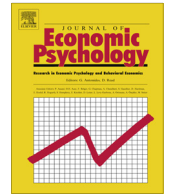Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

How unaware are the unskilled? Empirical tests of the "signal extraction" counterexplanation for the Dunning–Kruger effect in self-evaluation of performance

Thomas Schlösser [a], David Dunning [b,*], Kerri L. Johnson [c], Justin Kruger [d]

[a] University of Cologne, Department of Sociology and Social Psychology, Albertus-Magnus-Platz, 50923 Cologne, Germany
[b] Cornell University, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853, USA
[c] University of California, Los Angeles, 2330 Rolfe Hall, Los Angeles, CA 90095, USA
[d] New York University, Stern School of Business, 40 West 4th Street, Suite 804, New York, NY 10012, USA

ARTICLE INFO

ABSTRACT

Previous work on the Dunning–Kruger effect has shown that poor performers often show little insight into the shortcomings in their performance, presumably because they suffer a double curse. Deficits in their knowledge prevent them from both producing correct responses and recognizing that the responses they produce are inferior to those produced by others. Krajč and Ortmann (2008) offered a different account, suggesting instead that poor performers make performance estimates with no more error than top performers. Floor effects, coupled with the assumption of a backwards-J performance distribution, force their self-evaluations errors to be frequently positive in nature. Krajč and Ortmann, however, offered no empirical data to test their "signal extraction" account. In three studies, we assessed their theoretical model by examining whether (1) the data producing the Dunning–Kruger effect fit the statistical assumptions considered by Krajč and Ortmann necessary to produce it, and (2) to see if their framework reproduced Dunning–Kruger errors in a data set that fit their statistical assumptions. We found that the Krajč–Ortmann framework failed to anticipate self-evaluative misperceptions on the part of poor performers, but that it does much better at accounting for misperceptions among top performers. Paradoxically, the model suggests that Kruger and Dunning (1999) may have underestimated the accuracy of top performers, even though their account asserts such accuracy.

© 2013 Elsevier B.V. All rights reserved.

* Corresponding author. Tel.: +1 607 255 6391; fax: +1 607 255 8433.
E-mail addresses: t.schloesser@uni-koeln.de (T. Schlösser), dad6@cornell.edu (D. Dunning), kerri.johnson@ucla.edu (K.L. Johnson), jkruger@stern.nyu.edu (J. Kruger).

1. Introduction

Ever since Akerlof's (1970) groundbreaking paper, economics and other behavioral theorists have worked to delineate the implications of information asymmetry for economic and social behavior. In situations of information asymmetry, people differ in the amount of information they know. A used car salesman knows whether the car being sold is a "lemon," whereas a buyer is not necessarily in a position to know. Thus, to understand such situations, one must understand the circumstances surrounding the person in the "low information" position.

One key question a theorist can ask is whether people know when they are that low information person. One of the most documented biases in self-judgment is the tendency for people to overrate their skill, expertise, and performance. People provide overinflated views of themselves in a variety of settings (for reviews, see Dunning, 2005; Dunning, Heath, & Suls, 2004). Students overpredict how well they will do in upcoming course exams (Helzer & Dunning, 2012). Competitors engaged in debate or chess tournaments overestimate how well they will place (Ehrlinger, Johnson, Banner, Dunning, & Kruger, 2008; Park & Santos-Pinto, 2010). In two software development firms, 32% and 42% of engineers placed their skill among the top 5% of their colleagues at the firm (Zenger, 1992).

1.1. The Dunning–Kruger framework

One analysis of this pervasive overconfidence credits it to the Dunning–Kruger effect (Dunning, 2011; Kruger & Dunning, 1999). According to the effect, not everyone overrates his or her ability and performance. Rather, it is exactly low information individuals—that is, the incompetent—who supply much of the overestimation. They do so because they suffer a dual curse. First, gaps and corruptions they suffer in their expertise lead them to make many mistakes and errors. Second, and perhaps equally important, these exact same gaps and corruptions leave them unable to recognize that they are making errors, and to appreciate that the responses of their peers are more appropriate and wiser than their own. Because poor performers make what they believe are the most reasonable choices available, they think they are doing just fine when their actual performance is anything but.

In effect, the Dunning–Kruger framework asserts that the expertise needed to judge performance in many intellectual and social skill domains is exactly the same expertise necessary to produce good performance in the first place. Thus, those failing to achieve good performance are also those the least able to judge when it has been attained or avoided—and they will fail to recognize just how incompetent their performances are. More than that, because of their imperfect expertise, they are simply not in a position to recognize the depths of their deficiencies.

Extant data supports this dual-curse explanation of mistaken self-evaluation of performance. Among people asked to judge their performances on a test or a task just after finishing it, poor performers (taken as the worst 25% of performers) dramatically overestimate how well they have done, in both absolute terms and in comparisons with their peers (Dunning, Johnson, Ehrlinger, & Kruger, 2003; Ehrlinger et al., 2008; Kruger & Dunning, 1999; Sheldon, Dunning, & Ames, in press; Williams, Dunning, & Kruger, 2013). Poor performers are worse than better performers at judging quality of performance and often think they have outperformed a majority of their peers when, in fact, their performance puts them among the bottom 25% of performers.

Such a pattern of gross self-overestimation extends to real world settings, such as students taking classroom exams (Dunning, et al., 2003; Ehrlinger et al., 2008; Ferraro, 2010), competitors engaged in debate tournaments (Ehrlinger et al., 2008), lab technicians quizzed about everyday work tasks and knowledge (Haun, Zeringue, Leach, & Foley, 2000), players at chess tournaments (Park & Santos-Pinto, 2010), members of a bridge club (Simons, 2013), drivers failing their initial driving test in the Netherlands & Finland (Mynttinsen et al., 2009), and medical students practicing an intake interview (Hodges, Regehr, & Martin, 2001).

Consistent with the dual-curse account, poor performers also do worse than their more competent counterparts when explicitly asked to judge performances, that is, to separate superior from poor performance, whether those performances be their own or someone else's (Kruger & Dunning, 1999; see Dunning, 2005, 2011). Most important, perhaps, teaching poor performers the skill they need to solve logic problems causes them to see previous errors that they have made, and thus paradoxically prompts them to rate their logical reasoning skills as worse than they believed before (Kruger & Dunning, 1999). Apparently, it takes skill to see lack of skill, and thus people will not see their own deficits until those deficits have been eradicated.

Further, the Dunning–Kruger framework makes predictions about biased self-judgment among top performers. According to the framework, top performers judge their decisions accurately, but fall prey to a false consensus effect (Ross, Greene, & House, 1977), overestimating just how well other people perform on the same tasks (Kruger & Dunning, 1999). As such, they roughly know how well they are doing in an absolute sense, but they underestimate how special or unique their performance is among their peers, thus underestimating how well they are doing comparatively to others. In short, whereas errors among poor performers involve misevaluations of the self, errors among top performers involve more misestimates about other people. Data support this analysis of underestimation among the very competent (Ehrlinger et al., 2008; Kruger & Dunning, Study 4; Hodges et al., 2001).

1.2. Alternative accounts

Despite the evidence above, several researchers have proposed alternative accounts for the Dunning–Kruger effect that do not involve a dual curse. For example, Krueger and Mueller (2002; see also Ackerman, Beier, & Bowen, 2002; Burson, Larrick,

& Klayman, 2006) have asserted that regression to the mean, coupled with the above average effect, would produce the basic relationship between objective performance and self-perception attributed to the Dunning–Kruger effect. That is, no two variables are ever perfectly correlated, including the perception and the reality of performance, simply because each is measured with some level of error that degrades the correlation between them. Thus, it is not a surprise that people perceive their performance as too high when it is in reality very low and their performance too low when it is, in fact, very high. These alternative accounts have inspired vigorous theoretical replies and empirical refutations (see Dunning, 2011; Ehrlinger et al., 2008; Kruger & Dunning, 2002).

