Depression and Accuracy: Evidence from the 2010 FIFA World Cup

Kriti JAIN
J. Neil BEARDEN
Allan FILIPOWICZ
2011/19/DS/OB
Depression and Accuracy:
Evidence from the 2010 FIFA World Cup

Kriti Jain *

J. Neil Bearden**

Allan Filipowicz***

* PhD Candidate in Decision Sciences at INSEAD 1, Ayer Rajah Avenue, Singapore 138676, Singapore. Email: Kriti.jain@insead.edu
Corresponding author.

** Assistant Professor of Decision Sciences at INSEAD 1, Ayer Rajah Avenue, Singapore 138676, Singapore. Email: neil.bearden@insead.edu

*** Assistant Professor of Organisational Behaviour at INSEAD 1, Ayer Rajah Avenue, Singapore 138676, Singapore. Email: allan.filipowicz@insead.edu

A Working Paper is the author’s intellectual property. It is intended as a means to promote research to interested readers. Its content should not be copied or hosted on any server without written permission from publications.fb@insead.edu
Click here to access the INSEAD Working Paper collection
Abstract

Before and during the 2010 Soccer World Cup, participants made probabilistic forecasts for the outcomes of the tournament. We examine the relationship between depression levels and performance in this real-world forecasting task. Across two different waves of predictions, for both continuous and categorical classifications of depression, and with multiple measures of prediction accuracy, we find that depressed forecasters were less accurate. The poorer accuracy amongst the more depressed forecasters was mediated by neglect for base-rate probabilities: The depressed participants assigned probabilities that departed more substantially from the baserates, particularly for low base-rate events. Given the high incidence of depression in the workforce, the importance of probabilistic forecasting as a judgment task, and that we may be the first to look at forecasting accuracy on an ecologically valid task with objective outcomes over which the forecaster has no control, these findings may have important implications for both theory and practice.

Keywords: Depression; Judgment; Prediction; Depressive Realism
Depression and Accuracy: Evidence from the 2010 FIFA World Cup

Research on depression, a more intense and chronic affective state than is typically produced in laboratory settings, has long been used to elucidate a variety of basic social cognitive processes (Weary & Edwards, 1994; Weary, Edwards & Jacobson, 1995). One area that has attracted considerable research attention has been the influence of depression on judgment and in particular forecasting (e.g. Alloy & Abramson, 1979; Alloy & Ahrens, 1987; Dunning & Story, 1991; Strunk, Lopez & DeRubeis, 2006). This has important practical ramifications for three reasons: There is a high incidence of depression in the workplace, with estimates ranging from 5-10% of the workforce suffering clinical depression each year (Gabriel & Liimatainen, 2000; Kessler, Merikangas & Wang, 2008). Sub clinical levels of depression (which are even more prevalent) have important effects on judgment and decision making (e.g. Weary et al. 1995). And we have never properly measured the influence of depression on a critical judgment task, the ability to make accurate probabilistic forecasts, in spite of long standing critiques of the depression-accuracy research (e.g. Ackerman & DeRubeis, 1991; Dunning & Story, 1991; Haaga & Beck, 1995)

There has been a great deal of interest in the impact of depression on the correspondence between beliefs and reality. According to Beck (1967), depressed people tend to misperceive reality in a systematically negative or pessimistic way, thereby feeding their depression. Others have argued that it is the beliefs of non-depressed people that are most at odds with reality – in particular, that non-depressed people tend to have beliefs that are excessively optimistic (e.g. Taylor & Brown, 1988). A key assumption guiding much of the previous work on depression is
that the relative mismatch between beliefs and reality will in turn cause one group to do better or more poorly on judgment and forecasting tasks.

Alloy and Abramson (1979) reported that depressed subjects were better able to judge the extent to which they exercised control in a laboratory contingency judgment task. The view that depressed peoples’ beliefs are in greater correspondence with reality has come to be known as depressive realism (see Ackerman & DeRubeis, 1991; Haaga & Beck, 1995; Weary & Edwards, 1994 for reviews; Au, Chan, Wang & Vertinsky, 2003 for a recent variation). And though the idea has received mixed empirical support, with depressives showing a tendency towards a negative bias and inaccuracy when one moves away from laboratory tasks towards more realistic, personally relevant stimuli (e.g. Pacini, Muir & Epstein, 1998), the idea still has broad popular currency, as this line in a recent *New York Times* article suggests, “And then there's ‘depressive realism’: several studies have found that people with depression have a more accurate view of reality and are better at predicting future outcomes” (Lehrer, 2010, p. 38).

