# ARE THE UNSKILLED REALLY THAT UNAWARE? AN ALTERNATIVE EXPLANATION

Marian Krajč
Andreas Ortmann

# CERGE-EI

# Are the Unskilled Really That Unaware?
# An alternative explanation

Marian Krajč
Andreas Ortmann

# Are the Unskilled Really That Unaware?
# An alternative explanation

Marian Krajč and Andreas Ortmann

CERGE-EI*

April 2007

## Abstract

In a series of articles and manuscripts (e.g., Kruger & Dunning, 1999, Dunning et al., 2003, Ehrlinger et al., 2005), Dunning, Kruger and their collaborators argued that the unskilled lack the metacognitive ability to realize their incompetence. We propose that the unskilled-and-unaware problem – rather than being one of biased judgements – is one of unbiased judgements based on biased information.

## Abstrakt

V sérií článkov a manuskriptov (napr. Kruger & Dunning, 1999, Dunning a spol., 2003, Ehrlinger a spol., 2005) Dunning, Kruger a ich kolegovia arugumentovali, že netalentovaným chýba poznávacia schopnosť nato aby si uvedomili svoju neschopnosť. My navrhujeme, že unskilled-and-unaware problém nevyplýva z vychýleného úsudku, ale môžu vzniknúť aj pri správnom úsudku na základe nepresných informácií.

---

## 1. Introduction

In a recent series of articles and manuscripts, it has been argued that the unskilled are, in addition, unaware of their incompetence: "Not only do these people reach erroneous conclusions and make unfortunate choices, but their incompetence robs them of the metacognitive ability to realize it." (Kruger & Dunning, 1999[1], p. 1121). The unskilled, thus, are allegedly afflicted by a "double curse" (Dunning, Johnson, Ehrlinger, & Kruger, 2003, p. 84). In a recent manuscript (Ehrlinger, Johnson, Banner, Kruger, & Dunning, 2005), the authors replicated their earlier results, addressing various published critiques of their work (Ackerman, Beier, & Bowen, 2002; Burson, Larrick, & Klayman, 2006; Krueger & Mueller, 2002; Krueger & Funder, 2004). By and large, the authors found support for their original contention.

The results by Dunning and Kruger are intriguing for a number of reasons. Never mind the dire implications the alleged lack of metacognitive skills have for the unskilled, the results are an interesting methodological puzzle in that they seem to be at odds with the results reported by Juslin, Winman, and Olsson (2000), who – through a meta-analysis of 130 data sets – demonstrated that over- and underconfidence disappear for general-knowledge questions that employ representative stimuli. According to these authors, people are well-calibrated.[2]

---

[1] As of April 6, 2007, the Krueger & Dunning (1999) article has attracted more than 300 references on scholar.google.

[2] Admittedly, general-knowledge questions are different from the person-oriented tasks (asking, typically, about one's capabilities, in both an absolute and relative manner) that are used to generate the unskilled-and-unaware problem. Juslin et al. (2000, p. 394) note that "one sensible but as yet untested hypothesis" is that there are important differences between these two paradigms. We are not aware of any recent studies that speak to that issue but would find it surprising indeed if there were such differences.

We argue that the subject pools used by Dunning, Kruger, and their collaborators were not distributed uniformly, or at least symmetrically, but rather skewed toward the bottom. We show below with a simple model that the unskilled, rather than being more unaware than the skilled, face a tougher inference problem which, at least partially, explains their alleged lack of metacognitive ability.

The remainder of this article is organized as follows: In the following section we provide a brief review of the literature. We then sketch out the intuition guiding our model. In section 4, we provide a simple numerical example meant to illustrate our intuition. In section 5, we complicate that numerical example and show that our alternative explanation can explain the three stylized facts of Kruger and Dunning (1999) that constitute the unskilled-and-unaware problem. In section 6, we discuss the conditions under which we expect the inference problem that the unskilled face to disappear and sketch out possible experimental tests. Section 7 concludes.

## 2. A brief review of the relevant literature

Juslin et al. (2000) demonstrated that over- and underconfidence, at least for an important experimental paradigm in psychology (general-knowledge questions), is an artifact: People tend to be calibrated reasonably well in situations that they have had a chance to experience repeatedly (Cosmides & Tooby, 1996) – general-knowledge questions almost by definition fulfilling that criterion – and that are fairly described by the stimuli materials. (Here "fair" describes whether the selected general-knowledge questions, say in city comparison tasks, reflect the ecological validity of the cues; see Hertwig & Ortmann, 2005).

