Duttle, Kai; Shichijo, Tatsuhiro

Working Paper

Default or reactance? Identity priming effects on overconfidence in Germany and Japan

Working Papers on East Asian Studies, No. 103/2015

Provided in Cooperation with:
University of Duisburg-Essen, Institute of East Asian Studies IN-EAST

Suggested Citation: Duttle, Kai; Shichijo, Tatsuhiro (2015) : Default or reactance? Identity priming effects on overconfidence in Germany and Japan, Working Papers on East Asian Studies, No. 103/2015, University of Duisburg-Essen, Institute of East Asian Studies (IN-EAST), Duisburg

This Version is available at:
http://hdl.handle.net/10419/121473
Default or Reactance? Identity Priming Effects on Overconfidence in Germany and Japan

KAI DUTTLE
TATSUHIRO SHICHIJO

WORKING PAPERS
ON EAST ASIAN STUDIES
SEPTEMBER 2015
KAI DUTTLE

Doctoral Fellow at the DFG Research Training Group 1613 Risk and East Asia and the Mercator School of Management, University of Duisburg-Essen.

W www.in-east.de/riskandeastasia
E kai.duttle@stud.uni-due.de

TATSUHIRO SHICHIJO

Professor at the Graduate School of Economics, Osaka Prefecture University.

W http://www.eco.osakafu-u.ac.jp/~shichijo/
E shichijo@eco.osakafu-u.ac.jp
# CONTENT

1 Introduction 5  
2 Experimental Design 6  
   2.1 National identity priming 6  
   2.2 Money-incentivized evaluation of performance-based overconfidence 7  
   2.3 Money-incentivized evaluation of calibration-based overconfidence 7  
   2.4 Non-incentivized evaluation of the better-than-average bias 9  
3 Results 9  
   3.1 Better-than-average bias 10  
   3.2 Performance-based overconfidence 10  
   3.3 Calibration-based overconfidence 12  
4 Supplemental Survey Study 13  
5 Discussion 15  
References 16  
Appendix: Experiment Instructions 18
KAI DUTTLE / TATSUHIRO SHICHIJO:

Default or Reactance?

Identity Priming Effects on Overconfidence in Germany and Japan

WORKING PAPERS ON EAST ASIAN STUDIES, NO. 103, DUISBURG 2015

Abstract
This experimental study measures three types of overconfidence in the decision behavior of participants from Germany and Japan. In the first stage of the experiment subjects completed a Raven Progressive Matrices test and subsequently assessed their test performance in absolute and relative terms. During the second stage subjects provided probability forecasts by confidence intervals to artificially generated price paths. Furthermore subjects’ better-than-average bias was assessed during a post-experimental questionnaire. We find that monetary incentives as a reason to honestly self-evaluate reduce cultural differences in overplacement and in overestimation of own performance. Over-precision in probability judgment accuracy on the other hand significantly differed across ethnical groups. To analyze national identity priming effects on overconfidence, a pre-experimental questionnaire made this identity salient to a randomly selected treatment group. Previous studies found that primes of certain identities can trigger behavior that is consistent with the stereotypes associated with that identity, but can also cause psychological reactance leading to counter-stereotypical behavior. We find that in a setting where there are no incentives to provide honest performance self-evaluations the identity prime reinforces behavior consistent with a default strategy which helps adapting to social norms. In incentivized decisions on the other hand the prime causes counter-stereotypical self-perception. Reasons for this phenomenon are discussed.

Keywords
Overconfidence, interval estimates, identity priming, culture, stereotype effect

JEL Classification
C91; G02; F00
1 INTRODUCTION

In the research of psychological judgment biases overconfidence is one of the most robust findings (De Bondt / Thaler 1995) and also finds frequent application in the economics and finance literature (Glaser / Weber 2010). Three different types of overconfidence have been identified (Moore / Healy 2008), and although often used interchangeably, should be carefully distinguished since empirical studies typically find them hardly related (Glaser / Langer / Weber 2013). The first type is overestimation of own absolute performance. The second type is overplacement of own performance in relation to others. The third type is labeled overprecision, which describes an excessive certainty in the precision of own beliefs. In the cross-section of overconfidence studies, this last type appears to be more persistent than and hardly correlated with the former two types.

