The Point Illusion: Incorrect Weighting of Absolute Performance in Self-assessments

Scott A. Jeffrey, Brian P. Cozzarin

Abstract- **People spend much of their life in an attempt to assess their aptitude for numerous tasks. For example, students expend a great deal of effort to determine their academic standing given a distribution of grades. This research finds that students use their absolute performance, or percentage correct as a yardstick for their self-assessment, even when relative standing is much more informative. An experiment shows that this reliance on absolute performance for self-evaluation causes a misallocation of time and financial resources. Reasons for this inappropriate responsiveness to absolute performance are explored.**

I. Introduction

It has been demonstrated that people fall prey to a "money illusion" [1], where they evaluate money in nominal rather than real (inflation adjusted) terms. A similar effect has also been found with foreign currency, where people are overly sensitive to the nominal value of the foreign currency rather than the value in home currency [2]. These biases have implications for savings behavior [3], spending behavior [4], satisfaction with pay outcomes [5], and fairness perceptions [6]. For example, people believe raising salaries 5% in the presence of 12% inflation is more fair than cutting salaries 7% in the presence of no inflation and workers and trade-unions are far more likely to accept real wage cuts provided nominal wages do not decrease [5]. This paper introduces the concept of a "point illusion", whereby people place too much weight on an absolute evaluations of their performance when relative standing is the more relevant benchmark.

Beyond comparisons in the financial realm (e.g. how much do I earn?), people face many other opportunities for self-evaluation. For example, employees receive regular performance evaluations and students receive grades that evaluate their academic performance. Both of these evaluations often serve the purpose of determining someone's future earnings potential via the availability of good jobs and promotions. These evaluations can be based on absolute performance (e.g. meeting a sales or production quota), relative performance (e.g. performing in a specific percentile of the workforce), or some combination of the two. In most evaluative situations, people will learn of both their relative and absolute performance. For example, in a golf match, a player will know their absolute score and where that score placed them in the final rankings. In order for a person to accurately evaluate their performance, they must choose the appropriate weight for each type of performance feedback. In the golf tournament example, your absolute performance has no bearing whatsoever on your earnings from the match, only your relative position in the final standings.

In the first section of this paper, we will review the literature on self-evaluation and the predictions this literature would make. Then we will explore some reasons why some people (students in these studies) will rely too much on their absolute performance. We then report a study that show an over-reliance on absolute performance when evaluating performance satisfaction and the decision to expend resources towards self-improvement. The final section closes with some implications and directions for future research.

II. The Psychology of Self-evaluation

There has been a great deal of psychological research on how people select the criteria and comparators to use when evaluating they are evaluating their own abilities and performance. Beginning with work by Festinger [7], researchers have proposed that people would seek out objective information when attempting to evaluate their abilities. In the absence of objective performance information, they would compare themselves to similar others, since it was hypothesized that the main motivation of social comparison was accurate self-knowledge [8]. As this stream of research developed, other additional motives were explored. Other research on social comparison has extended this to include two additional motives that would drive the selection of comparators. When people were attempting to self-enhance, they would use downward comparisons [9]. When people had a self-improvement motive, they often would compare upwards, giving them a target or goal [10]. In an academic setting, most students

ISBN: 978-988-98671-0-2

use upwards social comparison in order to motivate themselves to do better. [11, 12].

We hypothesize that if students' self-esteem was threatened by low absolute performance, then the natural tendency predicted by social comparison would be to try to find a downward comparison target [13]. This would cause students to ask for distributional information, as is often the case. By seeing that they were above a certain number of students in the class would appease this negative reaction. Note that this would predict a focus on relative performance levels when reporting satisfaction, particularly for low absolute/high relative performers.

Other research on social comparison has shown that students do not choose downwards social comparison but rather choose upwards comparisons provided they feel that they are able to improve their performance [14, 15]. This would be consistent with a prediction that students would use absolute performance levels to evaluate their abilities. It would also support a prediction that students would use absolute performance levels to allocate resources, given the assumption that they have adopted an improvement motive.

A related stream of research hypothesized that people also compare themselves to previous versions of themselves or future ideal versions [16]. If students were using previous versions of themselves, then a low absolute score would be discouraging, especially for students who had consistently received high absolute scores. Therefore, low absolute scores would trigger a self-improvement motive that would cause them to dedicate more effort to areas in which their absolute score was low. This could also trigger the common request students make for distributional information to handle short term disappointment. However, once the self-improvement motive becomes dominate, students will expend effort where their absolute performance is lower, relative performance held constant.