In contrast, Kruger and Dunning (1999; see also Dunning et al., 2003) suggested that, across many intellectual and social domains, the subjects that perform the poorest (the unskilled) also lack the metacognition that would allow them to assess their deficiencies. This double curse of being unskilled and unaware induces, in their interpretation, the unskilled to dramatically overestimate their expertise, knowledge, skills, talents, etc.[3] The authors also suggested that the very skilled are somewhat, but less so, unaware of their skills. The ability-perception divergence is, however, much less prominent at the upper tail than at the lower tail; the authors attribute this phenomenon to "undue modesty" (Dunning et al., 2003, p. 85) which strikes us as an unpersuasive argument. There are therefore three stylized facts that beg for explanation: first, the alleged overconfidence of the unskilled; second, the alleged underconfidence of the very skilled; and, third, the asymmetry of the alleged miscalibrations.

The original findings are built on student subjects' knowledge of grammar and logical reasoning, and their self-assessment of how humorous they are; these findings have since been replicated with different tasks (e.g., Dunning et al., 2003: classroom exams) and also different subject pools (Edwards, Kellner, Sistrom, & Magyaria, 2003: clerks evaluating their performance; Haun, Zeringue, Leach, & Foley, 2000: medical lab technicians evaluating their on-the-job expertise; Parikh, McReelis, & Hodges, 2001: medical students assessing their interview skills).

A number of authors have questioned the results by Kruger and Dunning (1999). Krueger and Mueller (2002) proposed that regression to the mean (RTM) and the better-

---

[3] Below we often use these terms interchangeably.

than-average (BTA) effect could, jointly, explain the three stylized facts constituting the unskilled-and-unaware problem. RTM, to recall, is a statistical artifact that occurs when variables such as ability and the perception of ability are imperfectly correlated, possibly because of measurement errors. The imperfect correlation between ability and perception of ability, and a regression slope of less than 1 (as observed in Kruger & Dunning, 1999), imply that not all of those in the lower quartile in ability are actually in the lower quartile in perception. Thus, the expected value of the lower quartile in perception will be greater than the average ability of the lower quartile in abilities. While RTM explains the first two stylized facts, it cannot explain the third. Krueger and Mueller (2002) therefore appeal, in addition, to the BTA effect. Using the BTA effect as part of an explanation seems, however, problematic as it is the explanandum rather than an explanans.[4]

Recently, Ehrlinger et al. (2005) addressed a number of criticisms and alternative explanations of the results in Kruger and Dunning (1999) by using real-world settings and financial and social incentives. In the first part of their manuscript (study 1 and 2), Ehrlinger et al. (2005) investigated whether the performers in the bottom quartiles overestimated their relative and absolute ability after they controlled for measurement errors in real-world situations (in-class exams, debate tournaments); the authors argued that the results of these two studies confirm the original findings of Kruger and Dunning

---

[4] Krueger and Mueller (2002) also study the effects of various mediators and find that mediation has no, or lower, explanatory power than RTM and BTA together. Kruger and Dunning (2002) questioned the mediation results by Krueger and Mueller (2002), arguing they used unreliable tests and inappropriate measures of relevant mediating variables. They also point out that the results of Krueger and Mueller (2002) are true only if low or moderate levels of reliability are used and not in samples with highly reliable measures. Since this particular dispute is not of relevance to our argument, we do not pursue it in more detail here.

(1999) and undermine the RTM & BTA mean explanation of Krueger and Mueller (2002).

In the second part, the authors conducted three studies (study 3, 4, and 5) to examine whether insufficient incentives for accuracy are the reason for the overconfidence of poor performers. The participants of the third study were recruited at a Trap and Skeet competition and were asked to assess their confidence in the answers they gave to questions asked on a test of gun knowledge and safety. Participants were promised an additional $5 payment for average confidence responses within 5% of their actual score on the test.[5] Since the number of years of experience with firearm was reported to be 6-65, we can deduce that the participants of the Trap and Skeet tournament were not students. Hence this study seemed to show that the Kruger and Dunning (1999) results generalize to populations other than undergraduate students (and also other tasks.) In the fourth study, the authors conducted a similar experiment on a Logical Reasoning Test with undergraduate students who were promised an additional payment of $100 if their estimate of their performance was within 5% of their actual score on the test.[6] In the fifth study, the authors investigated the impact of social incentives (making students accountable for their self-assessment) on the results of Kruger and Dunning (1999). The "accountable" group was told that their professor would interview them regarding the rationale of their answers on a multiple choice test. The results of these three studies seemed to suggest that neither monetary nor social incentives affect the overestimation of performers in the bottom quartiles.