While overconfidence in self-assessment tasks is a robust finding among subjects from Western countries it is not as commonly observed in East Asian cultures. Japanese for instance are found to be less overconfident in their answers to general knowledge questions than Western subjects (Yates / Lee / Bush 1997), they perceive their performance as close to or even below average (Heine / Lehman 1999) and they are significantly more accurate in their probability judgments than Americans (Yates / Lee / Shinotsuka / Patalano / Sieck 1998). Tsukishima (1977; 1984) argues that this self-effacing tendency of Japanese does not reflect their true inner self-evaluation (Japanese: honne) but is a reflection of self-presentation to preserve a norm abiding façade (tatemae). Later research however revealed such self-effacing behavior also in studies with anonymously administered questionnaires. Yamagishi (2011) and Yamagishi et al. (2008; 2012) therefore propose the concept of a behavioral default strategy that helps adapting to collectivistic social norms. If there is no incentive to deviate from this strategy it is also applied in anonymous situations. Employing participants from the United States, Yamagishi et al. (2012) replicate an overconfidence experiment from Suzuki / Yamagishi (2004) which was conducted with Japanese subjects. The authors show that within an anonymous environment and with monetary incentives to provide an honest self-evaluation in place, cultural differences in relative performance placement indeed become insignificant.

The present study provides two main contributions to the existing literature. On the one hand it extends the work of Yamagishi and colleagues by taking all three different types of overconfidence into account. Overestimation, overplacement and overprecision are measured in a single money-incentivized cross-cultural experiment utilizing a within-subjects design. Additionally, the so-called better-than-average effect is measured in a non-incentivized questionnaire. Thereby subjects evaluate their skills in relation to a hypothetical group of people with similar characteristics as them. The second main contribution is the analysis of national identity priming on overconfidence of German and Japanese subjects. Priming is an implicit memory effect in which exposure to one stimulus influences a response to another stimulus. Previous studies found that primes of certain identities can trigger behavior that is consistent with the stereotypes associated with that identity. Shih et al. (1999) for instance find that compared to a control group, Asian American women perform better in a math test when their cultural identity is primed (as Asians are stereotyped to have superior quanti-
tative skills), but worse if their gender identity is primed (as women are said to have inferior quantitative skills). This effect has been replicated with African Americans stereotyped as poor performers on intellectual tasks (Steele / Aronson 1995), with Whites stereotyped as poor athletes (Stone / Lynch / Sjomeling / Darley 1999), and with Latinos stereotyped as poor students (Aronson / Quinn / Spencer 1998). On the other hand, in some cases identity priming was also found to trigger counter-stereotypical behavior. Zou et al. (2008) for instance find that a shift towards more stereotypical responses with regard to a cultural norm are driven by identification motives with the respective culture, while contrastive responses are driven by disidentification motives. Similarly, Jamieson / Harkins (2012) find that a stereotype threat (compared to a prime) motivates females to perform better on quantitative problems in order to disconfirm the stereotype. Performance-inhibiting consequences of stereotype threat can also be inverted when the threat is subtly reframed as a challenge (Alter / Aronson / Darley / Rodriguez / Ruble 2010). Such a counter-stereotype effect is in line with psychological reactance (Brehm 1966), a motivational state that leads people to acting in the opposite of what they feel they are being cued to do by social influences, and which often occurs preconsciously (Chartrand / Dalton / Fitzsimons 2007).

2 EXPERIMENTAL DESIGN

The experiment was programmed in z-Tree (Fischbacher 2007). It was conducted at ELFE, the Essen Laboratory for Experimental Economics of the University of Duisburg-Essen in Germany, and at KEEL, the Kyoto Experimental Economics Laboratory of Kyoto Sangyo University in Japan. In total 150 student subjects (100 Germans and 50 Japanese) were recruited with ORSEE (online recruitment system for economic experiments; Greiner, 2004) from a pre-selected subject pool that only included students of the respective ethnical groups. This pre-selection considered last names as ethnical identifiers. Male proportions among participants were 50 % in Germany and 68 % in Japan. We conducted six sessions during October 2013 and May 2014. Each session included exactly 25 participants and one session lasted approximately 70 minutes. No feedback concerning performance or payout was given within or between the stages of the experiment to exclude order effects.

In line with common practices in cross-cultural experimental research instructions were back-translated from German to Japanese by two independent native speakers. Furthermore local sessions were run only by the author of respective nationality in order to minimize potential experimenter demand effects. All payouts were labeled with points instead of monetary values, with an exchange rate of 10 points = 1 Euro (150 Yen) that was stated in the instructions at the beginning. The currency exchange rate was based on average wages of student assistants in the two countries. Participants received a show-up fee (5 Euros / 750 Yen) and on top earned 9.92 Euros on average in Germany, 1413 Yen in Japan, with a minimum of 0 Euros (300 Yen) and a maximum of 40 Euros (4350 Yen).