We believe that in circumstances where people receive both relative and absolute performance feedback, people will place too much weight on their absolute performance when it is less important than relative performance. For example, consider grades in university. For students, grades carry nearly as much, if not more, importance than money. This is reasonable since better grades tend to be correlated with higher salaries [17]. Grades are determined by performance on exams, but in courses that are curved on a forced grade distribution, performance relative to others is the more important metric. The same is true of a golf tournament, where relative standing will determine prize earnings as well as competitions for promotions within organizations [18]. The important performance benchmark in al of these examples is being above a certain number of one's rivals.

Continuing with the grading example, suppose a student is taking two classes and has received his or her midterm grade. One course is known for a hard teacher who gives hard exams and grades them strictly. The other course is taught by a far more lenient professor. In the difficult class, the student receives a score of 65 out of 100 while in the easy class he or she scores 85 out of 100. Further suppose that this student receives information that both of these scores place him or her in the 80th percentile of performance. Since both classes are graded on a curve, he or she should be equally satisfied with the performance on both exams; however we predict that our hypothetical student will be much less satisfied with the 65. While much research has shown that people do use relative comparison when evaluating their own performance [15, 19, 20], we believe it is difficult to ignore absolute performance, even when it is completely irrelevant.

The final reason it is difficult to ignore absolute performance is because often times it is a good signal of performance. In school, we are socialized to learn that absolute performance is a signal of learning, even though there is some arbitrariness in the grading process; some teachers create difficult tests, and sometimes grade assignments more strictly. This fact should make relative class standing is a better indicator of performance. All through school we are taught that 90% is an "A", 80% is a "B", all the way down to 50% or below which was an "F". The natural reaction to scoring 65 out of 100 is that this is a low C or even a D. This could lead a student to question their knowledge of the material in the course and lower satisfaction with performance. A related explanation for this is that students may feel that a test should be representative of the knowledge a student should have received in class. In other words, getting 65% right could be interpreted that the student only understood 65% of the material. This would also lead to below average self-assessments, especially if the student felt he or she knew the material well. All of these reasons suggest that people will have a hard time ignoring absolute performance, even in situations where it is not important or even irrelevant.

> Hypothesis 1: Relative performance held constant, people will be more satisfied with higher absolute performance.

While satisfaction with results is important, the behavioral consequences of those feelings are more important. Suppose that instead of equal relative ranking, the higher absolute score put our hypothetical student at a lower relative standing than the score on the more difficult test. If his or her intention to study is based upon her satisfaction level, and satisfaction is inappropriately affected by absolute performance, then he or she will spend too much time studying for the class in which they are in better relative position. It is probably not the case that a student evaluates his performance solely on absolute performance. However, in cases where relative performance is far more important, an over-reliance on absolute performance will cause a misallocation of improvement resources.

> Hypothesis 2: More resources (time, money, etc.) will be allocated to improving performance in areas where absolute performance is lower, relative standing held constant

III. Method

Fifty-nine engineering students at a large Canadian university were asked to complete a survey that asked participants to imagine that they had just received feedback regarding their performance on a midterm exam. 30 participants were told that they were taking a class from a notoriously difficult professor with a reputation for tough grading standards, while the other 29 participants were told that they were taking a class from a more lenient professor. Students in the more difficult section were given lower absolute scores than those in the more lenient professor's class. All participants were told that the professor used a grading curve in which the top 30% of students receive an A; the middle 60% receive a B; with the rest of the students receiving a C or D. So that the participants could better understand the performance feedback, they were told that there were 100 students in the class, so that these percentages could be easily translated into the number of students who would receive each grade. They were all told that their performance on the test placed them in the 53rd percentile.

All participants were asked to state their satisfaction with their performance. Participants were then told that they were planning how to allocate their time to studying for the final exam in the same course. Specifically, participants were asked how long they believed that they would study for the final exam, using a closed end scale from 2 to 16 hours in increments of 1 hour. After answering this question, they were informed of an internet course that would help them prepare for the final and asked their level of interest in the course and how much they would be willing to pay for that course. To avoid problems with modulus mapping (extreme values), they were asked to provide a number between zero and $100. They were told that if they would not take the course regardless of price, they should enter zero.

To control for order effects between satisfaction ratings and the decision to expend resources studying for the final, the satisfaction question was counterbalanced. In half of the surveys, this question was asked first, in the other half, this question was asked after the resource allocation questions. Finally, demographic information was collected from students, to see if there were any effect of gender and age.

Because of the cover story regarding the professor, participants should realize that their absolute performance on the midterm is a nearly meaningless signal as to how they are doing in the course and how much time and other resources they should expend in preparing for the final. This would imply that there should be no differences in responses between conditions. However, if participants are sensitive to their absolute performance as predicted, then participants in the difficult course will predict spending more time studying for the exam and be willing to pay more money for the Internet preparation course.