---

[5] However, the authors do not report the expected value of having a correct prediction, or what fraction of subjects actually earned the extra money. It is quite possible that the expected value was too low and that the financial incentives were simply insufficient (e.g., Hertwig & Ortmann, 2001).

[6] Again, the authors did not report the expected value of a hit. Even though the reward for accuracy was large nominally, the expected value might still have been low and financial incentives therefore insufficient.

In the third part of their manuscript, Ehrlinger et al. (2005) investigated the sources of inaccuracy in performance estimates. Towards that end, the authors computed via regression analysis how people weigh their estimates of their raw score and estimates of the raw score of the average person when estimating how well they performed relative to others. They then did a "what if" exercise (i.e., conducted a counterfactual regression analysis), asking what each participant's percentile ranking would be if her or his raw score (or, the average person's score) estimate were replaced with the real value. The results of this analysis suggest that the participants are inaccurate due to mistaken beliefs about their own performance, rather than due to a misconception about the performance of others.

In a related article on the unskilled-and-unaware problem, Burson, Larrick, & Klayman (2006) also questioned the results of Kruger and Dunning (1999). They conducted three studies to examine their hypothesis that task difficulty matters and suggested a noise-plus-bias model.

For the first study, students were asked to answer quizzes with either 20 hard or 20 easy questions about the University of Chicago. The students were then asked to estimate the number of correct answers they gave, their percentile rank, and the difficulty of their quiz.[7] For the easier questions, the authors replicated the results of Kruger and Dunning (1999); for the harder questions they found that both the estimates of low and high performers decreased. This result leads to a decrease in bottom performers' overestimation and an increase in top performers' underestimation. These changes are

---

[7] The authors paid only a flat participation fee in all three studies. Therefore, the usual caveats about (the lack of) performance-based financial incentive applies (e.g., Hertwig & Ortmann 2001).

such that the asymmetry (stylized fact three) found for easier questions in this study as well as in Kruger and Dunning (1999) and Ehrlinger et al. (2005) is observed for harder questions no longer; concretely, the low performers are as aware as the high performers in percentile estimates and even more aware in their score estimates. For the second study, Burson et al. (2006) varied domains (5), question sets (10), and difficulty (2) for student subjects. The participants were asked to estimate their percentile ranking and the task difficulty for themselves and for the other participants. In the third study, the student subjects were asked to create as many 4-, 5-, and 6-letter words as possible from a 10-letter word. Again, the participants were asked to estimate their percentile ranking, their number of points, and the task difficulty. The results of the second and third study support the results of the first study by Burson et al. (2006): the skilled and unskilled are similarly unaware of how they perform relative to others and the top performers are better calibrated in the easier tasks and the bottom performers in the harder tasks.

Burson et al. (2006) proposed a noise-and-bias model that is, according to the authors, sufficient to explain the observed behavior. In their model, the noise is caused by e.g. task randomness (e.g. random variation, luck, distraction, fatigue) and diagnosticity of feedback (what kind of feedback people get during the experiments) – this part of the explanation resembles the RTM argument (and the earlier work by Erev, Wallsten, & Budescu, 1994); the bias in their model captures the task difficulty – this part of the explanation resembles the BTA argument for easier questions.

Krueger and Mueller (2002) and Burson et al. (2006) agree that the reliability of measures plays an important role in the analysis. In addition, the task difficulty is the key in explaining the asymmetry in Burson et al. (2006). The relationship between task

difficulty and task reliability is, however, unclear, as Krueger and Mueller (2002) report higher reliability in easier tasks and lower in harder tasks; Burson et al. (2006) report it vice versa.

To sum up, the experiments conducted by Kruger and Dunning (1999) and Ehrlinger et al. (2005) suggest that the unskilled overestimate their absolute abilities as well as their relative abilities. The authors argue that the unskilled are overconfident about their abilities. This overconfidence is explained as resulting from a lack of metacognitive ability to realize their deficiencies. The meta-analysis of Ehrlinger et al. (2005) suggests, furthermore, that the lack of insight into participants' skills is the reason for the excessively optimistic self-assessments of poor performers. It seems widely acknowledged that noise is an indispensable part of any sensible explanation.