2.1 NATIONAL IDENTITY PRIMING

We use a similar priming technique as in Shih et al. (1999). Stimuli were introduced in a short questionnaire before the experiment started. Approximately 50 percent of subjects (48 Germans, 24 Japanese) were randomly selected to complete a set of five-level Likert scale items which were designed to make their ethnic identity salient. These read as follows. "I see myself as German (Japanese),", "I identify myself with Germany (Japan),", "I am like other Germans (Japanese)." The control group on the other hand was faced with the following set of questions
that were designed to be identity neutral. “I often watch television.”, “I often eat out.”, “I often go to the movies.”

2.2 MONEY-INCENTIVIZED EVALUATION OF PERFORMANCE-BASED OVERCONFIDENCE

The first stage of the experiment measured subjects’ performance self-estimation abilities in absolute and relative terms. Overestimation of one’s actual performance is usually assessed by comparing subjects’ estimates of their own performance in a task to the actual results, where on average they typically provide estimates higher than these values. Overplacement occurs when people perceive their relative performance in a group as better than it actually is. In experiments, overplacement is usually measured by comparing actual subject performance to the ranks, mostly quantile estimates, participants provide assessing their relative performance.

In this experiment subjects’ task performance was measured in a Raven Progressive Matrices test (Raven 1981) which is a so-called culture-free IQ test that does not require subjects to have a common level of education or set of information but only tests the ability to understand logic patterns. The advantage of using this task is that we can compare results from experimental sessions conducted in different countries while avoiding systematic errors in the data driven by diverse prior education and general knowledge. Participants completed ten Raven test items of varying difficulty levels where the order of test items was randomized for each subject but equal over sessions. Each test item consists of a 3 × 3 pattern matrix with one element missing. Subjects were asked to identify the missing element that completes the pattern among eight possible choices. The matrices and response options were shown on an imbedded picture and a time limit of one minute per item was given. Correct answers were rewarded with a payout of 10 points each (which equals 1 Euro or 150 Yen) while both a wrong answer and no answer resulted in zero payout for the respective test item. At the beginning of this stage subjects answered two training test items without time limit in order to get familiar with the task.

After they had completed the test, subjects evaluated their own absolute performance. For this purpose they were instructed to provide the best estimate for their number of correctly answered test items (0 to 10). The correct choice resulted in a payout of 40 points. Subsequently participants estimated their relative performance within the group by choosing from five quintiles, each labeled with a subgroup of 20 percent, in order to place themselves among all subjects in the respective session. The correct quintile choice also resulted in a payout of 40 points.

2.3 MONEY-INCENTIVIZED EVALUATION OF CALIBRATION-BASED OVERCONFIDENCE

The second stage of the experiment measures overprecision which describes the tendency of people to be excessively certain about the accuracy of their beliefs. In experimental studies this type of overconfidence is commonly measured by (usually 90 percent) confidence intervals given to general knowledge questions or forecasts. Typically subjects are overconfident in the sense that they provide too narrow confidence intervals such that the true answer is actually outside of the interval more than 10 percent of times (when using 90 percent confidence intervals). With these methods however it is not possible to distinguish between “true” and “apparent” overconfidence. Apparent overconfidence can occur due to a random error in subjects’ judgment process while in fact judgment is unbiased (Juslin / Winman / Hansson 2007; Soll / Klayman 2004). Only recently two methods were suggested that allow for measuring judgmental overconfidence at the item level rather than at the aggregate of multiple items. Budescu / Johnson (2011) compare subjects’ estimates in binary choice questions to probability estimates obtained from statistical models in order to evaluate the level of in-
individual overprecision. Glaser et al. (2007; 2013) measure subjects’ overprecision in a forecasting task utilizing artificially generated price paths that evolve based on a pre-specified probability distribution. This allows for directly calculating the implied probability mass of interval estimates.

The method we use in this study is based on the experimental design from Glaser et al. (2007, 2013). Subjects’ overprecision is measured in a forecasting task utilizing artificially generated price paths that evolve based on a pre-specified probability distribution. This allows for directly calculating the implied probability mass of interval estimates at the item level rather than at the aggregate of multiple items. The authors show that their measure is significantly correlated with other well established measures of overprecision (general knowledge interval questions, financial time series forecasts). It is independent from differences in the degree of knowledge of subjects, which is an additional advantage in the context of our cross-cultural study, and there exist stable individual differences over time. Thus it measures “true” overprecision that is not the product of a random error in predictions.