1.3. The "signal extraction" account

Krajč and Ortmann (2008) propose a different alternative account for the pattern of self-evaluation errors observed in the Dunning–Kruger studies. These researchers reject the dual-curse account, specifically the second half of the curse that poor performers are worse at judging good versus bad performance. Instead, they presume that high and low performers have roughly the same level of skill at assessing performance quality. Thus, people have a general idea of where their skill levels lay, but, of course, there is some random error in their judgments.

Where high and low performers differ is the external situation that makes inferring that skill level vis-à-vis others more difficult for low than for high performers. Krajč and Ortmann (2008) assume a backwards *J*-distribution (similar to a chi-square distribution) for skill level, with many people piled up at the bottom rungs of skill level and a distribution from that pile that narrows down to a thin tail of high performers. In fact, they specifically asserted that the original data from Kruger and Dunning (1999) was likely characterized by a backward-*J* distribution:

"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. ... Clearly, because of Cornell's reputation, 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 typically 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." (Krajč & Ortmann, 2008, p. 729).

According to Krajč and Ortmann (2008) this piling on the bottom of the performance distribution based on Cornell's selectivity makes extracting the signal of ability level, and placing it in the proper point in the social distribution of performances, more difficult for low performers than for their high performing counterparts. First, low performers are at the floor of performance, and so any errors they make about their skill level will be shifted upward. Being at the bottom, they cannot underestimate themselves, so the only possible errors that can arise are those in a positive direction. The same is true for very top performers, but in the opposite direction. They cannot overestimate themselves, so any errors will show a negative bias. However, there are far fewer top performers near the ceiling of performance level to cancel out what is happening at the low end. Thus, the major type of error that will be seen by an observer is for low performers to generally overrate their performance.

Second, because there are so many people at the bottom of the performance dimension, any error or shift upward in skill estimate will be displaced more in social comparison terms. That is, if a person at the low end misjudges his or her ability in a favorable direction, it will mean a bigger shift in percentile estimate (i.e., so many more people will now be assumed to be outperformed) than would be the case with an unfavorable error made by a top performer at the high end. At the rarified and sparsely populated high end, an error in an unfavorable direction will not change much how many people the top performer outperforms. Thus, the overestimation seen among poor performers (in percentile terms) will be greater in magnitude than any countervailing shift in the opposite direction among top performers.

There are many critiques one can make to the Krajč and Ortmann assumptions about the performance distributions underlying the data in the Kruger and Dunning (1999) paradigm. The first is that the Dunning–Kruger pattern of responses has been replicated in far less selective environments presumably capturing a far wider range of performances, such as competitors at a skeet-shooting contest or a debate tournament (Ehrlinger et al., 2008), those seeking pharmacy licenses in Canada (Austin, Gregory, & Galli, 2008), lab techs working in hospitals (Haun et al., 2000), and people seeking their first driver's license (Mynttinsen et al., 2009). Kruger and Dunning have also tested their hypotheses along a non-academic skill, sense of humor, that they can assure any reader that elite college students are not selected for.

In addition, another worry for the Krajč–Ortmann model is that the Dunning–Kruger effect arises even for estimates that involve no comparison with others—that is, where there is no comparative signal to extract. When respondents are asked to estimate their raw score on a test, poor performers tend to grossly overestimate how well they do, whereas top performers are largely unbiased in their estimates (Dunning et al., 2003; Ehrlinger et al., 2008).

More important, however, is that Krajč and Ortmann's (2008) assertion of a backwards-*J* performance distribution among elite American college students represents a oversimplified view of how such students are selected for admittance. If students were accepted solely on the basis of one skill, such as performance on a single IQ test, then one would see a *J*-distribution for skill along that one dimension. Students at elite institutions, however, are chosen along many criteria that

only moderately correlate with one another, if at all. They are chosen on the basis of verbal and math skills, grades in classes, and participation in extra-curricular activities. They are also chosen for special talents, such as musical ability or athletic skill. Depending on the school, 10–30% of Ivy League students are "legacies," and are chosen, in part, because their family has a history or relationship with the university (The Economist, 2004; Worland, 2011). Adding to this are selections used to create ethnic, social, or geographic diversity.

All these considerations guarantee that on any one specific skill (e.g., math ability), there were be a negative tail as well as a positive one in the performance distribution. To be sure, the mean of this more normal distribution may lay higher than it does in the general population, but there will be likely no "piling up" of people performing at the bottom of the distribution.

1.4. Overview of present studies

With their assumptions and observations in place, Krajč and Ortmann (2008) formally propose that the self-estimates of skill among low performers would be systematically shifted upward toward overestimation, whereas the estimates of top performers would display just a minor shift toward underestimation—the pattern typically associated with the Dunning–Kruger effect. However, they have not put their ideas to empirical test, and so that is what we set out to do. We examined three data sets with differing characteristics and gauged how well the Krajč–Ortmann signal extraction account explained the pattern of self-evaluation biases of high and low performers observed in those data sets.

The first data set (Kruger & Dunning, Study 4) was used because participants could sort themselves into only eleven different levels of performance, and Krajč and Ortmann (2008) suggested that their analysis should be more relevant to the extent that performance levels were few. It is also one of the data sets that Krajč and Ortmann asserted must be backwards-*J* shaped. The second data set comes from an unpublished study of a class of nearly 400 students judging how well they were doing on a course exam where 23 different levels of performance were realized. This data set is ecologically valid (it came from a university class actually being taught), and we can look to see if the performances from this Ivy League student data set fit the assumptions of the signal extraction account.

This second data set also allowed us to examine another assertion made by Krajč and Ortmann (2008; see also Ryvkin, Krajč, & Ortmann, 2012)—that people naturally, over time, would learn their true skill level and show less bias as a result. A similar study has suggested that the Dunning–Kruger effect evaporates with experience, but that feedback and incentives are needed in combination to produce better estimates (Miller & Geraci, 2011). However, there are also extant data suggesting that poor performers fail to provide more accurate self-assessments even when given experience and feedback about their deficits (Ferraro, 2010; Hacker, Bol, Horgan, & Rakow, 2000). Examining our data set allowed us to provide another test of whether poor performers become more accurate with experience.

The third data set is an unpublished pilot study we conducted focusing, on performance along Wason selection tasks. In contrast to Study 1, these data came from a wide sample of performances collected on the Mechanical Turk crowd sourcing facility provided by Amazon.com. The key to this data set is that it fit the basic statistical assumption Krajč and Ortmann (2008) asserted was necessary for the Dunning–Kruger effect to emerge, namely, a backwards-*J* performance distribution with many poor performers but only a few top performers.

Thus, using these three data sets, we then asked the following questions. The first was whether our data sets fit the statistical assumptions that the Krajč and Ortmann (2008) said were necessary to produce the Dunning–Kruger effect. For example, in data sets that produce the Dunning–Kruger pattern, is objective performance as positively-skewed as they assumed?

Second, do high and low performers (once one removes observed biases in their judgments) show an equivalent capacity to judge their own performances? That is, do they show equal levels of random error in the self-evaluations they make—demonstrating an equivalent ability to extract a signal out of noise—or do low performers show less of an ability (e.g., more random error in their judgments)?

The third question was what participants' performance estimates would look like if we took their actual performances and then predicted what their estimates should look like according to the Krajč–Ortmann model. By applying their model, would we recapitulate the pattern of responses typically seen in a Dunning–Kruger study, with low performers dramatically overestimating how well they had actually done?

Finally, with the second data set, in which participants estimated their performance on two exams, we could see if students showed any evidence of learning to provide more accurate self-assessments over repeated tasks.

2. Method

In all, we examined our questions within three data sets.

2.1. Study 1: Reanalysis of Kruger and Dunning (1999, Study 4)

2.1.1. Participants

Participants were 140 Cornell University students in psychology and human development courses who received extra credit toward their course grade for participating.

