberate practice is *necessary* for advanced expertise in informal reasoning. Some people have truly excellent informal reasoning skills: Top-level philosophers and trial attorneys may be examples. The deliberate practice hypothesis, in its strongest form, maintains that such people must have engaged in extensive deliberate practice. We conjecture that such people would have been very extensively involved in activities possessing many or most of the features of deliberate practice. This conjecture might be evaluated by studying their training and professional activities, but cannot be evaluated by studies like ours.

It might be objected that it is hardly surprising that reasoning skills improve with practice. We agree, but suggest our findings are still noteworthy for a number of reasons.

First, these studies are, to our knowledge, the first attempt to systematically investigate the acquisition of high-level expertise in informal reasoning, and they confirm that such expertise is acquired in much the same way as in other domains, despite informal reasoning being in many respects quite unlike those other domains. This outcome is surprising insofar as it conflicts with the naïve intuitive view that high-level expertise in reasoning is largely due to special talent or gifts such as high intelligence.

Second, the deliberate practice hypothesis is not the banal claim that reasoning skills improve with practice. Rather, it asserts that high-level skills result from practice of a very special sort. Activities that might be called practice but do not amount to deliberate practice, such as simply engaging in reasoning and argumentation, are predicted not to help people go beyond ordinary competence. For example, most academics routinely exercise their informal reasoning skills as a standard part of their professional lives. They might loosely be described as practicing their skills, but like social tennis players, they should not expect much change in their level of expertise.

Third, our findings demonstrate that specific sorts of practice exercises result in improved performance on the quite different tasks constituting the tests used to measure gains. In other words, the skills developed in deliberate practice *transferred* to the test items and context. Those familiar with the extensive literature on transfer of cognitive skills (e.g., Detterman, 1993) might have been pessimistic, in advance, about the prospects of such transfer. The occurrence of transfer suggests that deliberate practice was inducing gains in quite generic informal reasoning skills rather than "tricks" (procedures, rote memorization, etc.) specific to the Reason! approach.

Fourth, our findings are noteworthy because they contrast with the way we actually go about cultivating informal reasoning skills. The value of strong informal reasoning skills is universally recognized. Further, as illustrated by the way we train elite performers in so many other domains, we know what it takes to build advanced expertise. Strangely, however, we do not apply those same methods to building informal reasoning skills. With very rare exceptions, people do not, and are not required to, deliberately practice informal reasoning. The implicit assumption is that advanced informal reasoning skills are acquired not through deliberate practice but incidentally in the course of tertiary study. Our findings suggest that this unstated orthodoxy is misguided.

Fifth, the gain we observed in informal reasoning was substantial, and of practical and not just statistical significance. The estimated extra gain, over what otherwise might have been expected in one semester, was about 0.5 standard deviations, which is comparable with the gain observed in a full undergraduate education, and was achieved in less than an average of 100 hours of class time and Reason!Able work over 12 weeks. Employing the deliberate practice approach thus accelerates gains in one of the key outcomes of higher education.

Implications for Education

Our findings have practical implications for educators aiming to promote informal reasoning skills, especially at higher levels in the educational system.

First, instruction aimed at enhancing informal reasoning skills ought to based on deliberate practice. For example, standard one-semester college-level informal reasoning subjects should become less didactic and more based on a structured practice regime.

Second, educational programs or institutions hoping to cultivate high-level informal reasoning ability ought to include some explicit instruction in informal reasoning, based on deliberate practice. Tertiary institutions usually claim to promote general thinking skills, including informal reasoning. Currently, they overwhelmingly adopt an indirect approach. Only a very small proportion of students receive direct instruction (e.g., taking a one-semester reasoning subject). It is expected that most students' skills will naturally improve as they deploy their informal reasoning competence in a range of learning activities in specific academic domains. Informal reasoning is rarely practiced in its own right; it is usually practiced only in the context of engaging in some activity that requires informal reasoning (e.g., a political science essay). Students' skills do improve this way, but only slowly; much faster improvement could be achieved by targeted instruction based on deliberate practice. If this takes place early on (e.g., in the first year), subsequent informal reasoning activity may further enhance gains.

Future Directions

As a first foray into largely unexplored terrain, our research leaves open many paths for future research.

First, proper experimental studies would provide better insight into the relationship between practice and gain than our quasi-experimental design. Our studies had to contend with at least two major problems. One problem is that participants themselves determined their practice levels, opening the door for moderating variables to attenuate any relationship between practice and gain. Second, we were forced to measure (rather than specify) the amount of practice being done. Even with our objective measures, this was an imprecise business. An experimental study would control the independent variable (practice), and look at how amount of gain depends on varying and tightly specified amounts of practice. Of course, such a study would take considerable resources. These resources were not available to us at the time, but our results may help justify investing such resources in the future.