Budescu / Du (2007) find a strong correlation between subjects’ probability judgment accuracy and the confidence intervals used in the assessment. Letting subjects forecast future price movements of real stock prices by different confidence intervals the authors find overconfidence for 90%-interval estimates, well-calibrated forecasts for 70%-interval estimates, and even underconfidence for 50%-interval estimates. Subjects in our experiment therefore provided forecasts to a total of three different randomly generated artificial price paths by confidence intervals of 90 percent, 70 percent and 50 percent, respectively. These price paths were different for each subject in a session but paths were equal across sessions. The artificial process generating the price paths in our experiment is based on a simple probability distribution which was also shown to subjects on screen.

Price changes of [-2, -1, 0, 1, 2] are equally likely in each period. Subjects observed a price chart showing price changes over 50 periods and were taught the period 50 price of their individual chart also in text. Then they were asked to forecast the price in period 100 by a given confidence interval where they had to state an upper and a lower bound of the interval. One after another, they provided intervals of 90 percent, 70 percent and 50 percent for three different graphs drawn from a large set of various randomly generated charts. The order of the three requested confidence intervals and that of all graphs were randomly selected for each subject but were equal across sessions. The best-calibrated participant in each forecast thereby earned a bonus of 100 points, respectively.

Fig. 1 Discrete probability distribution and cumulative probability distribution of absolute price changes within 50 periods
In order to evaluate subjects’ performance in this task a score to measure overprecision is necessary. With help of the cumulative probability function of absolute price changes within 50 periods the stated confidence intervals can be translated into probability mass. An overprecision score for (e.g.) the 90 percent interval is defined as \( OC = 90\% - (p_{high\%} - p_{low\%}) \). Positive values correspond to overconfidence, negative values imply underconfidence. In order to calculate a subject’s overprecision score, the provided upper and lower bound of the confidence interval, in terms of absolute changes relative to the period 50 price level, have to be translated into probability mass values \( p_{high} \) and \( p_{low} \). These values can be looked up using the discrete cumulative probability distribution of absolute price changes within 50 periods.

### 2.4 Non-Incentivized Evaluation of the Better-Than-Average Bias

In order to provide a context for our data that was generated using a money-incentivized laboratory experiment, we also assessed the better-than-average bias by means of traditional methods from psychology. Therefore we used the non-incentivized survey method from Lee et al. (1995). In this questionnaire participants evaluated two categories of personal skills in relation to a hypothetical group of 99 fellow students of the same sex who entered university the same year they did. Along with a number of socio-economic questions these assessments were part of the final questionnaire after the experiment. The first out of two questions read as follows.

Imagine a random sample of 100 University of Duisburg-Essen (Kyoto Sangyo University) students the same sex as you and who entered university the same year you did. Assume that you yourself are one of those 100 students. Suppose that all 100 students in the sample are ranked according to mathematical skills. What is your best estimate of the number of students in the sample (0–99) who would be more mathematically skilled than you are?

This first category considered primary skills, defined as ego-engaging and career-significant. In our case these were mathematical abilities. The second category considered secondary skills, defined as ego-engaging and personally significant, such as physical fitness. In the original study Japanese participants displayed much less overconfidence in both categories compared to subjects from the United States.

### 3 Results

Four types of overconfidence are analyzed in this section. Overestimation of absolute performance was determined in the first stage of the experiment when subjects completed ten Raven Progressive Matrices test items. It is defined as the difference between a subject’s estimated number of correct test items and his actual performance in the test. Relative overplacement within the group was determined in the first stage as well. It is defined as the difference between a subject’s estimated relative performance ranking and the actual quintile within the group. Overprecision was determined during the second stage of the experiment. Subjects had to provide different confidence interval forecasts to artificially generated price paths. A subject’s overprecision score is thereby defined as the difference between the requested confidence interval and the size of his stated confidence interval. Moreover, during a post-experimental questionnaire the better-than-average bias was evaluated, which just like overplacement is also a comparative measure of overconfidence. It is defined as the difference between a subject’s percentile rank estimate of his skill and the total average of 50 percent. All measures are designed such that positive values refer to overconfidence and negative values denote underconfidence.
3.1 BETTER-TAN-AVERAGE BIAS

The better-than-average bias was found to be a stable pattern of human behavior in Western societies. Social psychologists however demonstrated that this does not necessarily extend to people from East Asia. Lee et al. (1995) for instance find remarkable differences in the better-than-average bias among participants from Taiwan, Singapore, India, Japan, and the United States. The present study replicates parts of their method and evaluates the better-than-average bias of German and Japanese subjects. In a post-experimental questionnaire participants evaluated two categories of personal skills in relation to a hypothetical group of 99 fellow students of the same sex who entered university the same year they did. No monetary incentives were provided in this part. A thereby measured better-than-average bias is indexed by

\[
\text{better-than-average bias} = \text{percentile rank estimate} - 0.50.
\]

A positive value refers to the number of percentiles the respective subject ranks himself above group average. In total, a positive mean bias in the population indicates overconfidence. Correspondingly, a negative mean bias indicates underconfidence. The responses of our subjects in the control treatment were in line with the results from Lee et al. (1995), as shown in the first part of Tab. 1. Japanese did not display overconfident behavior. Their average bias was below zero for both categories. German participants’ responses were more in line with those of the original study’s subjects from the USA, demonstrating a positive better-than-average bias both for primary and secondary skills.