Alternatively, depressed people may simply have a systematic pessimistic, or negative distortion (Beck, 1967). The depressive pessimism view has received a range of empirical support (e.g. Alloy & Ahrens, 1987; Pietromonaco & Rook, 1987; Strunk, Lopez & DeRubeis, 2006), and suggests that the best accuracy can be obtained by combining the (negatively skewed) forecasts of those who are depressed with the (positively skewed) forecasts of those who are not. However, these studies are subject to the critiques of depression-accuracy research that we will return to below.

A third possibility is that depressed participants’ predictions are simply inaccurate. Dunning and Story (1991) found that depressed subjects’ predictions about self-relevant future actions and outcomes were less accurate. The depressed subjects assigned higher likelihoods to
Depression and Accuracy

the occurrence of low base-rate events (e.g., being the victim of a crime), and they also tended to be less accurate when making relatively optimistic predictions. Strunk, Lopez and DeRubeis (2006) also found that depressed individuals were less accurate when predicting the occurrence of negative real-world self-relevant events. Importantly, the events for which subjects made forecasts in these studies were, to some extent at least, subject to self-fulfilling prophesies; that is, the outcomes could be influenced by the forecaster (e.g., “I will end a major relationship.”).

Accurate probabilistic forecasting, the ability to correctly assign high probabilities to events that will occur, and low probabilities to events that will not, is an important and organizationally relevant skill. By correctly assessing the probability that interest rates (or the stock market, or currency exchange rates, or even product demand) will move in a certain direction, one can adequately take advantage of and also protect against such movement. And it is possible that one tenth or more of the workforce is better at this task, is biased at this task in a way that balances the remainder of the workforce, or is worse at this task. We attempt to adjudicate among these possibilities.

We are interested in the role of depression in predicting events with genuine, exogenous uncertainty. Our prediction task – in which subjects predict the outcomes of the 2010 Soccer World Cup – satisfies a number of conditions that depression-accuracy research should have (for a discussion of these desiderata, see Ackerman & DeRubeis, 1991; Dunning & Story, 1991; and Haaga & Beck, 1995). First, the task has ecological validity: making judgments about teams winning sporting events is something that (many) people actually do. Second, the task allows us to objectively assess the accuracy of people’s forecasts. Finally, participants in our prediction task cannot influence the outcomes. In short, our task allows us to address the question of whether depressed people are better at making predictions using the term “prediction” in a sense
in which it is used with high frequency – and which has been neglected in previous research on depression and accuracy.

The Present Studies

The FIFA World Cup – the “Soccer World Cup” – is a men’s soccer tournament held every four years. Thirty-two national teams from around the world compete in a tournament that runs over a roughly one month period. The tournament consists of a series of stages through which teams get sequentially eliminated until a single winner, the “champion,” remains. In 2010, the FIFA World Cup was held in South Africa from June 11 to July 11. By some estimates, the final match alone was viewed by 700 million people. We had soccer fans make probabilistic forecasts for teams making it to various stages of the tournament, and also complete a standard depression measure.

Method

Participants and Procedure

In the Main Round of the study, during the week preceding the World Cup, 1110 soccer fans (906 males, 204 females) took part in an online survey for the chance to win cash prizes. Participants were drawn from current and former students and staff at INSEAD, as well as from the broader community of soccer fans in Singapore.

Each participant from the Main Round was also invited to take part in a “Second Round” just after the Quarter-Finals (when only 8 of the 32 teams remained in the tournament). They were given a 24 hour window to complete the survey. Four-hundred and sixty-one participants took part in the Second Round. The Second Round serves as a simultaneous (quasi-)replication of the Main Round.
Financial incentives. For the Main Round, participants were informed that five out of every 100 participants would be selected at random, and paid up to SGD 100 based on performance. In addition, the two top players would earn SGD 500. One out of every 100 participants in the Second Round was selected randomly to receive cash payoffs linked to forecasting performance. In total, we paid out nearly SGD 5000 (USD 3800) in performance-based incentives. The incentive mechanism is described below.