IV. Results

Prior to analyzing the results, the data were checked to see whether asking people to rate their satisfaction prior to study plans changed their answers. None of the comparisons approached significance and so responses were combined for all subsequent analysis. Table 1 shows the means of study variables across conditions (high vs. low absolute performance)

Subjects were more satisfied with higher levels of performance when their absolute score was higher, $t(58) = 2.87$, $p < .01$. With respect to the willingness to allocate resources, participants expressed a higher willingness to pay when their absolute performance was lower, $t(58) = 2.39$, $p < .05$. Their intent to study was also higher (although only marginally so), 2-tailed $t(58) = 1.80$, $p = .08$.

TABLE I
SUMMARY STATISTICS

	Absolute Score		
	Low	High	Difference
Satisfaction	2.69 (0.26)	3.58 (0.18)	0.89**
Intent to Study (Hours)	9.03 (0.71)	7.35 (0.61)	1.68†
Willingness to Pay	26.98 (4.56)	13.29 (3.53)	13.69*

† p < .10, * p < .05, ** p < .01 in two-tailed t-test.
Standard errors in parentheses

A more thorough analysis was provided by an OLS regression, which also included the demographic information of age and gender. These results can be seen in Table 2. The addition of demographic information does not change the result that satisfaction is positively impacted by absolute score, with relative position held constant, $\beta = 0.84$, $t(55) = 2.663$, $p < .01$. Since satisfaction with performance might also predict the willingness to allocate resources we included that in the two regressions (WTP and Willingness to study). The coefficients on this variable did not approach significance for either resource allocation question. In addition, it did not cause the significance of the level of absolute score to change. This rules out the change in resource allocation being mediated by satisfaction [21].

A second means by which to test the willingness to take an internet preparation course is by checking the proportion of students in each condition that would not take the course regardless of price. To analyze this question, a binary variable was created such that any participant responding zero dollars to the willingness to pay question was coded as someone who would not enroll in the course. Those that entered any positive number were coded as someone who would enroll in the course. The proportion of participants reporting a positive willingness to pay was 86.2% when absolute performance was low, and only 51.6% when it was

high, $t(58) = 2.88$, $p < .01$. This difference was confirmed in a logistic regression also reported in Table 2. None of the control variables predicted the willingness to take the course, nor did the overall satisfaction level with the performance. Only the coefficient on absolute performance was significant.

TABLE II
REGRESSION RESULTS

Variable	Sat	Study Plans	WTP	Positive WTP
Age	.01 (.10)	.68* (.29)	.77 (1.91)	.11 (.21)
Gender	.24 (.36)	-.62 (1.03)	-3.53 (6.69)	.09 (.69)
Absolute Score	.84** (.32)	-1.83† (0.97)	-15.2* (6.26)	-1.7* (.69)
Satisfaction (Sat)	NA	.07 (.39)	2.48 (2.51)	.001 (.28)
F(3,55)	2.52†	2.55*	1.61	NA
R-square	.12	.16	.11	.18

† $p < .10$, * $p < .05$, ** $p < .01$
Standard errors in parentheses

V. DISCUSSION

This study showed that engineering students are more satisfied with higher absolute levels of performance, even when relative standing is held constant, and all that matters for their overall evaluation. Beyond the difference in satisfaction level, which is highly understandable, students used their irrelevant absolute performance level to determine the allocation of study resources. Students in our study were willing to spend more time studying and more money for preparation courses when their absolute performance was lower, even when their relative position was all that mattered. We believe that this is a mistake, with long term negative consequences, since resources should be allocated to the item that will do the most long term good. If studying resources are misallocated, then overall grades will suffer.

One limitation of this study is that it is hypothetical, asking students to self-report their intention. However, the authors believe it is reasonable to assume that students have adequate psychological access to their future behaviors with respect to studying behavior. Also, there is little reason to believe that students would not respond accurately to these questions, reducing the potential negative consequences of this limitation.

From a practical standpoint, if a professor wants their students to spend more time studying the material in their course, then they should be strict graders on assignments during the term. The problem is that strict graders tend to receive poor teaching evaluations. Therefore, there is a mixed incentive for faculty members. In order for their students to take their topic seriously, they should grade strictly, yet their own performance evaluations may suffer, affecting salary decisions as well as tenure and promotion decisions.

In the end, it is in the professors best interest to try to effectively communicate the grading criteria, and be certain that students understand exactly where they stand with respect to their overall final performance. While this seems easier said than done, to be true educators, professors must try to avoid inflating students' marks so that they can increase their own teaching evaluations. In the end, the best strategy would be to create assignments and tests that reflect true understanding of the material so that students "curve themselves". This would then make absolute performance proxy nicely for relative standing.