The second part of Tab. 1 summarizes better-than-average biases in the priming treatment group. Participants in this group were primed such that their respective nationality was made salient during a pre-experiment questionnaire. Previous priming literature demonstrated stereotypical behavior representative of the primed identity to emerge. In line with these findings, Japanese participants who were primed for their ethnical identity engage more heavily in self-effacement compared to the control group. In fact, behavioral tendencies of Japanese and German subjects are qualitatively much more pronounced among members of the national identity primed treatment group, suggesting that there exist opposing stereotypes about the level of self-confidence in both cultures. As a result, compared to the control group, cross-cultural differences in better-than-average assessments are more distinctive in the treatment group and statistically significant for both evaluated skill sets.

### Tab. 1 Mean values of the better-than-average bias (%) of German (GER) and Japanese (JP) subjects

<table>
<thead>
<tr>
<th>Evaluated skill set</th>
<th>Control treatment</th>
<th>Priming treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GER</td>
<td>JP</td>
</tr>
<tr>
<td>Primary: Math</td>
<td>2.40</td>
<td>-0.65</td>
</tr>
<tr>
<td>Secondary: Physical fitness</td>
<td>5.54</td>
<td>-4.35</td>
</tr>
</tbody>
</table>

Info: p-values are obtained from a two-sample Mann-Whitney U test.

3.2 PERFORMANCE-BASED OVERCONFIDENCE

Yamagishi et al. (2012) propose that cross-cultural differences in comparative overconfidence that were found in previous studies by social psychologists are due to a behavioral default strategy of Japanese. This strategy allows adopting to collectivist social rules in their environment. In an experimental study the authors find that given anonymity of choices and monetary incentives to provide honest self-assessments, participants from Japan and those from the United States do not significantly differ in relative self-confidence.
In addition to the better-than-average bias described in the last section, the present study also measured three types of overconfidence that were elicited under same experimental conditions as described in Yamagishi et al. (2012). Tab. 2 shows the results from the first stage of the experiment which measured performance-based overconfidence types, overplacement and overestimation in the Raven test. The first part of the table summarizes the proportions of subjects in the control treatment without national identity priming who were overconfident, underconfident or assessed their performance correctly. Although Japanese were less overconfident than German participants, this distributional difference is statistically not significant. Our results therefore support the hypothesis of Yamagishi and colleagues that the strong self-effacing tendency that Japanese display in non-incentivized surveys is due to a behavioral default strategy and is not a display of honest self-assessment.

Tab. 2 Biases in relative performance placement and absolute self-estimation of Germans (GER) and Japanese (JP) in the Raven Progressive Matrices test

<table>
<thead>
<tr>
<th></th>
<th>Control treatment</th>
<th>Priming treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GER</td>
<td>JP</td>
</tr>
<tr>
<td>Placement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% overconfident</td>
<td>48.1</td>
<td>38.4</td>
</tr>
<tr>
<td>% correct self-assessment</td>
<td>23.1</td>
<td>30.8</td>
</tr>
<tr>
<td>% underconfident</td>
<td>28.8</td>
<td>30.8</td>
</tr>
<tr>
<td>mean</td>
<td>0.42</td>
<td>−0.08</td>
</tr>
<tr>
<td>Self-estimation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>% overconfident</td>
<td>53.8</td>
<td>38.5</td>
</tr>
<tr>
<td>% correct self-assessment</td>
<td>13.5</td>
<td>23.0</td>
</tr>
<tr>
<td>% underconfident</td>
<td>32.7</td>
<td>38.5</td>
</tr>
<tr>
<td>mean</td>
<td>0.56</td>
<td>−0.15</td>
</tr>
</tbody>
</table>

Info: p-values are obtained using a two-sided Pearson χ² test

The second part of Tab. 2 summarizes performance self-estimations of the priming treatment group. Participants in this group were primed such that their respective nationality became salient during a pre-experiment questionnaire. This identity prime worked like an intensifier for the application of a behavioral default strategy in the non-incentivized better-than-average assessment questionnaire. However, in this money-incentivized performance evaluation task we observe the exact opposite. Japanese participants in the priming treatment group demonstrate strong self-enhancement in their performance assessments. This treatment effect is significant for subjects’ relative assessments (p = 0.03, MWU test). Moreover this effect cannot only be observed in the behavior of Japanese subjects. Where German participants in the control group demonstrated a self-enhancing tendency for both types of overconfidence, the primed treatment group displayed much less confidence in their Raven test performance. This treatment effect is significant for absolute assessments (p = 0.03, MWU test).