Measures

Depression. Each participant completed the Beck’s Depression Inventory (BDI-II; Beck, Steer & Brown, 1996). The BDI-II has 21 items that measure a range of depressive symptoms. Each item is scored from 0 to 3, and the total score is obtained by summing the individual item scores. Higher scores indicate greater severity of depressive symptoms (for our sample, $M = 8.11$; $SD = 7.30$; $\alpha = 0.89$). Beck et al. reported that the test-retest reliability of the BDI-II administered seven days apart was 0.93. Henceforth, we will refer to the BDI-II simply as the BDI.

Probabilistic forecasts. In the Main Round, each participant gave probabilistic forecasts for 20 of the 32 World Cup teams. The 20 teams were selected at random for each participant. Using a subset of the teams served two methodological purposes. First, the random selection of teams helped keep the time required to complete the survey reasonable. Second, the random selection helped partially control for the effects of team favoritism, which we worried could introduce confounding influences into our analyses. For each of the 20 teams, each participant in the Main Round estimated the probabilities of the following five events:

- Stage 1: The team will make it to the Round of 16.
- Stage 2: The team will make it to the Quarter-Finals.
Depression and Accuracy

- Stage 3: The team will make it to the Semi-Finals.
- Stage 4: The team will make it to the Finals.
- Stage 5: The team will win the World Cup.

In total, each participant gave 100 probabilistic forecasts (20 teams × 5 events). Just after the Quarter-Finals, participants who took part in the Second Round gave estimates for the remaining eight teams making it to the Semi-Finals, Finals, and winning the tournament (8 teams × 3 events = 24 forecasts).

**Scoring the probabilistic forecasts.** We used a *quadratic scoring rule* to incentivize the participants to give their best probabilistic estimates for each event (Winkler, 1969). Specifically, for assigning probability \( p \) to an event with outcome \( x \) (where \( x = 1 \) if the event happens, and \( x = 0 \) otherwise) participants earned \( \pi(p, x) \) points according to the following function:

\[
\pi(p, x) = [1 - (x-p)^2] \times 100.
\]

Suppose, for example, that a participant assigned an 80% chance \( (p=0.80) \) to the event “France will make it to the Round of 16.” Since France did not make it to the Round of 16, the participant would earn 36 points for this forecast. If, in contrast, she had estimated the chances to be 20%, her payoff would have been 96 points. The scoring rule rewards assigning high probabilities to events that happen and low probabilities to events that do not. More accurate probabilistic estimates earn forecasters higher points, which then translate to higher earnings. This scoring rule has been used extensively to evaluate professional forecasters (e.g., Winkler & Murphy, 1968), and also as an incentive-compatible mechanism in judgment research (e.g., Camerer, 1987).

To make the scoring system plain, participants had access to a simple table that showed how their probability judgments translated to points depending on whether the event they were
forecasting occurred or not. Under this system, a participant can maximize her score by responding truthfully for each event, that is, by giving her best estimate \( p \).

**Naïve bookmaker benchmark.** We wish to evaluate forecasting performance across tournament stages; so we must adjust for task difficulty across stages. We do so by evaluating each participant’s forecasts within a stage relative to a simple benchmark: the score of the naïve bookmaker who assigns the base-rate probabilities to each team in that stage (ignoring team quality, rankings, and so on). Since there are 32 teams at the start of the tournament and only half can make it to the first stage (the “Round of 16”), the naïve bookmaker would assign probability 0.50 to each team making it to that stage. Likewise, he would assign probability 0.25 to each team making it to the second stage (the Quarter finals). And so on.