## 3. Our alternative explanation – the intuition

Drawing on empirical data relevant to the subject pools used, we propose an alternative explanation of the results of Kruger, Dunning, and their collaborators. Our key observation is that for almost all studies by Kruger, Dunning, and their collaborators traditional but hardly representative subjects – undergraduate (psychology) students from Cornell --  were employed, i.e., a convenience sample rather than a representative sample of the population.

Students at Cornell, and similar schools such as the University of Chicago (e.g., Burson et al, 2006), are drawn from the outer upper tail of the normal distribution of student talent. Take the example of Cornell University: According to U.S.News & World

9

Report the percentage of applicants admitted to Cornell University is 29%. Clearly, because of Cornell's reputation,[8] this is already a sample from a self-selected pool. It seems unlikely that high-school students from the lower half of the talent distribution would apply.

The talent distribution of the subject pool used in the experiments is therefore highly asymmetric and can be approximately captured by the J-distribution, which one can think of as a truncated (from below) normal distribution. This pattern can be seen in the IQ distribution, the most general measure of a person's cognitive abilities. The convex part of the upper outer tail of that distribution represents approximately the top 15% of the population. Since Cornell University accepts only 29% of the applicants and since its pool of applicants is self-selected, it seems likely that, save a couple of legacy cases, almost all Cornell students are located in the convex part of the upper outer tail of the normal distribution.

In addition, it is well-known from studies of grade inflation (Avery, Fairbanks, & Zeckhauser, 2003; Johnson, 2003; Lewis, 2006, chapter 5; Sabot & Wakeman-Linn, 1991) that grades at the undergraduate level have – with the notable exception of the natural sciences – become less and less differentiating over the years: more and more students are awarded top grades. For example, between 1965 and 2000 the number of A's awarded to Cornell students has more than doubled in percentage while the percentage of grades in the B, C, D and F ranges has consequently dropped (in 1965, 17.5 percent of grades were A's, while in 2000, 40 percent were A's).[9] This data

---

[8] According to the U.S.News & World Report web site, Cornell University is currently the 13th best university in the U.S.
[9] http://www.thehoya.com/news/041202/news7.cfm

strongly suggests that Cornell University experiences the same phenomenon of (differential) grade inflation that Harvard experiences (Lewis, 2006) and the schools discussed in Sabot & Wakeman-Linn (1991). The dramatic grade inflation documented for the humanities and social sciences devalues grades as meaningful signals specifically in cohorts of students that are newly assembled and typically draw on the top of high school classes. Inflated grades complicate the inference problem of student subjects that, quite likely, were students in their first year or in their first semester (Ortmann & Hertwig 2002).

To model the lack of feedback resulting from grade inflation and, possibly, the fact that student subjects were students in their first year or in their first semester, we introduce an error term in own-ability perception. The presence of noise in people's ability assessment has already been acknowledged in RTM explanations. Noise is likely to be correlated with familiarity (and hence feedback about one's own standing) with a particular domain. If one is not that familiar, one is likely to use one's self-assessment from other domains as a proxy, which adds to the error. We assume this error to be identical across subjects although one could reasonably argue for larger errors at the lower end. Ultimately, noise is a function indeed of task randomness as well as diagnosticity of feedback. Note that by assuming homogeneity of error, we handicap ourselves. Our results would be stronger if we were to assume skills to be being inversely related to error.

Below we show how abilities and perception of abilities drift apart as a function of both the noise (error) and the asymmetric distribution of talent. The error in our analysis is assumed to be normally distributed; our argument is robust to other specifications of the

error term. In the next section, we will illustrate the basic idea with very simple numerical examples and illustrate how our model fits the results by Kruger and Dunning (1999) and Ehrlinger et al. (2005).

The upshot of our model is that the students in the bottom quartile(s) face a tougher inference problem: it is more difficult for students in the bottom quartile(s) to estimate correctly their relative standing from the feedback that tightly clustered signals (grades) provide. By way of example, the A++ student does not face much of an inference problem, an average B student does, especially if the task pertains to domains where he or she is not likely to have had much previous feedback about their relative standing (e.g., tasks on grammar, logical reasoning, humor, and the like).