2.1.2. Procedure

A description of the entire study may be found in the original Kruger and Dunning (1999) report. Herein, we focus only on the pre-test phase of that study, in which students confronted a 10-item Wason selection task test (Wason, 1966). In Wason tasks, participants are presented four cards and told there is information on both sides of the cards. For example, if students see cards marked "A", "B", "3," and "6", they are told that each card has a letter on one side and a number on the other. They are then told that should specify which cards have to be turned over to determine whether the following logical rule is being followed: *If there is a vowel on one side of the card, then there is an even number on the other*. Participants are told to indicate the minimal set of cards that must be turned over to verify that the rule is being followed. (Here, the correct answer is A and 3.)

After completing all 10 items, participants were asked to evaluate their performance. They were asked to judge, in percentile terms, both their logical reasoning ability in general among Cornell students, as well as how well they had done on the quiz specifically. Participants were reminded that a percentile scale focused on the percentage of Cornell students involved in the experiment that participants believed they had outperformed. For example, a self-rating of 60 meant that participants felt they outperformed 60% of their peers when it comes to logical reasoning, in general. In terms of quiz performance, a rating of 25%, for example, meant that participants felt they had outscored only 25% of their peers on this particular 10-item test.

After providing percentile estimates, participants next estimate how many out of the 10 items they had answered correctly. They then moved onto the second phase of the study.

2.2. Study 2: Classroom exam

2.2.1. Participants

Participants were 414 students enrolled in an intermediate-level, large lecture class in psychology.

2.2.2. Procedure

The course presented students with two preliminary (i.e., mid-term) examinations. Examinations were in-class, during a regular 75-min class session. The first exam took place roughly five weeks after the start of the course; the second, five weeks after that. Each test contained a total of 45 points.

At the end of each test, students were asked to rate, in percentile terms, how well they had done on the test relative to other students in the room. They were also asked to estimate their raw score (out of 45 points) on the test. There was one variation of raw score estimation across tests. For the first exam, participants estimated their score as a percentage of points possible (from 0% to 100%), after indicating that a perfect score (100%) would mean gaining 45 points. We then converted participants' percentages to a raw score. On the second exam, participants directly estimated the number of points (out of 45) they thought they had attained.

After the tests were graded, a teaching assistant for the course added participants' actual raw test scores to our dataset. Of the 414 students enrolled, we had complete data (self-estimates plus objective performance on the two tests) for 394 when examining responses to the first exam and 344 students when examining responses to both.

2.3. Study 3: Wason selection redux

2.3.1. Participants

Participants were 103 individuals recruited via Amazon.com's Mechanical Turk crowd-sourcing facility. Each participant was paid US$0.50 for participating. Respondents were constrained to those accessing the experiment from a United States IP address.

2.3.2. Procedure

Participants completed a 10-item Wason selection task quiz similar to that used in Study 1. At the end of the quiz, they estimated the number of items they had responded to correctly. They also evaluated their "logical reasoning ability" and "specific test performance" in percentile terms, relative to "other participants taking part in the experiment."

3. Results

3.1. Evidence for the Dunning–Kruger effect

Before assessing the success of the Krajč–Ortmann model (Krajč & Ortmann, 2008) to account for the Dunning–Kruger effect, we first established that participants tended to overestimate themselves, and that these misestimates fit the Dunning–Kruger pattern. Table 1 shows average perceived versus actual performance that participants posted across the three studies. For brevity, we include data from only the first exam that students completed in Study 2, although data from the second exam produced (as we shall see) virtually identical results. Of the 6 self-evaluation measures examined, participants

Table 1
Average perceived versus average actual performance.

Study/self-assessment measure	Perceived average	Actual average	p
Study 1			
Ability percentile	63.8	50.0	<.0001
Test score percentile	60.6	50.0	<.0001
Raw score	6.6	5.0	<.0001
Study 2 (exam 1)			
Test score percentile	69.9	49.9	<.0001
Raw score	37.2	37.2	ns
Study 3			
Ability percentile	68.6	50.0	<.0001
Test score percentile	65.4	50.0	<.0001
Raw score	7.2	3.0	<.0001

significantly overestimated their performance on 5 ($ps < .0001$). Participants failed to overestimate their performance only on estimates of their raw exam score in Study 2.

Fig. 1–3 examine the relationship between self-perceptions versus reality of performance for our three data sets. In Fig. 1, focusing on data from Study 1, we split participants into four groups, as close as we could of equal size, based on their objective performance, from bottom to top performance quartiles. Fig. 2 does the same for the first exam that students faced in Study 2. Here, to prevent sizes of groups that were lop-sided, we split students into five rather than four groups. In Fig. 3, concentrating again on the Wason task in Study 3, we return to splitting people into four groups.
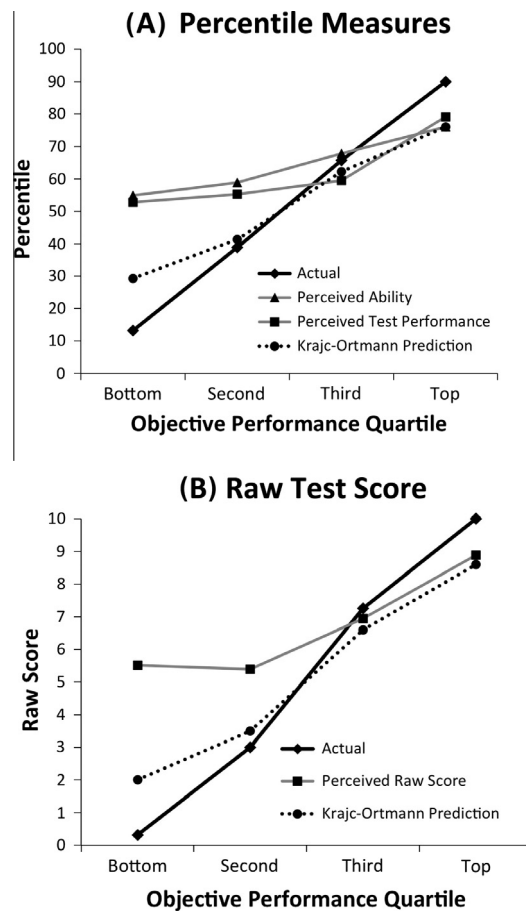

Fig. 1. Actual versus perceived performance (Study 1; Wason task). Dotted line represents where perceived performance should lay according to the Krajč–Ortmann model.
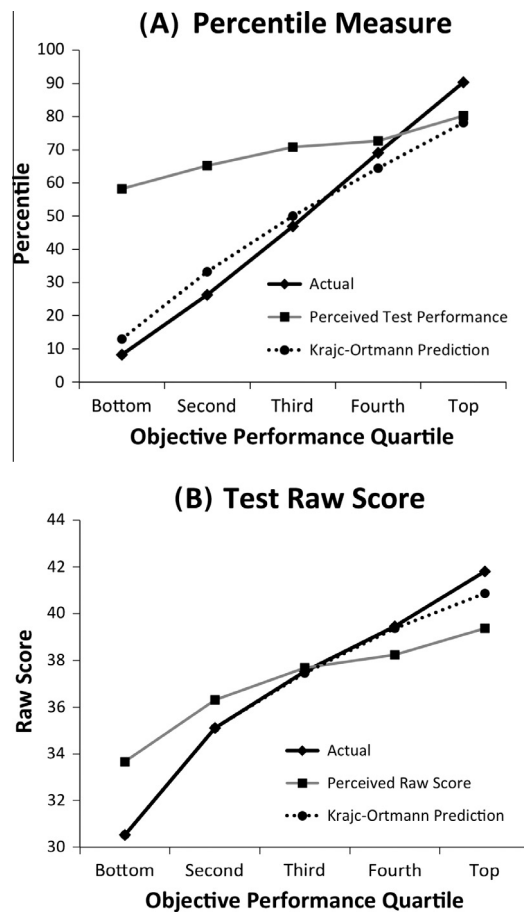
Fig. 2. Actual versus perceived performance (Study 2; classroom exam). Dotted line represents where perceived performance should lay according to the Krajč–Ortmann model.