Tab. 3 Linear regression analyses of overplacement and overestimation in the Raven Progressive Matrices test

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Overplacement</th>
<th>Overestimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japanese subject (dummy)</td>
<td>−0.59 (0.080)</td>
<td>−0.80 (0.086)</td>
</tr>
<tr>
<td>Identity priming</td>
<td>−0.36 (0.188)</td>
<td>−0.77 (0.046)</td>
</tr>
<tr>
<td>Identity priming * Japanese subject</td>
<td>1.36 (0.005)</td>
<td>1.43 (0.033)</td>
</tr>
<tr>
<td>Male subject (dummy)</td>
<td>0.42 (0.071)</td>
<td>0.43 (0.183)</td>
</tr>
<tr>
<td>Adj. R²</td>
<td>0.0481</td>
<td>0.0219</td>
</tr>
</tbody>
</table>

Info: 150 observations, p-values in parentheses
Tab. 3 shows the results of an additional ordinary least squares regression analysis on the two performance-based overconfidence types. The proportion of male subjects in our sample was higher in Japan (68%) than it was in Germany (50%). As men are typically assumed to be more overconfident than women, our analysis also controls for the gender factor. In fact, if we take gender into account, we do find significant cross-cultural differences in overconfidence also within the control group. Partial effects of priming are negative for German subjects. The comparatively much stronger priming effect on Japanese subjects is positive and significant for both performance-based overconfidence types, and it is about double the size in magnitude as the negative culture effect.

The counter-stereotypical self-perception of primed subjects is the result of an incentive to deliberately reflect on the requested performance evaluation. Answers to the non-incentivized better-than-average questionnaire were given more intuitively while unconsciously applying a behavioral default strategy that, in the treatment group, was reinforced by the national identity prime. In the deliberate and conscious evaluation of Raven test performance however the prime was seemingly rather perceived as a stereotype threat triggering reactance. Primed subjects evaluated their performance in the opposite of what they are usually cued to by society.

3.3 CALIBRATION-BASED OVERCONFIDENCE

This section describes the results from the second stage of the experiment which measured calibration-based overconfidence. Overprecision is found to be hardly related to other types of overconfidence (Glaser et al. 2013). Indeed we find overplacement and overestimation highly correlated (0.69 for Germans, \( p < 0.001 \) and 0.75 for Japanese, \( p < 0.001 \)), yet there is no correlation of overprecision with any of the two other types for neither of the requested confidence intervals. Also with respect to the effect of priming national identities on experiment participants overprecision poses a different case. Apparently priming does not have an effect at all here. All p-values obtained from a Mann-Whitney U test comparing populations among treatment and control groups are well above 10 percent. A behavioral default strategy seemingly does not extend to this complex form of overconfidence as there do not exist any relevant social norms. For that reason we neglect the treatment variable for the remainder of this section and analyze the whole set of observations.

**Tab. 4 Median values of overprecision scores for all requested confidence intervals**

<table>
<thead>
<tr>
<th>Requested interval</th>
<th>all observations</th>
<th>GER</th>
<th>JP</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>90 % confidence interval</td>
<td>***7.75</td>
<td>***7.50</td>
<td>***11.35</td>
<td>0.8254</td>
</tr>
<tr>
<td>70 % confidence interval</td>
<td>1.90</td>
<td>2.10</td>
<td>-16.15</td>
<td>0.0305</td>
</tr>
<tr>
<td>50 % confidence interval</td>
<td>-2.60</td>
<td>11.9</td>
<td>***-18.10</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

Notes: 150 observations. \( p \)-values are obtained from a two-sample Mann-Whitney U test. *** marks significant difference from zero on the 1% level, based on a Wilcoxon signed-rank test.