Therefore, the *expected points* to the naïve bookmaker who uses the base-rate probabilities for all events is given by:

\[
E[\pi(q_s, \cdot)] = (q_s[1 - (1-q_s)^2] + (1-q_s)[1-(q_s)^2]) \times 100,
\]

where \( q_s \) is the base-rate probability for stage \( s \). The naïve bookmaker’s expected payoffs are displayed in Table 1. Our primary performance measure will be the *gain score*, or percent improvement from the naïve bookmaker’s expected payoffs. For each participant, we calculate average points for each stage \( s \), denoted \( \pi_s \), and compute the gain score for that stage:

\[
g_s = (\pi_s / E[\pi(q_s, \cdot)]) - 1) \times 100.
\]

If \( g_s = 0 \), then the judges’ estimates produced points no better than those that one would get by assigning the base-rate probability \( q_s \) to each team in stage \( s \). Positive (negative) \( g_s \) indicate that the judge did better (worse) than the naïve bookmaker. The gain score allows us to control for task difficulty across stages, and also puts scores into an interpretable metric. In short, higher gain scores represent greater forecasting accuracy.
Results

Accuracy and Payoffs

To examine the forecasting quality of the participants, we ran a repeated measures ANCOVA with gain score as the dependent variable, and BDI Score (between-subjects) and stage (within-subjects) as independent variables. For the Main Round, the BDI relationship was significant, $F(1, 1104) = 18.43, p < 0.001, \eta^2 = 0.02$, with higher BDI scores associated with lower gain scores. There was also a main effect of stage, $F(2.40, 2653) = 60.44, p < 0.001, \eta^2 = 0.05$: the gain scores tended to decrease with stage.$^1$ The interaction was not significant, $p = 0.15$.

The pattern replicated in the Second Round: The BDI relationship was again significant, $F(1, 459) = 8.19, p = 0.004, \eta^2 = 0.02$; there was a significant effect of stage, $F(1.5, 692) = 11.61, p < 0.001, \eta^2 = 0.03$; but the interaction was not significant, $p = 0.60$. Most relevant, higher BDI scores tend to be associated with poorer accuracy and therefore lower earnings.

For purposes of visualization, we partitioned the participants into depressed and non-depressed groups based on BDI scores. Participants with BDI scores greater than nine were classified as depressed, while those with scores of nine or less were classified as non-depressed (Non-depressed: $M = 4.3, SD = 2.9$; Depressed: $M = 16.5, SD = 7.1$).$^2$ The conclusions from an ANOVA on gain scores are (qualitatively) the same under this binary classification of the BDI – except that the BDI by stage interaction comes out significant under the binary partitioning in the Main Round, $p = 0.04$. Using the average gain score across rounds for each subject as the dependent variable, 40% versus 47% of the non-depressed and depressed participants,

$^1$ We use the Greenhouse-Geiser correction to the degrees-of-freedom on all of our repeated measures tests. Hence, according to Huynh and Feldt (1976), our tests may be somewhat conservative.

$^2$ We used a cutoff commonly used in research on depressive realism (e.g. Alloy & Abramson, 1979; Msetfi, Murphy & Simpson, 2007), adjusted for the difference in score between the BDI and BDI-II, as per Dozois, Dobson and Ahnberg (1998).
respectively, scored less than the naïve bookmaker in the Main Round, a significant difference \((z = 2.15, p = 0.03)\); in the Second Round 67% of the non-depressed and 79% of the depressed participants scored less than the naïve bookmaker, also a significant difference \((z = 2.54, p = 0.01)\). Figure 1 displays the gain scores across stages for each group for both the Main Study and the Second Round. It is clear that the participants generally tend to do more poorly when forecasting later stages of the tournament. Most importantly, the depressed tend to do more poorly across stages in both the Main Round and the Second Round.

**Departure from base-rate probabilities.** We have just seen that participants with higher BDI scores tend to predict more poorly and consequently to earn less. Now, we examine whether there are systematic patterns of responding that can help explain this effect. For each participant at each stage, we calculated the difference between average judged probability \((p_s)\) and the base-rate probability \((q_s)\), a measure we refer to as the *average deviation from base-rate* and denote by \(d_s\):

\[
d_s = (p_s - q_s) \times 100.
\]

The average deviation from base-rate in stage \(s\) indicates the extent to which the participant deviated from the naïve bookmaker’s forecasts.