The Figs. 1–3 and Table 2 display two obvious trends. First, perceived performance correlated significantly with actual performance for all measures. This was revealed in Tobit regression analyses in which we predicted perceived performance from its cognate in actual performance (i.e., perceived percentiles from all actual percentiles, perceived raw score from actual raw score), all bs > .22, ps < .005 (Table 2, Model 1). We used a Tobit regression rather than a regular ordinary least squares regression in order to avoid any bias due to floor or ceiling effects. That is, perceptions of performance were bounded at the bottom by 0 and at the top by 100 in the case of percentile estimates or the maximum raw score possible, thus potentially producing a biased flattening of the regression slope. Tobit regressions are designed to minimize any bias due to such effects (Amemiya, 1973).

Importantly, across the three studies, this strong relationship between perceived and actual performance was qualified for 5 out of 8 self-evaluation measures by a quadratic relationship between actual and perceived performance (see Model 2, Table 2). These quadratic relationships showed up in both studies using the Wason task, but did not emerge in the study focusing on a classroom exam. As Figs. 1 and 3 show, these quadratic trends mean that the relationship between perceived and actual performance flattened as performance got worse.

However, despite these strong relationships between actual and perceived performance, a Dunning–Kruger pattern emerged in each data set. In Study 1, participants in the bottom quartile drastically overestimated how well they had done by roughly 40 percentile points. Top performers underestimated how well they had done, but not to the same degree of their poor performing counterparts, underestimating themselves by roughly 10–12 percentile points, depending on the measure. In terms of raw score, bottom performers drastically overestimated how many items they got right, thinking they got 5.5 items right when their actual score was barely 0.3. Top performers, who all achieved perfect scores, underestimated their raw score, but not to the same degree, roughly by only 1 item.

In Study 2, bottom performers facing that first classroom exam again drastically overestimated their performance in both percentile and raw score measures. They overestimated their raw exam score by almost 4 points ($p < .0001$) and their percentile standing by over 50 points ($p < .0001$). Top performers underestimated how well they did, but again not to the same degree, underestimating themselves by 2.2 raw points and 10 percentile points, both ps < .0001.
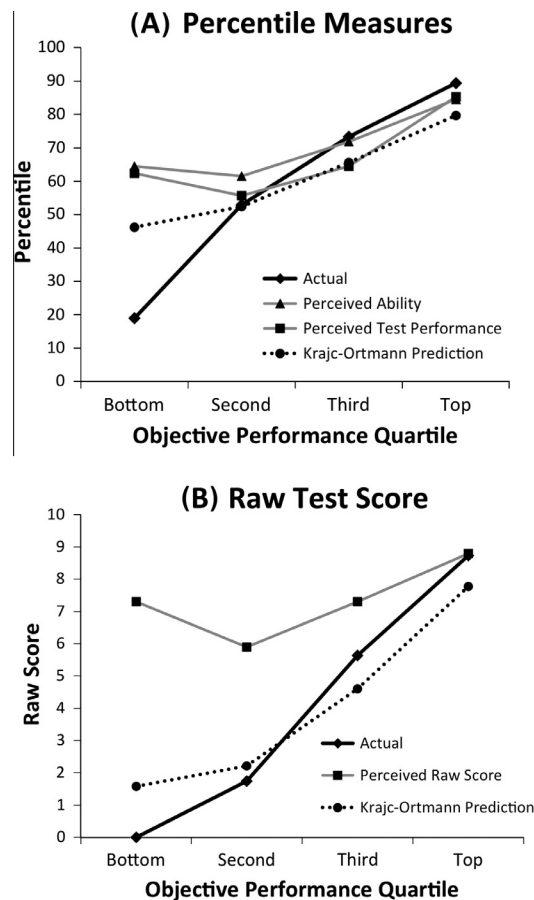
Fig. 3. Actual versus perceived performance (Study 3; Wason task with backwards-*J* performance distributuion). Dotted line represents where perceived performance should lay according to the Krajč–Ortmann model.

Table 2
Relation between actual performance, perceived ability and specific test scores.

Study/self-assessment measure	Model 1	Model 2		Correlation between actual performance and self-assessment error[a]
	Linear	Linear	Quadratic	
Study 1				
Ability percentile	.26***	.26**	.00	−.08
Test score percentile	.28***	.28**	.01**	−.01
Raw score	.37***	.34**	.08**	−.35***
Study 2				
Test score percentile	.23***	.23**	−.00	−.11*
Raw score	.49***	.46**	−.01	−.26***
Study 3				
Ability percentile	.27***	.21*	.01***	−.36***
Test score percentile	.27***	.25*	.02***	−.28***
Raw score	.23***	−.02	.09**	−.19*

* $p < .05$.
** $p < .01$.
*** $p < .005$.
[a] Self-assessment errors are measured by taking the absolute value of the residual remaining for each participant in Model 2. Smaller values indicate less random error, and thus more accuracy.

In Study 3, bottom performers overestimated their performances by over 40 percentile points. They thought, on average, they had gotten 7 out of 10 items right when the actual figure was 0. Top performers underestimated their performance by only 5 percentile points, and should virtually no underestimation when it came to raw score. In sum, all three studies

depicted the usual pattern of self-assessments comprising the Dunning–Kruger effect (not a surprise, since the first study is a re-reporting of one of the original studies introducing the effect).

3.2. Testing assumptions underlying the Krajč–Ortmann model

Studies 1 and 2 showed the pattern traditionally associated with the Dunning–Kruger effect. Did this pattern arise because the underlying performance distribution mimicked the proposals of Krajč and Ortmann (2008), with many performers bunched down at the bottom end of the performance scale and only a few reaching its top rungs?

Fig. 4A and B depicts the underlying raw score performance distributions for the two studies. As seen in the figure, both distributions fail to show the positive skew assumed by Krajč and Ortmann (2008). Their example distribution had a positive
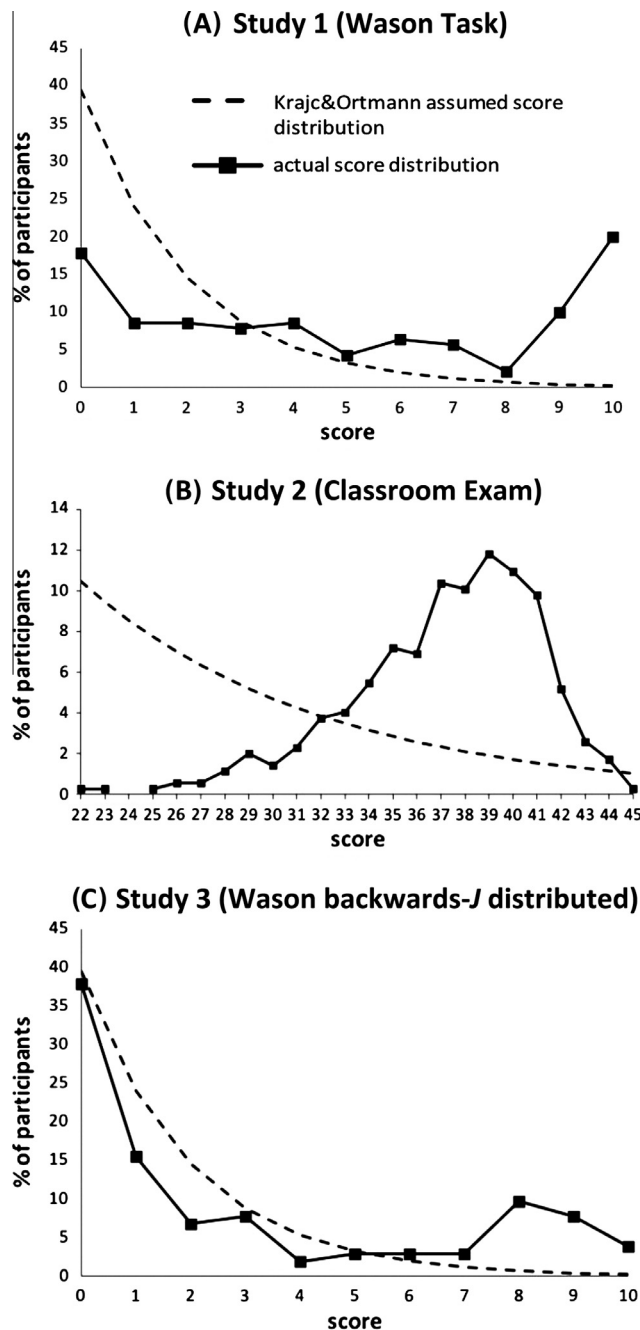