Budescu / Du (2007) find that overprecision of their subjects from the United States decreases with smaller requested confidence intervals. On top of the most commonly requested 90 percent interval our experiment therefore also incorporates 70 percent and 50 percent confidence intervals in this second stage of the experiment. Tab. 4 summarizes median values of overprecision scores for all three confidence intervals along with significance test results. An overprecision score gives the percentage point deviation of a probability forecast interval from the respective confidence interval that was requested. Positive values correspond to overconfidence, negative values imply underconfidence. Budescu / Du (2007) find that overprecision of
their US subjects decreases with smaller requested confidence intervals, participants being overconfident for 90 percent intervals, underconfident for 50 percent intervals, and well-calibrated for 70 percent intervals. In our experiment we find the same pattern in the decisions of Japanese subjects. However, Germans’ probability judgment accuracy seems to be much less affected by a change in requested confidence intervals. Where there are no significant differences in overprecision among ethnical groups for the 90 percent interval estimation, differences are getting highly significant for smaller intervals. Thus we find that sensitivity to a change in requested confidence intervals does differ across countries. Only Japanese behaved according to the pattern observed by Bdescu / Du (2007) while German subjects were statistically well-calibrated for intervals smaller than 90 percent.

**Fig. 2** Box plots of all observations’ overprecision scores by country. This type of box plot reports the median and quartiles of the distribution in form of the box. Whiskers indicate either minimum and maximum values or 1.5 times the interquartile range. In the latter case outliers are reported as dots

Previous studies found Japanese participants to be significantly less overconfident in their probability judgment accuracy than subjects from the USA (Lee et al. 1995; Yates et al. 2010). Overprecision of participants in our study does not differ across cultures for commonly requested confidence intervals of 90 percent. However if we accumulate all observations independent of the requested confidence interval, the cross-cultural difference in median overprecision scores is highly significant (Fisher’s exact test, \(p = 0.004\)). While the median score of 6.2 percentage points is significantly positive for German subjects (one-sided sign test, \(p = 0.01\)), the median Japanese value accounts for −10.0 points (one-sided sign test, \(p = 0.01\)).

### 4 Supplemental Survey Study

Our cross-cultural experimental study of national identity priming effects on overconfidence produced some surprising results. Identity primes intensified the application of culture-specific behavioral default strategies in non-incentivized questions while it triggered reactance behavior when questions were financially incentivized. In order to support these findings and to further analyze the cognitive processes that are driving them, we designed a supplemental questionnaire and conducted an additional survey study with a total of 114 Japanese students at Kansai University during January 2015.

On the first page of the questionnaire sheet subjects answered several questions concerning their nationality. All participants were primed in a similar fashion. Subsequently they assessed their
physical abilities compared to other students of their university by choosing from 10 deciles, labeled as “top 10%” through “bottom 10%”. In addition they were asked how, in their opinion, a typical Japanese student would answer this question. On the second page subjects were explicitly instructed to be careful not to let other students know about their answers. They were asked to assess their hypothetical performance in a newly developed test of cognitive abilities. Instructions stated that they would complete the cognitive abilities test and also their performance self-estimations in a completely anonymous environment, and that they would get a financial bonus if their answer was correct. Again, they were also asked to estimate the self-assessment of a typical Japanese student. The better-than-average biases measured in this questionnaire are indexed by

\[
\text{better-than-average bias} = \text{decile rank estimate} - 5.50.
\]

A positive value refers to the number of deciles a representative subject ranks himself above group average and thus indicates overconfidence. Correspondingly, a negative value indicates underconfidence.

**Fig. 3 Distributions of answers to the survey self-assessment questions. The vertical line marks the 50th percentile rank. Assessments left of the line imply underconfidence, those on the right imply overconfidence**

Figure 3 displays distributions of answers to all four questions of the survey. The left-hand graph shows subjects’ estimates of typical Japanese self-assessments of IQ test performance and physical fitness. Both distributions are highly condensed around the 50th percentile. The majority of subjects expects a typical Japanese student to assess both his test performance and physical fitness around the mean. The right-hand graph shows participants’ own self-assessments. As expected from previous research and also from our own experimental study, the distribution of non-incentivized self-assessments of physical fitness is tilted to the left. Subjects (on average) assess their fitness as below average, and the measured better-than-average bias accounts for a negative −0.45 deciles \((p = 0.08, \text{signed-rank test})\). On the other hand, the distribution of (hypothetically) incentivized self-assessments of performance in the IQ test is tilted to the right. In this question, subjects assess their performance to be above average with a mean better-than-average bias of 0.39 deciles \((p = 0.02, \text{signed-rank test})\). Distributions of both self-assessment types are different from the distributions of estimates for a typical Japanese’ self-evaluation \((p < 0.01 \text{ for IQ test performance, signed-rank test})\).