A second direction for research is to look more closely at the shape of the relationship between practice and gain. Even in a correlational design, measuring informal reasoning skills repeatedly (rather than just at the beginning and end) may show that the relationship is not a matter of linear gain with hours of practice. Such research is important for the design of educational programs intended to cultivate expertise through deliberate practice. For example, if there is an S-shaped (e.g., sigmoidal) relation between practice and gain, a cost-effective educational program should require enough deliberate practice to secure the steep gains that only appear after an initial period of slow progress.

Third, future studies should observe the outcome of much larger amounts of deliberate practice. Nothing in our results, or our informal observations, suggests that our students have reached their full potential. Indeed, we suspect that they have only just begun to scale the ladder of serious expertise in informal reasoning, and are optimistic that a training program akin to those found in other domains, such as elite sports or music, would produce levels of expertise only rarely found today.

This research was supported by a grant from the Australian Research Council. We would like to thank Andrew Bulka and Angela Donohue for assistance with the research, and the reviewers for helpful suggestions on the manuscript.

Correspondence concerning this article should be sent to Tim van Gelder, Department of Philosophy, University of Melbourne, Parkville VIC 3010 Australia (E-mail: tgelder@unimelb.edu.au).

References

Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science, 26*, 85-112.

Buckingham Shum, S. (2002). The roots of computer supported argument visualization. In P. Kirschner, S. Buckingham Shum, & C. Carr (Eds.), *Visualizing argumentation: Software tools for collaborative and educational sense-making* (pp. 3-24). London: Springer Verlag.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale NJ: Lawrence Erlbaum Associates.

Detterman, D. K. (1993). The case for the prosecution: Transfer as an epiphenomenon. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cogni-*

tion and instruction (pp. 1-24). Norwood, NJ: Ablex.

Dreyfus, H. L., & Dreyfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: The Free Press.

Ericsson, K. A., & Charness, N. (1994). Expert performance. *American Psychologist, 49*, 725-747.

Ericsson, K. A., Krampe, R. T., & Tesche-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363-406.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology, 47*, 273-305.

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus*. Millbrae, CA: California Academic Press.

Facione, P. A., & Facione, N. C. (1992). The California critical thinking skills test and manual. Millbrae, CA: California Academic Press.

Felton, M., & Kuhn, D. (2001). The development of argumentative discourse skill. *Discourse Processes, 32*, 135-153.

Graf, G. (2003). *Clueless in academe: How schooling obscures the life of the mind*. New Haven, CT: Yale University Press.

Hatcher, D. L. (1999). Why critical thinking should be combined with written instruction. *Informal Logic, 19*, 171-183.

Hatcher, D. L. (2001). Why Percy can't think: A response to Bailin. *Informal Logic, 21*(2), 171-181.

Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.

Lehmann, I. J. (1963). Changes in critical thinking, attitudes and values from freshman to senior years. *Journal of Educational Psychology, 54*, 305-315.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability and knowledge levels. *Cognition and Instruction, 14*, 139-178.

Pascarella, E. (2003). *Cognitive skills and intellectual Growth*. Unpublished manuscript.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco, CA: Jossey-Bass.

Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology, 77*, 562-571.

Perkins, D. N., Allen, R., & Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell & J. Bruner (Eds.), *Thinking: The expanding frontier* (pp. 177-189). Philadelphia, PA: The Franklin Institute Press.

Stein, N. L., & Miller, C. A. (1993). The development of memory and reasoning skill in argumentative contexts: Evaluating, explaining, and generating evidence. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 285-335). Hillsdale, NJ: Erlbaum.

van Gelder, T. J. (2001). How to improve critical thinking using educational technology. In G. Kennedy, M. Keppell, C. McNaught, & T. Petrovic (Eds.), *Meeting at the crossroads: Proceedings of the 18th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (pp. 539-548). Melbourne: Biomedical Multimedia Unit, The University of Melbourne.

van Gelder, T. J. (2002). Enhancing deliberation through computer-supported argument visualization. In P. Kirschner, S. Buckingham Shum, & C. Carr (Eds.), *Visualizing argumentation: Software tools for collaborative and educational sense-making* (pp. 97-115). London: Springer-Verlag.

van Gelder, T. J., & Bulka, A. (2000). Reason!Able (Version 1.1). Melbourne: The Reason Group (http://www.goreason.com).

Voss, J. F., & van Dyke, J. A. (2001). Argumentation in psychology: Background comments. *Discourse Processes, 32*, 89-111.

Sommaire

La plupart des gens acquièrent ce que les auteurs conviennent d'appeler une compétence en raisonnement informel par des processus standard de développement, de socialisation et d'éducation. La compétence s'entend d'un degré d'aptitude plus ou moins stable, suffisant pour la plupart des besoins courants et comparable à celui des pairs, quoique en deçà du niveau de la maîtrise. L'article est centré sur les façon dont les gens peuvent dépasser un degré de compétence courant et parvenir à des niveaux d'habileté supérieurs.