We ran repeated-measures ANCOVAs on the average deviation from base-rate scores using stage (within-subjects) and BDI scores (between-subjects) as independent variables. There is a significant effect of stage, \(F(1.5, 1689) = 135.51, p < 0.001, \eta^2 = 0.11\). The deviation from base-rate was greater for later stages of the tournament. There was a significant effect of BDI score, \(F(1, 1104) = 16.22, p < 0.001, \eta^2 = 0.01\), with higher BDI scores associated with greater deviations from the base-rates. And the BDI by stage interaction also is significant, \(F(1.5, 1689)\)
Depression and Accuracy

\[ F(1.3, 591) = 131.90, \ p < 0.001, \ \eta^2 = 0.22; \] and also of BDI, \( F(1, 459) = 7.40, \ p = 0.007, \ \eta^2 = 0.02. \) The stage by BDI interaction was significant, \( F(1.3, 591) = 12.80, \ p < 0.001, \ \eta^2 = 0.03. \) In short, we find that higher BDI scores are associated with greater departures from the base-rate probabilities.

Partitioning the participants into depressed and non-depressed based on BDI scores (BDI > 9 = depressed; BDI \( \leq \) 9 = non-depressed) makes the nature of the effects quite apparent: For most stages, on average, the participants assign higher than base-rate probabilities (i.e., \( d_s > 0), \) the depressed participants have an even greater departure from the base-rate probabilities, and the difference in the departure from base-rate probabilities between the depressed and non-depressed is greater at lower base-rates (i.e., at later stages of the tournament). See Figure 2.

We tested whether the increased deviation from base-rate probabilities among the depressed mediated the relationship between BDI scores and gain scores in the forecasting task. Recall that in the analyses presented above the gain scores and deviation from base-rate scores were computed separately for each stage of the tournament. Here, to examine the potential mediating influence of the departure from base-rate on the gain scores, for each participant we computed the average gain score and average deviation from base-rate across stages. In both the Main Round and Second Round, all three of Barron and Kenny’s (1986) preconditions for mediation were met. The predictor (BDI) predicts the outcome (average gain score) (Main Round: \( \beta = -0.13, \ p < 0.001, \ R^2 = .02; \) Second Round: \( \beta = -0.13, \ p = 0.004, \ R^2 = .02). \) The predictor (BDI) predicts the mediator (departure from base-rates) (Main Round: \( \beta = .12, \ p < 0.001, \ R^2 = .02). \) The
Depression and Accuracy

0.01; Second Round: β = .13, p = 0.007, \( R^2 = 0.02 \). And when both BDI and departure from base-rates are included, coefficient of BDI drops (Main Round: β = -.05, p = 0.02; Second Round: β = -0.05, p = 0.12) and the coefficient of departure from base-rates remains significant. Therefore, the increased departure from base-rate probabilities partially mediated the effect of BDI scores on the gain scores in the Main Round (Sobel Test Statistic = -3.99, p < 0.001), and fully mediated the effect in the Second Round (Sobel Test Statistic = -2.68, p < 0.01).

**Robustness Checks**

An alternative measure of forecast accuracy. It is easy to show that if we replace the naïve-book maker base-rate probability estimates with estimates that deviate from the base-rate probabilities by a constant \( r_s = q_s + \delta \), where \( \delta \neq 0 \), then

\[
E[\pi(r_s, \cdot)] = (q_s[1 - (1-r_s)^2] + (1-q_s)[1-(r_s)^2]) \times 100 < E[\pi(q_s, \cdot)].
\]