## 4. Our alternative explanation – a simple numerical illustration

We propose the following simple formula for perceived ability (y):

$y = x + \varepsilon,$

where x is the real ability (distributed according to J-distribution) and $\varepsilon$ is the error term (normally distributed). Roughly, and somewhat simplifying (but inspired by the "binning" into four skills levels in Krueger and Dunning, 1999), we represent the J-distribution of true skills as follows (1 = the worst, 4 = the best):

| Skill rank | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| # of subjects | 8 | 4 | 2 | 1 |

We represent the normal distribution of the error term as follows:

| Misclassification | -2 | -1 | 0 | +1 | +2 |
|---|---|---|---|---|---|
| Probability | 0.13 | 0.22 | 0.3 | 0.22 | 0.13 |

where 0.3 stands for the probability of a correct self-assessment for each of the skill levels and 0.22 (0.13) for the probability of under- or overestimating their ability by one (two) level(s).

Since we assume boundaries on skill categories (min.=1, max=4), for subjects in the corner categories it is possible to misclassify their ability only into one direction (e.g. someone with the lowest real skill level (1) can perceive herself as being of the same level (1), or of better level (2 or 3) only).[10] A similar logic applies to skill levels that are close to boundaries. We therefore truncate the probabilities for corner and relevant interior categories by removing impossible events (skills levels outside the range 1-4) and by normalizing the remaining probabilities to sum up to one for each category. The renormalized probabilities for the four skill levels after truncation are:

| Misclassification | -2 | -1 | 0 | +1 | +2 |
|---|---|---|---|---|---|
| Probabilities for skill rank 1 | | | 0.46 | 0.34 | 0.2 |
| Probabilities for skill rank 2 | | 0.25 | 0.35 | 0.25 | 0.15 |
| Probabilities for skill rank 3 | 0.15 | 0.25 | 0.35 | 0.25 | |
| Probabilities for skill rank 4 | 0.2 | 0.34 | 0.46 | | |

With this distribution of errors, the perceived distribution of skills is:[11]

---

[10] This argument can be justified by the selective nature of the subject pool as well as the typical distribution of grades at most colleges and universities.

[11] The mass of subjects in each skill level was computed according to the simple formula for perceived ability (y) using the probabilities truncated errors. For example, for perceived skill level 2, 4.82 = 0.35*4

| Skill rank | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Mass of subjects | 4.98 | 4.82 | 3.64 | 1.56 |
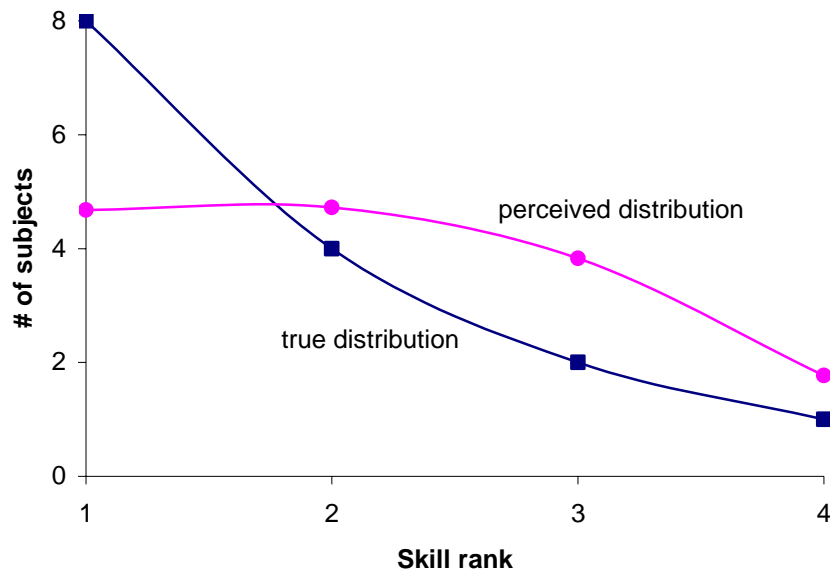


**Figure 1**

Figure 1 illustrates that as a consequence of the truncated error term, subjects perceive themselves as more skilled than they really are: the average perceived skill level increases from 2.88 for the true distribution to 3.26 for the perceived distribution. We see that this increase is driven mostly by subjects in the bottom quartiles – they perceive themselves as having better skills. This is the essence of what Kruger and Dunning (1999) and Ehrlinger et al. (2005) observed: about half of the subjects of skill rank 1 will move up (leaving four subjects – or .46 of the eight subjects in that category -- in the bottom skill rank in perception), while one quarter of the subjects of skill rank 2 (or .25 of the four subjects in that category) will move down to the bottom. Plus a fraction of a subject of skill rank 3 (or .15 of 2 subjects) will also move down to the bottom.