While the participants in this study were undergraduate students, talking about studying, we believe that this has practical implication in a number of other areas. Organizations often suffer from ratings inflation in their performance evaluation, where few employee are rated "below average", even when they are relatively low performers. This can cause employees to think they are doing better than they really are, creating two potential negative outcomes. First, they may feel treated unfairly if they receive a below average raise, since raises are generally determined by relative standing. In addition, if the firm moves to a forced curve rating system, then their subsequent evaluations will be much lower, leading to bitterness.

VI. CONCLUSION

In most situations, people receive both types of feedback, relative and absolute performance, and must make judgments about how to combine this feedback. The results of this study suggest that people will put too much emphasis on their absolute performance, and underweight their relative position. This could have negative consequences such as the allocation of resources (time and money) to improving in the wrong areas.

VII. REFERENCES

[1] E. Shafir, P. Diamond, and A. Tversky, "Money illusion," *Quarterly Journal of Economics,* vol. 112, pp. 341-374, MAY 1997.

[2] P. Raghubir and J. Srivastava, "Effect of face value on product valuation in foreign currencies," *Journal of Consumer Research,* vol. 29, pp. 335-347, Dec 2002.

[3] E. Koskela and M. Viren, "Inflation and Savings: Testing Deaton's Hypothesis," *Applied Economics,* vol. 14, p. 579, Dec 1982 1982.

[4] W. H. Branson and A. K. Klevorick, "Money illusion and the aggregate consumption function," *American Economic Review,* vol. 59, pp. 832-849, 1969.

[5] G. A. Akerlof, W. T. Dickens, and G. L. Perry, "The macroeconomics of low inflation," *Brookings Papers on Economic Activity,* vol. 1, pp. 1-76, 1996.

[6] D. Kahneman, J. Knetsch, and R. H. Thaler, "Fairness and the assumptions of economics," 1985.

[7] L. Festinger, "A theory of social comparison processes," *Human Relations,* vol. 7, pp. 117-140, 1954.

[8] C. Sedikides and M. J. Strube, "Self evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better," in *Advances in experimental social psychology*. vol. 29, M. P. Zanna, Ed. San Diego, CA: Academic Press, 1997, pp. 209-269.

[9] L. G. Aspinwall and S. E. Taylor, "Effects of social-comparison direction, threat, and self-esteem on affect, self-evaluation, and expected success," *Journal of Personality and Social Psychology,* vol. 64, pp. 708-722, May 1993.

[10] J. F. Ybema and B. P. Buunk, "Aiming at the Top - Upward Social-Comparison of Abilities after Failure," *European Journal of Social Psychology,* vol. 23, pp. 627-645, Nov-Dec 1993.

[11] H. A. Wayment and S. E. Taylor, "Self-evaluation processes: Motives, information use, and self-esteem," *Journal of Personality,* vol. 63, pp. 729-757, 1995.

[12] M. Foddy and I. Crundall, "A field study of social comparison processes in ability evaluation," *British Journal of Social Psychology,* vol. 32, pp. 287-305, 1993.

[13] B. P. Buunk, S. E. Taylor, G. A. Dakof, R. L. Collins, and N. W. Vanyperen, "The Affective Consequences of Social-Comparison - Either Direction Has Its Ups and Downs," *Journal of Personality and Social Psychology,* vol. 59, pp. 1238-1249, Dec 1990.

[14] F. X. Gibbons, H. Blanton, M. Gerrard, B. Buunk, and T. Eggleston, "Does social comparison make a difference? Optimism as a moderator of the relation between comparison level and academic performance," *Personality and Social Psychology Bulletin,* vol. 26, pp. 637-648, May 2000.

[15] N. Michinov and J. M. Monteil, "Upward or downward comparison after failure. The role of diagnostic information," *Social Behavior and Personality,* vol. 25, pp. 389-398, 1997.

[16] S. Albert, "Temporal comparison theory," *Psychological Review,* vol. 84, pp. 485-503, 1977.

[17] P. L. Roth and R. L. Clarke, "Meta-analyzing the relation between grades and salary," *Journal of Vocational Behavior,* vol. 53, pp. 386-400, Dec 1998.

[18] E. P. Lazear and S. Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy,* vol. 89, p. 841, Oct 1981 1981.

[19] L. Festinger, "The motivating effect of cognitive dissonance," in *Assessment of human motives*, G. Lindzey, Ed. New York: Holt, Rinehart, & Winston, 1958.

[20] D. J. Lane and F. X. Gibbons, "Social comparison and satisfaction: Students' reactions after exam feedback predict future academic performance," *Journal of Applied Social Psychology,* vol. 37, pp. 1363-1384, Jun 2007.

[21] R. M. Baron and D. A. Kenny, "The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations," *Journal of personality and social psychology,* vol. 51, pp. 1173-1182, Dec 1986.