Fig. 4. Objective performance distributions compared with the backwards-*J* distribution (dashed line) assumed by Krajč–Ortmann model.

skew of 1.07, but the same statistic for these two studies, 0.11 (Fig. 4A) and −0.79 (Fig. 1B) respectively, revealed little skew for the Wason task and a visible negative skew for the course exam. Thus, the Dunning–Kruger pattern of self-misperceptions arose even though the underlying performance distributions differed quite a bit from the type assumed by Krajč and Ortmann.

Our examination revealed another assumption from the Krajč–Ortmann model that proved false across all three studies. Krajč and Ortmann (2008) asserted that high and low performers possess equal skill at evaluating their own skill and performance. We decided to test this assumption. Figs. 2–4 show that high and low performers obviously show different *biases* in their self-judgments, with low performers overestimating theirs and high performers underestimating theirs. But, once that difference in bias in accounted for, do both groups show equal levels of *random error* in their self-judgments?

Thus, in each study, and for percentile and raw scores separately, we regressed perceived performance on objective performance—thus, accounting for this level of bias in perceived performance related to objective performance. In Tobit regressions, we controlled for both linear and quadratic components of objective performance. We then took the absolute value of the residual left for each participant in each regression. According to the Krajč–Ortmann model, this measure of error should be unrelated to the participant's objective performance because their skill at judging their own performances is the same across all performance levels. However, if high performers are better at estimating their performance than their low performance counterparts, they should produce smaller errors than their low performing counterparts.

More specifically, we should see a negative correlation between this measure of error and objective performance. As the last column of Table 2 shows, the self-estimates of high-performers displayed much less random error than did those of poor performers. All 8 correlations are negative ($p < .005$, by sign test), meaning the discrepancy between actual and perceived performance is diminishing as performance goes up. For 6 of the 8 correlations, the relationship is statistically significant (i.e., $p < .05$). Thus, overall, high performers showed more evidence of capturing the signal of their performance relative to poor performers.

3.3. Modeling self-evaluation of performance via the Krajč–Ortmann model

Recall that Krajč and Ortmann (2008) conjectured that errors would be truncated for participants at the upper and lower end of the distribution of actual performance: Bottom performing participants could not estimate that they did any worse than they really did (due to floor effects) and top performing participants could not overestimate themselves (due to ceiling effects). Along this reasoning Krajč & Ortmann defined perceived ability as $y = x + \varepsilon$, where x denotes real ability and ε the perception error. They assumed this error to be normally distributed and homogenous across subjects independent of performance level.

Following these assumptions, we modeled what the forecasting errors of top and bottom performers would look like if their errors were truncated at the bottom for poor performers and at the top for best performers. In the Study 1, based on the Tobit multiple regression analysis between perceived and actual performance described above for raw score estimates, the residual error in participants' estimates across all performance levels, as indexed by SD (i.e., square root of the error term), was 2.34. For the exam study, the SD = 3.44. For Study 3, the SD = 2.50. Using these estimates, we constructed the probabilities that individual participants would be exactly right in their raw score estimate versus off by a certain number of ranks. Recall in the Wason task that there were 11 ranks of performance, in that a participant could score from 0 to 10. In the exam study, there were 46 ranks, representing scores from 0 to 45. However, the lowest score observed in the exam study was 23, meaning that scores ranged over 23 of the possible 46 ranks.

Thus, in Study 1, assuming a normal curve and using the observed SD, we calculated that there was a 17% chance that the participants would estimate their score correctly, a 15.5% chance of overestimating by one rank, 12% by two, 7.5% by 3, 4% by 4, 2% by 5, and 0.5% by 6. Chances of underestimating in a negative direction were considered symmetric (i.e., the chance of underestimating by one rank was 15.5%, by 2 was 12%, and so on). From these probabilities, we then calculated what each participant's perceived percentile rankings should be. That is, for each rank the participant's estimate could fall in, we took the actual percentile of performance associated with that rank and multiplied it by the probability that the participant's estimate would fall in that rank. We then totaled up the results from all possible ranks for that participant to create a projection of what their overall percentile estimate should be.

Following Krajč and Ortmann (2008), when participants' score fell near the floor or ceiling of performance, we truncated estimates that were no longer possible and renormalized all remaining probabilities. For example, if actual performance fell at the very bottom, we dropped all probabilities associated with underestimating that score and renormalized the errors that were left so that they added to 1.0, meaning that people at the very bottom of performance had a 17% chance of estimating their score correctly (remember, the bottom half of the distribution of judging correctly would be truncated off), 31% of overestimating by one rank, 24% by 2, 15% by 3, 8% by 4, and 4% by 5, and 1% by 6. We followed a similar procedure for Studies 2 and 3.

Fig. 1A shows the results of computing participants' percentile performance estimates in Study 1 following this Krajč–Ortmann procedure, accounting for truncation of error for those near the top and bottom of the performance ladder. For illustrative purposes, and to provide continuity from past work, results the original analysis based on 11 ranks were compressed to the four performance quartiles displayed by Kruger and Dunning (1999). As seen in the figure, the Krajč–Ortmann projection corrects participants' estimates in the right direction for participants in the bottom quartile of ability, but that correction accounts for only a small degree of bottom performers' overestimation of their performance. Participants' actual

performance fell in the 13.2 percentile, and the Krajč–Ortmann model predicts that estimates of ability and test performance should instead be at 29.3. However, Kruger and Dunning (1999) observed self-estimates that were much higher: 52.3 for estimates of test performance and 54.9 for estimates of logical ability in general. In short, the Krajč–Ortmann framework produced a correction among bottom performers that was quite insufficient, accounting for roughly only 40% of the difference between true performance levels and perceptions of performance. Perceptions of performance lay far higher than what the Krajč–Ortmann model would suggest, both $ts > 5.2$, $ps < .0001$. Thus, their model leaves the vast majority of the original Dunning–Kruger effect unexplained.

That said, we found their model predicts self-evaluation errors among top performers rather well. Posting an average performance at the 90.0 percentile, top-performing participants tended to think that their own test score lay at 79.2 percentile and their general ability at 76.0. The Krajč–Ortmann model predicts a perceived percentile ranking of 76.0, a result that does not significantly differ from either self-estimations, $ts < 1$.

Raw score estimates painted a similar picture. As Fig. 1B shows, the Krajč–Ortmann model suggests that bottom performers would estimate their raw score to be roughly 2.45 out of 10, when actual estimates observed in Kruger and Dunning (1999) averaged 5.52, $t(36) = 6.22$, $p < .0001$, which means the model corrected for only 30% of the overestimation produced by bottom performers. However, for top performers, the estimate participants were projected to make, 8.6, was indistinguishable from the mean estimate (8.9) actually observed, $t < 1$.

Fig. 2A depicts the results of a Krajč–Ortmann projection for Study 2. For participants in the lowest quintile of ability, the projection corrects percentile estimates of test performance to only a minimal extent. Given an average actual performance ranking of 8.1, the projection predicted that participants would perceive their performance to be in the 13th percentile. However, in reality, participants tended to see their performance as falling in the 58.1 percentile. As such, the Krajč–Ortmann model explained only 9.8% of the difference between actual and perceived performance among bottom performers, and the projection severely underestimated how participants would rate their performance, $t(57) = -17.74$, $p < .0001$.