---

+ 0.34*8 + 0.25*2 + 0.2*1=(subjects with true skills 2 and perceived skills 2)+(true 1 and perceived 2)+(true 3 and perceived 2)+(true 4 and perceived 2).

While our simple example exhibits a similar pattern in the bottom quartiles as the results of Kruger and Dunning (1999), it does not capture behavior in the top quartile. As we will see presently, this is an artifact resulting from the low number of skill levels considered.

## 5. Our alternative explanation – the example generalized to more skill levels

In this section we generalize the simple example to more levels of abilities. To generate a J-shaped distribution of real abilities we use one of the J-shape distributions – the Chi-square distribution.[12] The error term is, as in our illustration, generated by the normal distribution and it is truncated where necessary (at the edges of the abilities range). Perceived ability is then computed as real ability plus (renormalized) error in one's self-assessment.

Perceived as well as true distribution of skills is shown in the Figure 2.
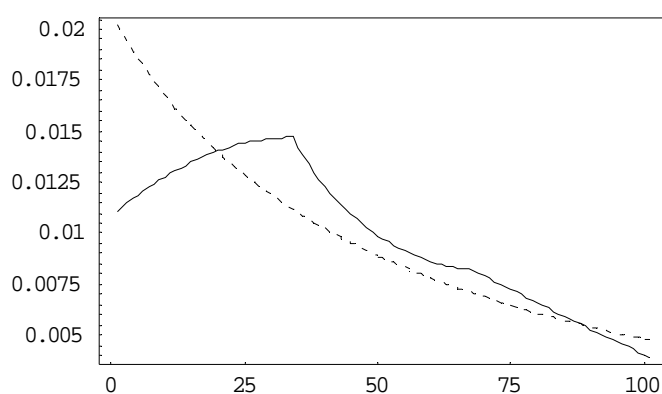


**Figure 2**

---

[12] We are using the Chi-square distribution as a proxy to the J-distribution because the Chi-square distribution is, unlike the J-distribution, parameterized. Another possibility is to use the Pareto distribution, which also has the desired shape. The results obtained using the Pareto distribution are very similar to the results we get using the Chi-square distribution.

On the x-axis abilities range from 1 (the lowest) to 101 (the highest). The y-axis corresponds to the number of subjects (where total mass of subjects is 1). The dashed line is the pdf of the real abilities and the solid line is the pdf of the perceived abilities constructed as described. We again observe a significant shift of the mass of subjects with lower abilities (dashed line) towards higher perceived abilities (solid line). We also, however, observe a shift of the mass of subjects with very high abilities towards lower perceived abilities. This captures the second stylized fact constituting the unskilled-and-unaware problem. The third stylized fact also emerges as the misperception of the unskilled is much larger than the misperception of the very skilled.

The distribution of perceived abilities is, of course, dependent on the choice of the key parameters – standard deviation and for how many ability categories we allow people to make errors (truncation). We have conducted various robustness tests and find that our basic result is robust to variations in the standard deviations and truncation of errors.[13]

Our model suggests what is intuitively clear; the less error we allow (i.e., the less complicated the inference problem that subjects face), the more accurate predictions people make and the less likely they are liable to fall victim to the unskilled-and-unaware problem. Note that our results are consistent with the results of the meta-analysis in Ehrlinger et al. (2005), where the authors show that correcting the errors in own perceived ability helps people to assess their percentile ranking more accurately.

## 6. Discussion

---

[13] These computations are available from the corresponding author.