À l'heure qu'il est, le cadre qui fait autorité pour ce qui est de la compréhension du développement d'un degré élevé de compétence est celui qui a été conçu par Ericsson et ses collègues. Ils ont constaté que, dans tous les domaines auxquels ils se sont intéressés, les niveaux d'habileté les plus élevés sont le fait de volumes importants d'exercice comportant des aspects particuliers, qu'ils qualifient d'« exercice délibéré ». L'application du cadre au raisonnement informel débouche sur ce que les auteurs appellent « hypothèse de l'exercice délibéré » : à savoir un degré élevé de compétence en raisonnement informel résulte de volumes importants d'exercice délibéré du raisonnement informel.

Les auteurs ont vérifié leur hypothèse par deux études, soit des quasi-expériences effectuées dans le contexte d'un cours de raisonnement informel d'un semestre destiné aux

étudiants de premier cycle. Les participants étaient des étudiants qui faisaient usage d'une méthode d'apprentissage originale fondée sur l'exercice délibéré. Ils s'exerçaient dans un environnement informatique développé pour la circonstance dans lequel le raisonnement était représenté par des schémas d'arguments. Les progrès étaient évalués avant et après l'expérience au moyen d'un test indépendant à choix multiples, soit le California Critical Thinking Skills Test. Le volume d'exercice effectué était mesuré par le logiciel, qui recensait toutes les activités et téléchargeait les données connexes vers un serveur. Les participants ont également rempli un questionnaire à cette fin, tandis qu'un deuxième a servi à recueillir des données sur des variables supplémentaires susceptibles d'influer sur le rapport exercice-progrès.

L'hypothèse de l'exercice délibéré prédisait que les participants allaient augmenter par une grande marge leur compétence en raisonnement informel et que les progrès réalisés seraient proportionnels au volume d'exercice délibéré effectué. Les auteurs ont constaté que les participants ont évolué d'environ 0,8 écart-type. La corrélation avec le volume d'exercice était d'environ 0,3 selon les objectifs de l'exercice recensés par l'ordinateur. Aucune influence de variables modérateurs potentielles n'a été constatée.

Les résultats susmentionnés donnent à entendre que l'aptitude au raisonnement informel des participants a effectivement augmenté sensiblement. Un effet de 0,8 est important (Cohen, 1988) et comparable aux progrès réalisés normalement au cours d'un premier cycle universitaire complet. Cela dit, la corrélation entre les progrès et les volumes d'exercice était plus faible que prévu. Il se peut que le rapport ait été freiné par des variables modérateurs autres que ceux que les auteurs ont pu mesurer et écarter. Qui plus est, l'exercice effectué par les participants n'étaient pas aussi intensif que l'exercice délibéré auquel se livrent les exécutants de premier rang dans des domaines tels la musique et le sport. Il est possible qu'un rapport étroit entre l'exercice et les progrès soit constaté lorsqu'il s'agit d'exercice délibéré.

Les résultats obtenus appuient l'hypothèse de l'exercice délibéré. Toutefois, ils ne montrent pas que des volumes importants d'exercice délibéré produisent les degrés les plus élevés d'aptitude au raisonnement informel, puisque les participants n'ont effectué qu'un volume relativement modeste d'exercice au cours d'une brève période (12 semaines). Ils ne révèlent pas non plus qu'un degré de compétence élevé passe obligatoirement par des volumes importants d'exercice délibéré, quoique les auteurs présument que quiconque possède une compétence comparable se soit livré à des activités qui s'apparentent à l'exercice délibéré.

Les constatations ne font pas que suggérer que l'exercice augmente de plusieurs façons la compétence en raisonnement informel. Elles indiquent que le domaine du raisonnement informel n'est pas exceptionnel en ce qui concerne l'acquisition d'un degré élevé de compétence, que la compétence est acquise non seulement par la pratique d'une activité, mais par un exercice de type particulier, que l'exercice délibéré développe des compétences transférables authentiques, et que l'approche conventionnelle du développement de compétences en raisonnement informel, qui ne fait pas appel à l'exercice délibéré, est fautive.

Les constatations laissent entendre également que les éducateurs qui souhaitent aider les étudiants à acquérir des compétences élevées en raisonnement informel devraient leur enseigner explicitement le raisonnement informel et que leur enseignement devrait être fondé sur l'exercice délibéré.

Les études futures pourraient reposer sur des expériences bien conçues, s'attacher davantage aux rapports entre l'exercice et les progrès et viser à cerner les effets de volumes importants d'exercice délibéré. Les études effectuées par les auteurs les portent à croire que des programmes de formation comparables à ceux qui sont prévus à l'intention des exécutants de premier rang dans des domaines tels échecs, musique et sport déboucheraient sur des degrés de compétence en raisonnement informel qui ne sont que rarement observés de nos jours.