For the top quintile, the picture is again different. Actual performance lay in the 90.5 percentile, although participants perceived that their performance lay in the 80.3 percentile. According to the Krajč–Ortmann model, the average self-evaluation should have fallen in the 78.1 percentile, a result not differing from estimates actually observed, $t(76) = 1.25$, ns. For raw scores (Fig. 2B), the Krajč–Ortmann model produced no correction for bottom performers, since there was no bottom boundary that would truncate their estimates. The estimates of top performers were partially corrected by 38%, but the projected estimate ($M = 40.9$) was significantly different from estimates actually observed ($M = 39.4$), $t(76) = 10.94$, $p < .0001$.

In sum, we found a mixed picture for how well the Krajč–Ortmann model accounted for the Dunning–Kruger pattern of self-evaluation errors. For bottom performers, the self-performance estimates suggested by the model did not recapitulate the dramatic overestimation seen in the original Kruger and Dunning (1999) report and since. However, the model did reasonably well in reproducing the degree of underestimation observed among top performers.

3.4. Testing the Krajč–Ortmann in more "ideal" conditions

The first two studies showed that the Krajč–Ortmann model failed to account for Dunning–Kruger errors in self-evaluation in data sets that produce those errors, at least among poor performers. However, because these two data sets did not fit the statistical assumptions underlying the model, it left open the possibility that the Krajč–Ortmann may work under some circumstances—that is, in data sets in which the performance distribution has a backwards-J shape. Thus, we explored this possibility by examining how well the Krajč–Ortmann framework anticipates self-assessment errors in Study 3, in a data set that clearly has a backwards-J distribution (see Fig. 4C). The positive skew in this data set was 0.82.

As seen in Fig. 3A, much more error made by bottom performers was corrected by the model. Whereas the actual performance of the bottom group lay at the 19th percentile, the model predicts that they would have perceived it as being near the 46th percentile. However, although a large correction, it is less than 65% of the real gap between perceived and actual performance, in that participants tended to see their ability and test performances as laying above the 62nd percentile, $ts > 4.0, ps < .0001$. For raw score, the correction is much more meager. Participants in the bottom group scored 0 but thought they had answered roughly 7.3 items correctly. The Krajč–Ortmann corrects their perceptions to only 1.58 items correct, still far away from what participants perceived, $t(38) = -15.04$, $p < .005$, thus failing to explain nearly 80% of the gap between perception and reality of performance.

For top performers, the model at times overcorrects for error. For percentile measures, the actual performance of the top group lay in the 89th percentile, but participants tended to see it as falling near the 85th. The model successfully anticipated that difference, projecting participants to see their performance to lay in the 84th, $ts < 1$. In terms of raw score (Fig. 3B), participants achieved in reality a score of 8.7 and their estimates almost matched that, $M = 8.8$. The model predicted their self-estimates would drop to 7.78, thus causing an over-correction, $t(21) = -3.18$, $p < .005$.

In sum, in the data set most "ideal" for the Krajč–Ortmann model that we could find, we observed that the model provided more correction for bottom performers—but that the level of correction was far from perfect even in these most ideal conditions. For top performers, the model over-corrected people's mistakes.

3.5. Learning

Does experience lead people to more accurate self-impressions of performance? To address this question, we examined whether Study 2 participants provided more accurate self-assessments on the second exam than on the first. Like Ryvkin et al., 2012, we first looked at students who performed on the bottom rung on both exams ($n = 58$). We found no difference in how much they overestimated their performance, both in percentile ($Ms = 45.1$ and 46.3 for exams 1 and 2, respectively, $t = .41$, *ns*) or raw score terms ($Ms = 2.2$ and 1.9 for exams 1 and 2, respectively, $t = -.48$, *ns*). When we look at the absolute value of their errors, we again find no improvement (for percentile estimates, $Ms = 45.3$ and 46.3 for exams 1 and 2, respectively, $t = .38$, *ns*; for raw scores, $Ms = 4.3$ and 3.7 for exams 1 and 2, respectively, $t = .93$, *ns*.

Next, we examined possible increases in accuracy within all performance levels. Before doing so, however, we had to recalculate who qualified as high and low performers. Assuming that overall performance on both exams was the best indicator of performance quality, we sorted participants into five groups based not just on their performance on the first exam but instead on their total performance on both exams taken together. We then calculated various measures of accuracy in self-assessment for these newly created performance groups. For assessments of raw score, we first calculated how much participants over- versus underestimated their score on each exam, thus creating a directional measure of judgmental error. We then took the absolute value of each participant's error, thus creating an absolute measure of judgmental error. We then submitted each of those measures (directional and absolute) to separate 5(Performance Quintile) × 2(Exam: First or Second) ANOVA, with the last variable serving as a within-subject variable.

As seen in Table 3, accuracy was no greater for the second exam for either the directional measure of error ($Ms. = -.04$ and $-.20$ for first and second exams, respectively, $F < 1$, *ns*) or the absolute value one ($Ms = 2.8$ versus 2.5, for first and second exams, respectively, $F(1,304) = 1.81$, *ns*). The direction and magnitude of error did, not surprisingly, differ across performance groups. The higher the performance group, the more it under- rather than overestimated its performance, $F(4,304) = 19.47$, $p < .0001$. The absolute magnitude of estimation errors were also smaller for higher performance groups than for poorer ones, $F(4,304) = 11.29$, $p < .0001$. We calculated similar measures of directional and absolute error for percentile estimates. Performance quintiles differed greatly both in terms of their directional errors, $F(4,304) = 115.68$, $p < .0001$, and absolute ones, $F(4,304) = 80.27$, $p < .0001$. Both these effects showed that high performers underestimated themselves more than did low performers (who dramatically overestimated their own performance), and that high performers showed less absolute error on average than their lower performing counterparts.

Importantly, across all participants we found no overall improvement in self-evaluation on the second exam, $Ms = 19.3$ and 21.3 for first exam, respectively, on the directional measure, $F(1,304) = 1.60$, *ns*, and $Ms = 27.3$ and 28.0, respectively, for first and second exams, $F < 1$, *ns*, on the absolute value measure. We did, however, find some whispers of self-judgmental improvement, but for only those in the top quintile. Students in this group went from underestimating their performance in percentile terms on exam 1 to slightly overestimating it on exam 2, $Ms = -7.6$ and 1.8, respectively, for exams 1 and 2, $t(52) = 2.93$, $p < .01$, although no improvement was shown on the absolute value measure, $t < 1$, *ns*. Students in the low performing quintile did not show any improvement ($Ms = 45.3$ and 46.3 for exams 1 and 2, respectively).

In sum, we found little evidence, if any, toward more accurate self-evaluations with more experience. If anyone, it was high performers who better assessed whether their achievement lay relative to their peers. But beyond that no convergence in perception toward the truth was observed, replicating other findings, but not the proposals made by Krajč and Ortmann (2008).

Table 3
Directional and absolute discrepancies of perceived estimates from actual performance for exam 1 and 2 (study 2).

	Objective performance quintile				
	Bottom	Second	Third	Fourth	Top
Percentile measure					
Directional discrepancy					
Exam 1	45.1	31.4	17.2	8.8	−7.6
Exam 2	46.3	33.9	21.5	3.1	1.8
Absolute discrepancy					
Exam 1	45.3	34.9	22.6	19.9	13.5
Exam 2	46.3	35.6	28.5	16.6	12.5
Raw score measure					
Directional discrepancy					
Exam 1	2.24	0.80	−0.27	−1.10	−1.87
Exam 2	1.88	0.58	−0.43	−1.40	−1.58
Absolute discrepancy					
Exam 1	4.29	2.75	2.04	2.50	2.40
Exam 2	3.70	2.58	2.40	1.80	2.19
n	58	65	64	70	52

4. General discussion

People tend to overrate their performances, with poor performers producing self-evaluations that only a little less positive than the estimates produced by those performing at the top (for a review, see Dunning, 2011). Across three data sets, we put this Dunning–Kruger pattern of performance estimates under close scrutiny, with an eye toward testing a signal-extraction account for it (Krajč & Ortmann, 2008). Across three studies, with one notable exception, we found that Krajč–Ortmann came up short as an adequate explanation.