The following table summarizes subject pools, financial incentives, and real-world stimuli in the two sets of studies that have motivated our inquiry (Kruger & Dunning, 1999; Ehrlinger et al., 2005).

| Study | Real-world stimuli? | Financial incentives? | Subject pool |
|---|---|---|---|
| K&D(1999)[14] – Study 1 | (humor) | No (extra credit) | CU undergraduates* |
| K&D(1999) – Study 2 | (logical reasoning) | No (extra credit) | CU undergraduates* |
| K&D(1999) – Study 3 | (English grammar) | Yes ($5 or extra credit) | CU undergraduates |
| K&D(1999) – Study 4 | (logical reasoning) | No (extra credit) | CU undergraduates |
| E(2005)[15] – Study 1 | (perf. on in-class exam) | No (extra credit) | CU undergraduates* |
| E(2005) – Study 2 | (debate tournament) | No | Students participating in a debate tournament |
| E(2005) – Study 3 | (Trap and Skeet) | Yes | ? |
| E(2005) – Study 4 | (logical reasoning) | Yes | CU undergraduates* |
| E(2005) – Study 5 | (logical reasoning) | No (extra credit) | CU undergraduates* |

* all psychology students

We have argued that the asymmetric distribution of talent we are likely to find in particular samples such as those drawn from elite colleges and universities like Cornell or Chicago is a key determinant of the experimental findings that constitute the

---

[14] Kruger and Dunning (1999)
[15] Ehrlinger et al. (2005)

unskilled-and-unaware problem. Indeed, with one exception,[16] all studies reported in Kruger and Dunning (1999) and Ehrlinger et al. (2005) featured student subjects.

What about participants like those in the Trap and Skeet competition (and other subject pools used in related studies)? There exists no absolute or relative measure in Trap and Skeet as exists in chess (ELO number) or in golf (handicap) where people therefore can assess reasonably well their absolute as well as relative position in the players' pool. In our view it is exactly this lack of information/feedback about one's own position that makes it difficult for student subjects as well as subjects in Trap and Skeet to make the appropriate inferences about where they stand in the grand scheme of things. We conjecture that Trap and Skeet competitions draw, in addition, on asymmetrically distributed populations and that the participants – like the student participants – did not have the kind of repeated feedback that could substitute for ELO numbers or golf handicaps (which of course require a series of competition to emerge/be computed). We suspect that these arguments also apply to those other populations and tasks that have been used in the literature: medical students assessing their interview skills; medical lab technicians evaluating their on-the-job expertise; clerks evaluating their performance.

Our explanation suggests that the alleged unskilled-and-unaware problem is to some extent an experimental artifact but for different reasons than conjectured in the literature

---

[16] Ehrlinger et al. (2005) recruited participants at a Trap and Skeet competition in some nearby club; these participants had 6-65 years of experience with firearms (mean=34.5), 96% owed at least 1 firearm and 89% had taken a course in firearm safety. We have no information about the actual distribution of skill in this study and we do not know how much the subjects in this competition could reasonably make inferences about their absolute and relative skills on a test that seems not have been related to the task they were in the process of performing.

and reviewed in the literature review. Unskilled students do not necessarily[17] lack metacognitive ability any more than more skilled students; they simply face a tougher inference or signal extraction problem.

Our suggestion fits well in an established body of literature that suggests that people, when being allowed and able to learn, will do so (e.g., Koehler, 1996; see also Kruger & Funder, 2004). Take a cause celebre to both economists and psychologists: Chu and Chu (1990) and Cox and Grether (1996) have shown experimentally that it takes only a couple of rounds of repetition coupled with financial incentives and feedback for the preference reversal phenomenon to be driven out. Needless to say that the precision (diagnosticity) of the feedback, as well as the financial and social incentives, will affect subjects' ability to learn.

Our explanation suggests a couple of (experimental) tests. First, an experimenter might want to ask early in the semester and late in the semester students in a class about their absolute and relative performance: It seems highly likely that we will see a relatively quick sorting of subjects in such a situation (although it will be moderated by the quantity and quality of feedback (e.g., two midterms, an informative distribution of grades, etc.) even when subjects start out with homegrown, and misleading, priors because they come from varied backgrounds. Second, we conjecture that we will see a significant difference in the self-assessments of first-semester/year humanities or social-science students on the one hand and last-semester/year natural sciences majors even at elite schools like Cornell and Chicago on the other hand. This is because of the time they have to assert their absolute and relative performance as well as the differential

---

[17] They might but this is a topic of current research.

diagnosticity of feedback that the well-documented differential grade inflation at US colleges and universities implies.

## 7. Concluding remarks

Dunning, Kruger, and their collaborators have proposed that the unskilled suffer from the "double curse" of being unskilled and being afflicted by a lack of metacognitive ability to realize their deficiencies. We have provided an alternative explanation that argues that the unskilled face a much tougher inference or signal extraction problem. In other words, we suggest that flawed self-assessments do not necessarily result from biased judgements but can be explained as unbiased judgements based on biased information.