That is, if we have two judges, one who uses the base-rate probabilities and another whose probabilities are displaced from the base-rate probabilities by a constant, the latter will score less under the quadratic scoring rule. (In other words, the left-side of the inequality is maximized when \( r_s = q_s \).) However, the *linear relationships* between the judges’ probabilities and the outcomes would carry the same slope; and both fits would carry the same \( R^2 \). So, in one sense they are equally accurate, even though one earns less under the quadratic rule. This fact demonstrates that assessments of accuracy are contingent – they depend on how one defines accuracy. To check the robustness of the BDI-forecasting relationship reported above, we also computed a linear measure of accuracy. For each participant, we calculated the Pearson product-moment correlation between the participant’s probability judgments and the binary event outcomes (1 if the event occurred; 0 otherwise) (Main Round: \( M = 0.51, SD = 0.11 \); Second Round: \( M = 0.30, SD = 0.18 \)). The judgment-outcome correlation was negatively and
significantly correlated with the BDI in both the Main Round ($r = -0.09, p = 0.002$), and the Second Round ($r = -0.15, p = 0.002$). Higher BDI scores were again associated with poorer accuracy. Further, as a measure of reliability, we computed the correlation between the judgment-outcome correlations in the Main Round and the Second Round, and found that they are significantly correlated ($r = 0.26, p < 0.001$). Our finding that depression is negatively related to forecasting accuracy is robust with respect to our definition of accuracy.

Using the probability-outcome correlation as our measure of accuracy, we tested whether the BDI-accuracy relationship was mediated by the average departure from the base-rate probabilities (i.e., the average $d_s$ taken across stages). Again the three Barron and Kenny (1986) preconditions were satisfied. The predictor (BDI) predicts the outcome (accuracy) (Main Round: $\beta = -0.09, p = 0.002, R^2 = 0.01$; Second Round: $\beta = -0.15, p = 0.002, R^2 = 0.02$). The predictor (BDI) predicts the mediator (departure from base-rates) (Main Round: $\beta = 0.12, p < 0.001, R^2 = .01$; Second Round: $\beta = .13, p = 0.007, R^2 = 0.02$). And when both BDI and departure from base-rates are included, coefficient of BDI drops (Main Round: $\beta = -0.07, p = 0.02$; Second Round: $\beta = -0.12, p = 0.007$) and the coefficient of departure from base-rates remains significant. Therefore, we find that the deviation from base-rates partially mediated the BDI-accuracy relationship in both the Main Round (Sobel Test Statistic = -3.30, $p < 0.001$) and the Second Round (Sobel Test Statistic = -2.23, $p = 0.02$). Hence, again the departure from base rate bias seems to be playing a role in the depression-accuracy relationship: Individuals with higher depression scores tend to show a larger departure from the base-rate probabilities, and this effect drives their poorer forecasting performance.

Discussion
When forecasting accuracy is judged using an ecologically valid task with objective outcomes over which participants have no control, in this case predicting the outcomes of the World Cup, we reach an unambiguous conclusion: depression is associated with less accurate forecasting. This claim holds for the initial set of 100 forecasts (Main Round), and also for a second set of 24 forecasts made two weeks later (Second Round). It stands whether we treat the BDI-II as a continuous measure of depression or use a cutoff score to form depressed and non-depressed subgroups, as many others working on depressive realism have done (e.g., Alloy & Abramson, 1979; Msetfi, et al. 2007). Further, the claim is robust to our definition of accuracy: using both quadratic and linear measures of accuracy, depressed forecasters tended to do worse.

The poorer accuracy among the more depressed forecasters is correlated with – and, in fact, is statistically mediated by – a tendency to assign probabilities that are significantly greater than the base-rate probabilities of advancement for most stages of the World Cup. On average, the depressed subjects tended to assign higher probabilities to most events – and this resulted in poorer accuracy. In fact, the probabilities tended to sum to considerably more than what is logically possible. For instance, in the Main Round the depressed group’s average probabilities for teams winning the World Cup deviated from the base-rate probabilities by around +8%. Hence, their probabilities, if interpreted literally, implied that they believed that 2.5 teams would win the World Cup. Consistent with much previous work on probability estimation, both groups’ estimates were subadditive (e.g., Bearden, Wallsten, & Fox, 2007; Tversky & Koehler, 1994) – but the depressed groups’ were even more so.