4.1. Testing the Krajč–Ortmann model

First, the Dunning–Kruger effect emerged in two data sets that failed to hold to statistical assumptions set forth in the model. The model assumes a backwards-*J* distribution of performance—that is, a distribution with profound positive skew—but Study 1 had no noticeable skew and Study 2 evinced a visible negative skew. In addition, the Krajč–Ortmann framework assumes that high and low performers display the same degree of random error (i.e., precision) in their performance estimates. However, after we regressed perceptions of performance on reality of performance, we found that the random error left over was significantly higher among poor performers than it was among top performers in 6 of 8 comparisons. In short, high performers along this metric showed superiority in their performance estimations, in that those estimations tended to show less random variation than those of poor performers.

Second, when we took the assumptions of Krajč–Ortmann model to predict what participants' performance estimates should be, we found that the estimates predicted by the model failed largely to match the actual estimates being given by poor performers. The model predicted that participants should estimate their performance fell in roughly in the 25th percentile in Study 1 and the 13th in Study 2. Actual estimates of poor performers, however, fell between the 52nd and 58th percentile, depending on the study and measure—at least 27 percentile points higher than predicted by the model. In a similar vein, projected estimates of raw score among bottom performers failed to reflect how much those performers overestimated their performance.

Third, when we assessed the Krajč–Ortmann model in an empirical setting that was the most congenial to it we could find (i.e., the data set in which the performance distribution fit a backwards-*J* curve), we found that the model did a much better job at anticipating the self-misestimates of bottom performers, but the model was still far from perfect, missing the self-estimates provided by bottom performing participants by nearly 20 percentile points.

To be sure, there may be some extant situations in which multiple circumstances conspire for the Krajč–Ortmann model to work. That is, there may be occasions when performance distributions are so extreme and self-insight good enough for the model to produce a Dunning–Kruger pattern. Our data suggest that those circumstances, if they exist, are likely few, and that the Krajč–Ortmann framework fails as a more general account of the Dunning–Kruger effect.

That said, there was one general circumstance, however, in which the Krajč–Ortmann model was a success in our data. When we look at underestimation of performance among top performers, we find that the Krajč–Ortmann model makes projections of self-evaluation that are indistinguishable from the actual estimates made by top performers. They tended to underestimate their performance by roughly 10–14 percentile points, and the model predicted that magnitude of underestimation. Although not as strong, projections of raw score estimates also revealed significant correction across studies. In short, although the Krajč–Ortmann model failed to account for overestimation by poor performers, the usual focus of the Dunning–Kruger effect, it provided a much better account of underestimation by top performers.

This result, as well as others, suggest that Dunning and colleagues (see Dunning, 2011; Dunning et al., 2003; Kruger & Dunning, 1999) may have undersold just how accurate top performers might be in assessing performance, even though a central tenant of their framework is that only top performers have the capacity to make accurate judgments of self- and peer performance. The fact that they underestimate themselves, then, may have less to do with any intellectual shortcomings but instead the irritants of statistical artifact (i.e. the one-sided error). Correct for any artifact and the accuracy of top performers reveals itself.

4.2. On other alternative accounts

Note that this analysis applies to the Krajč–Ortmann framework, but it also might apply to the regression to the mean account suggested by other authors (Ackerman et al., 2002; Krueger & Mueller, 2002). Dunning and colleagues have shown with data that a regression account does little to explain overestimation among poor performers (Ehrlinger et al., 2008; Krueger & Mueller, 2002; Sheldon et al., in press). However, it may explain more of the more minor error made by top performers (see Ehrlinger et al., 2008). For example, consider the data presented in Study 1 of Ehrlinger et al. (2008), in which 124 students estimated their exam score in a large-lecture psychology class. In that study, Ehrlinger and colleagues examined how much of the Dunning–Kruger effect would be corrected if the data were adjusted to account for regression to the mean due to measurement error. The data revealed that only 10% of the typical misestimate of bottom performers was corrected by this adjustment. However, among top performers, the data suggest that half of the underestimation along percentile measures and 88% along raw score measures could be eliminated by this adjustment.

That said, data already collected also demonstrates that statistical artifacts cannot be solely responsible for the underestimation by top performers. First, top performers have already been shown to directly display false consensus, in that they overestimate the performances of their peers in general. Second, when we look at underestimation, we find that it is much more systematic for comparative judgments (i.e., judgments of performance against peers) than for absolute judgments (i.e., estimates of raw score) (Ehrlinger et al., 2008). This is anticipated by a false consensus account, in which high performing people accurately assess their absolute performances but overestimate the absolute performances of others, leading to comparative judgments that are erroneous. Consequently, statistically correcting for any false consensus significantly reduces underestimation by top performers (Ehrlinger et al., 2008). Further, just showing top performers the choices of their peers causes them to recognize just how special or unique their own performances are (Hodges et al., 2001; Kruger & Dunning, 1999, Study 3). None of these results are anticipated by an analysis based solely on statistical artifact.

In short, empirical evidence suggests that the underestimation of top performers may importantly involve both statistical artifact as well as psychology dynamics such as false consensus. The question for future research to resolve is how much of the underestimation is due to the former rather than the latter. As such, although a feature of their framework, Kruger and Dunning (1999) may have underestimated the capacity of top performers to produce accurate self-estimates, even ones involving contrasting one's own performance against one's peers.

Finally, we should make one additional note about how these three studies provide a challenge to at least one alternative account for the Dunning–Kruger pattern. Burson et al. (2006) proposed that the pattern should ebb and flow according to the difficulty of the underlying task. When tasks were easy, participants should in general overestimate their performances relative to their peers—with low performers overestimating greatly and high performers overestimating very little, if at all (i.e., the typical Dunning–Kruger effect). However, when the task was difficult, participations should underestimate themselves in general, with high performers underestimating themselves greatly and low performers not much at all (i.e., a mirror Dunning–Kruger effect).

Study 3 contradicts this pattern. It is clearly a difficult task, with participants on average answering only 3 out of 10 items correctly (with a median correct of just 1 item, and over 70% of participants getting less than half the items right). Yet, the typical Dunning–Kruger effect emerged, with low performers overestimating *both* their raw scores and their percentile rank, and high performers slightly underestimating theirs. Other data sets elsewhere show the same pattern. Williams et al. (2013, Study 3) presented participants with a financial estimation task, in which they were asked to gauge the money they would earn on bonds with certain interest rates and maturities. Although participants on average answered only 42% of questions correctly, they still showed the Dunning–Kruger pattern of overestimation, with bottom quartile performers overestimating their performance by some 50–60 percentile points, and top performers underestimating theirs by 5–12.

4.3. Implications for economic behavior

These studies, as well as others in the Dunning–Kruger paradigm, repeatedly show that people with little expertise often grossly overestimate how much they know and how well they perform. Although a well-documented phenomenon, and obviously relevant to enduring psychological and philosophical questions about the self (e.g., do people ever really *know themselves?*), we suggest that the implications of this research for economic behavior should not be neglected.