Our results seem of importance to the unskilled and unaware as well as those that try to understand, and remedy, the situation. Our results also make the more fundamental methodological point that little can be said about (the lack of) metacognitive ability if one does not control for the distribution of real abilities, task randomness, the diagnosticity of feedback, and real financial and other incentives.

This insight applies to various other contexts.

## References

Ackerman L.P., Beier E.M., & Bowen R.K. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences, 33,* 587-605.

Avery, Ch., Fairbanks A., & Zeckhauser R. (2003). *The Early Admissions Game: Joining the Elite.* Harvard University Press.

Burson A.K., Larrick P.R., & Klayman J. (2006). Skilled or Unskilled, but Still Unaware of It: How Perceptions of Difficulty Drive Miscalibration in Relative Comparisons. *Journal of Personality and Social Psychology, 90,* 60-77.

Chu, Y., & Chu, R. (1990). The Subsidence of Preference Reversals in Simplified and Marketlike Experimental Settings: A Note. *The American Economic Review, 80,* 902-911.

Cosmides L., & Tooby J. (1996). Are Humans Good Intuitive Statisticians after all? Rethinking some Conclusions from the Literature on Judgment under Uncertainty. *Cognition, 58,* 1-73.

Cox, C.J., & Grether, M.D. (1996). The preference reversal phenomenon: Response mode, markets and incentives. *Economic Theory, 7,* 381-405.

Dunning D., Johnson K., Ehrlinger J., & Kruger J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12,* 83-87.

Edwards K.R., Kellner R.K., Sistrom L.Ch., & Magyaria J.E. (2003). Medical student self-assessment of performance on an obstetrics and gynecology clerkship. *American Journal of Obstetrics and Gynecology, 188,* 1078-1082.

Ehrlinger J., Johnson K., Banner M., Kruger J., & Dunning D. (2005). Why the Unskilled are Unaware: Further Exploration of (Absent) Self-Insight Among the Incompetent. Manuscript Under Review.

Erev, I., Wallsten, T.S., & Budescu, D.V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101,* 519-527.

Haun, D.E., Zeringue, A., Leach, A., & Foley, A. (2000). Assessing the Competence of Specimen-Processing Personnel. *Laboratory Medicine, 31,* 633–637.

Hertwig, R., & Ortmann, A. (2001). Experimental Practices in Economics: A Challenge for Psychologists? *Behavioral and Brain Sciences, 24,* 383-403.

Hertwig, R., & Ortmann, A. (2005). The Cognitive Illusions Controversy: A Methodological Debate in Disguise That Matters To Economists. In R. Zwick and A. Rapoport (eds.), *Experimental Business Research* (pp. 361-378). Boston, MA: Kluwer.

Johnson, E.V. (2003). *Grade Inflation: A Crisis in College Education.* New York: Springer.

Juslin P., Winman A., & Olsson H. (2000). Naïve Empiricism and Dogmatism in Confidence Research: A Critical Examination of the Hard-Easy Effect. *Psychological Review, 107,* 384-396.

Koehler, J.J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological changes. *Behavioral and Brain Sciences, 19,* 1-53.

Kruger J., & Dunning D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own incompetence Lead to Inflated Self-Assessment. *Journal of Personality and Social Psychology, 77,* 1121-1134.

Kruger J., & Dunning D. (2002). Unskilled and unware – but why? A reply to Krueger and Mueller. *Journal of Personality and Social Psychology, 82,* 189-192.

Krueger I.J., & Funder C.D. (2004). Towards a Balanced Social Psychology: Causes, Consequences and Cures for the Problem-seeking Approach to Social Behavior and Cognition. *Behavioral and Brain Sciences*, *27*, 313-376.

Krueger I.J., & Mueller A.R. (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,* 180-188.

Lewis, R.H. (2006). *Excellence Without a Soul: How a Great University Forgot Education*. PublicAffairs.

Ortmann, A., & Hertwig, R. (2002). The Costs of Deception: Evidence From Psychology. *Experimental Economics, 5,* 111-131.

Parikh A., McReelis K., & Hodges B. (2001). Student feedback in problem based learning: a survey of 103 final year students across five Ontario medical schools. *Medical Education, 35,* 632-636.

Sabot R., & Wakeman-Linn J. (1991). Grade Inflation and Course Choice. *The Journal of Economic Perspectives, 5,* 159-170.