One must be cautious in interpreting this result. One inclination might be to conclude that the depressed participants were quite “optimistic” about the teams’ chances of winning. However, that interpretation would seem to rest upon the implicit assumption that higher
probabilities represent, say, “hope” on the part of the forecasters. Less charitably, the understanding would be the result of a linguistic confusion (cf., Wittgenstein, 1953; Hutto, 2009). A less colorful, but descriptively more defensible account, would be that the depressed subjects tended to be less sensitive to base-rate probabilities. Across the stages of the tournament (with the exception of the Round of 16 where the base-rate probability was one-half), the depressed subjects tended to assign probabilities that deviated by a larger margin from the base-rate probabilities. Dunning and Story (1991) also found that depressed subjects were more likely to over-predict low base-rate events. Importantly, though, their subjects were predicting personal, self-relevant events. Here, we find that this pattern of greater base-rate neglect (Kahneman & Tversky, 1973) among depressed people obtained in forecasts for events over which the judges had no control.

We have shown that depressed forecasters performed more poorly in predicting truly uncertain, real-world events: their probabilistic forecasts were in less correspondence with the reality. We hope these findings contribute to our understanding of how affect influences judgment and decision making. And given the importance of probabilistic forecasting and the prevalence of depression in the workforce we hope they may also have practical relevance.
References


Figure Captions

Figure 1: Average Gain Score by tournament stages for depressed and non-depressed in Main Round (top panel) and Second Round (bottom panel). Error bars represent ±1 SEM.

Figure 2: Average Deviation from Base-Rate by tournament stages for depressed and non-depressed in Main Round (top panel) and Second Round (bottom panel). Error bars represent ±1 SEM.
### Table 1

Base rate probabilities and expected payoffs for the naïve bookmaker across tournament stages

<table>
<thead>
<tr>
<th>Tournament Stage</th>
<th>Round of 16</th>
<th>Quarter-Finals</th>
<th>Semi-Finals</th>
<th>Final Match</th>
<th>Win World Cup</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Naïve Base rate ((q_s)) for Main Round</strong></td>
<td>0.500</td>
<td>0.250</td>
<td>0.125</td>
<td>0.061</td>
<td>0.031</td>
</tr>
<tr>
<td><strong>Naïve Bookmaker Expected Payoff for Main Round</strong></td>
<td>75</td>
<td>81</td>
<td>89</td>
<td>94</td>
<td>97</td>
</tr>
<tr>
<td><strong>Naïve Base rate ((q_s)) for Second Round</strong></td>
<td>---</td>
<td>---</td>
<td>0.500</td>
<td>0.250</td>
<td>0.125</td>
</tr>
<tr>
<td><strong>Naïve Bookmaker Expected Payoff for Second Round</strong></td>
<td>---</td>
<td>---</td>
<td>75</td>
<td>81</td>
<td>89</td>
</tr>
</tbody>
</table>
Figure 1

![Graph showing average gain scores for non-depressed and depressed individuals across different rounds of a competition.

- Top graph: Round of 16, Quarter-finals, Semi-finals, Finals, Win.
- Bottom graph: Semi-finals, Finals, Win.

Bars represent average gain scores with error bars indicating variability.](image_url)
Figure 2

![Bar chart showing average deviation from base-rate for non-depressed and depressed groups across different rounds of a competition.](image)

- **Round of 16**
  - Non-Depressed: Average Deviation from Base-rate
  - Depressed: Average Deviation from Base-rate

- **Quarter-finals**
  - Non-Depressed: Average Deviation from Base-rate
  - Depressed: Average Deviation from Base-rate

- **Semi-finals**
  - Non-Depressed: Average Deviation from Base-rate
  - Depressed: Average Deviation from Base-rate

- **Finals**
  - Non-Depressed: Average Deviation from Base-rate
  - Depressed: Average Deviation from Base-rate

- **Win**
  - Non-Depressed: Average Deviation from Base-rate
  - Depressed: Average Deviation from Base-rate
Europe Campus
Boulevard de Constance
77305 Fontainebleau Cedex, France
Tel: +33 (0)1 60 72 40 00
Fax: +33 (0)1 60 74 55 00/01

Asia Campus
1 Ayer Rajah Avenue, Singapore 138676
Tel: +65 67 99 53 88
Fax: +65 67 99 53 99

Abu Dhabi Campus
Muroor Road - Street No 4
P.O. Box 48049
Abu Dhabi, United Arab Emirates
Tel: +971 2 651 5200
Fax: +971 2 443 9461

www.insead.edu