One can easily find broad implications for understanding the Dunning–Kruger effect for economics. Consider, again, the important issue of information asymmetry, that is, situations in which one person has more knowledge than another in an economic transaction (Akerlof, 1970; Rothschild & Stiglitz, 1976; Spence, 1973). Usually, these situations are modeled in terms of *uncertainty*. Each actor knows the relevant issues surrounding a transaction. One just happens to know more of the answers that address those issues (Égré & Bonnay, 2012). For example, in the classic example of a used car dealer selling a car, the seller is in a much better position to know whether the car being sold is a "lemon" (i.e., a vehicle with flaws), but the buyer does not know. To be sure, the buyer knows about the problem of lemons in general; that buyer is just uncertain about this particular car. Because the buyer is uncertain, he or she demands a lower price, which ultimately drives higher quality cars out of the market, causing that market to spiral into collapse (Akerlof, 1970).

The Dunning–Kruger approach, with its emphasis on meta-cognitive processes, suggests that high and low performers differ not so much in uncertainty as they do in *awareness* (Fagin & Halpern, 1988; Heifetz, Meier, & Schipper, 2006; Égré & Bonnay, 2012). Low performers overestimate themselves because they are not aware of all the issues and questions that matter. In the used car example above, they are not even aware of the existence of lemons. The idea of one fails to occur to them. Thus, they remain blissfully ignorant of how risky or troublesome a used car can be, and so they harbor no wariness about paying a higher price. One can speculate in this world that used car prices, thus, remain high and the market fails to collapse as often predicted.

In short, theorists modeling economic behavior under information asymmetry should be mindful about whether low information actors are characterized best as being uncertain (knowing what they do not know) or unaware (not aware of all that they do not know). The state of knowledge chosen may have significant impacts on what their models will reveal. Tellingly, the confident self-assessments of low performers in the Dunning–Kruger paradigm suggest that low performers often fall into the unaware camp. For example, when college students were asked to play a popular word game, then often estimated that they had missed an average of 18 solutions to the game in each round. In reality, they missed an average of 154 (Caputo & Dunning, 2005).

4.4. The impact of experience

Our second study on exam performance revealed no greater accuracy in performance self-estimates with greater experience. Students at the bottom of the performance ladder showed little increased realization of the true nature of their performance on the second exam relative to the first. Along all rungs of the performance ladder, we observed no increase in accuracy, even though the number of participants included in the study would have detected a .2 SD move in accuracy with 97% likelihood (and one rung was about .3 SD wide). As such, this study joins a growing list of studies showing no increase in self-evaluation insight as people gain experience (Ferraro, 2010; Hacker et al., 2000).

However, these studies cleave cleanly from two research programs that have shown increased self-insight with experience (Miller & Geraci, 2011; Ryvkin et al., 2012). The crucial difference seems to be whether researchers have included an incentive to enhance self-accuracy. It appears that without incentives, people will not show increased self-accuracy with mere experience. However, if experience is combined with incentives for accurate self-assessments, those accurate self-assessments emerge with greater frequency. To be sure, such speculative conclusions prompted by comparisons across studies is always risky, in that studies may differ in any number of ways (e.g., participant pool, feedback design, task difference) that could also produce differences that have nothing to do with incentives or with experience. As such, the specific efficacy of incentives versus efficacy should be put to more exacting rigorous test in future studies within the same experience.

4.5. Concluding remark

That said, there is one intervention that has also been shown to make people more accurate in their self-assessments. That is to make them more competent. Training people to rid them of the gaps and corruptions in knowledge has been shown to make them more accurate self-assessors. Train people in logical reasoning, and people become more accurate in knowing their skill level, and just how impoverished their skill level was before (Kruger & Dunning, 1999). This is perhaps the best evidence yet for the original dual curse explanation of the Dunning–Kruger effect, and maybe the best advice for people who find themselves dealing with it.

Acknowledgement

References

Ackerman, P. L., Beier, M. B., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences, 34*, 587–605.

Akerlof, G. A. (1970). The market for lemons: Quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*, 488–500.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica, 41*, 997–1016.

Austin, Z., Gregory, P. A. M., & Galli, M. (2008). "I just don't know what I'm supposed to know": Evaluating self-assessment skills of international pharmacy graduates in Canada. *Administrative Pharmacy, 4*, 115–124.

Burson, K. A., Larrick, R. P., & 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.

Caputo, D. D., & Dunning, D. (2005). What you don't know: The role played by errors of omission in imperfect self-assessments. *Journal of Experimental Social Psychology, 41*, 488–505.

Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press.

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In J. Olson & M. P. Zanna (Eds.). *Advances in experimental social psychology* (Vol. 44, pp. 247–296). New York: Elsevier.

Dunning, D., Heath, C., & Suls, J. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest, 5*, 69–106.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*, 83–86.

Economist, The. (2004). The curse of nepotism. *The Economist*, 27.

Égré, P., & Bonnay, D. (2012). Metacognitive perspectives on unawareness and uncertainty. In M. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (pp. 321–342). Oxford, UK: Oxford University Press.

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware? Further explorations of (lack of) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes, 105*, 98–121.

Fagin, R., & Halpern, J. (1988). Belief, awareness, and limited reasoning. *Artificial Intelligence, 34*, 39–76.

Ferraro, P. J. (2010). Know thyself: Competence and self-awareness. *Atlantic Economic Journal, 38*, 183–196.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92*, 160–170.

Haun, D. E., Zeringue, A., Leach, A., & Foley, A. (2000). Assessing the competence of specimen-processing personnel. *Laboratory Medicine, 31*, 633–637.

Heifetz, A., Meier, M., & Schipper, B. C. (2006). Interactive unawareness. *Journal of Economic Theory, 130*, 78–94.

Helzer, E. G., & Dunning, D. (2012). Why and when peer prediction is superior to self-prediction: The weight given to future aspiration versus past achievement. *Journal of Personality and Social Psychology, 103*, 38–53.

Hodges, B., Regehr, G., & Martin, D. (2001). Difficulties in recognizing one's own incompetence: Novice physicians who are unskilled and unaware of it. *Academic Medicine, 76*, S87–S89.

Krajč, M., & Ortmann, A. (2008). Are the unskilled really that unaware? An alternative explanation. *Journal of Economic Psychology, 29*, 724–738.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The contribution of social–perceptual skills and statistical regression to self-enhancement biases. *Journal of Personality and Social Psychology, 82*, 180–188.

Kruger, J. M., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*, 1121–1134.

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

Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition & Learning, 6*, 303–314.

Mynttinsen, S., Sundstrom, A., Vissers, J., Koivukoski, M., Hakuli, K., & Keskinen, E. (2009). Self-assessed driver competence among novice drivers: A comparison of driving test candidate assessments and examiner assessments in a Dutch and Finnish sample. *Journal of Safety Research, 40*, 301–309.

Park, Y.-J., & Santos-Pinto, L. (2010). Overconfidence in tournaments: Evidence from the field. *Theory and Decision, 69*, 143–166.

Ross, L. D., Greene, D., & House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279–301.

Rothschild, M. E., & Stiglitz, J. E. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics, 90*, 629–649.

Ryvkin, D., Krajč, M., & Ortmann, A. (2012). Are the unskilled doomed to remain unaware? *Journal of Economic Psychology, 33*, 1012–1031.

Sheldon, O. J., Dunning, D., & Ames, D. R. (in press). Emotionally unskilled, unaware, and uninterested in learning more: Reactions to feedback about deficits in emotional intelligence. *Journal of Applied Psychology*.

Simons, D. J. (2013). Unskilled and optimistic: Overconfident predictions despite calibrated knowledge of relative skill. *Psychonomic Bulletin & Review, 20*, 601–607.

Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics, 87*, 355–374.

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology* (pp. 135–151). Baltimore: Penguin Press.

Williams, E., Dunning, D., & Kruger, J. (2013). The hobgoblin of consistency: Algorithmic judgment strategies underlie inflated self-assessments of performance. *Journal of Personality and Social Psychology, 104*, 976–994.

Worland, J. C. (2011, May 11). Legacy admit rate at 30 percent. *The Harvard Crimson*.

Zenger, T. R. (1992). Why do employers only reward extreme performance? Examining the relationships among performance, pay, and turnover. *Administrative Science Quarterly, 37*, 198–219.