The results of this supplemental survey study strongly support our hypothesis about cognitive
processes that are driving the observed patterns of subjects’ self-assessments. Subjects expect a typical Japanese student to assess his performance or skill around the average, independent from incentives to honestly self-evaluate. When assessing their own features however, participants show the exact same behavioral pattern as it was found in our experimental study. The national identity prime intensifies the application of a behavioral default strategy prescribing underconfidence in a non-incentivized setting, when answers are given more intuitively. Where an (even hypothetical) incentive induces subjects to make a deliberate and conscious decision however, they evaluate their performance in the opposite of what they are cued to by society.

5 DISCUSSION

Our experimental study measures several types of overconfidence in self-estimations of participants from Germany and Japan. In the first stage of the experiment subjects completed a Raven Progressive Matrices test which is a so-called culture-free test for general intelligence. Subsequently they assessed their test performance in absolute terms as well as in relation to other participants in the respective session. During the second stage subjects provided probability forecasts to artificially generated price paths by three different confidence intervals. In a post-experiment questionnaire subjects further estimated their math skills and physical fitness in relation to a hypothetical group of fellow students at their university.

Previous studies from cultural psychology typically find Japanese subjects to be much less confident in their performance and knowledge than subjects from Western countries. Yamagishi et al. (2012) suggest that this self-effacing tendency of Japanese is a reflection of a behavioral default strategy adapting to collectivistic social norms in their society. It is used to avoid accruing a negative reputation by offending others. The authors find strong support for their hypothesis by introducing monetary incentives as a reason to honestly self-evaluate. Cultural differences in self-enhancement/effacement of US and Japanese participants disappeared in their experimental study of comparative overconfidence. As a first main contribution the results of our study back up and extend the findings of Yamagishi et al. (2012). Monetary incentives as a reason to honestly self-evaluate reduce cultural differences in comparative (overplacement) overconfidence as well as in absolute overestimation of own performance. Overprecision in probability judgment accuracy on the other hand significantly differed across ethnical groups also in our experiment, even when decisions were incentivized and thus payout relevant. We suggest that a behavioral default strategy does not extend to this complex form of overconfidence as there do not exist any relevant social norms.

This study also analyzes national identity priming effects on overconfidence. For that purpose a pre-experimental questionnaire made this identity salient to a randomly selected treatment group, while the control group answered identity neutral questions. Previous studies found that primes of certain identities can trigger behavior that is consistent with the stereotypes associated with the respective identity. In some cases however identity priming was also found to result in psychological reactance leading to counter-stereotypical behavior. The results of our study enable us to make reasonable assumptions about in how far priming leads to either a stereotype effect or a counter-stereotype effect. In a non-incentivized questionnaire measuring better-than-average biases in subjects’ self-estimations of personal skills, national identity priming reinforced behavioral (default) tendencies of the respective culture group. Japanese participants of the primed treatment group engaged in
self-effacing behavior much more heavily, while country primed Germans were even more overconfident than those of the control group. During the main experiment where we introduced a reason to disclose true beliefs by giving monetary incentives however behavioral tendencies of Germans and Japanese were exactly opposite. Japanese members of the primed treatment group were more overconfident than those of the control group, while Germans whose national identity was made salient to them were much less confident in their performance than those of the control treatment. Only overprecision in probability judgment accuracy, considered to be least related to other types of overconfidence, was not affected by our priming treatment. These findings suggest that identity primes intensify the effect of an existing behavioral default strategy in unconscious and intuitive decision making processes. Where conscious and deliberate thinking is applied however, a perceivable prime like it was used in our study may be interpreted as a stereotype threat and result in psychological reactance and counter-stereotypical behavior. The results of a supplemental survey study strongly support this hypothesis.

REFERENCES


References


Thank you very much for participating in this decision study. First of all we want to point out some basic rules in the laboratory. (1) Please do not talk during the experiment. (2) If you have questions, raise your hand and the instructor will come to your place. (3) You may not use the computer for any other purpose.

During the experiment you will have the chance to earn money. Your personal payoff depends on your decisions. Before you make a decision, you will be informed about the amount of points you can earn. 10 points equal 1 Euro. At the end of the experiment you will be informed on screen about your final payout. Please stay seated until your seat number is announced. You will receive your money in cash at the exit. Every decision you make here will be completely anonymous. Neither the instructor nor the fellow participants can monitor your inputs. Instructions to the single parts of the experiment will be shown directly on screen.

**Stage 1:** In this part of the study, you will see one pattern at a time where some element is missing. You will see 8 suggestions to choose from in order to complete the pattern, respectively. Your task thus consists of choosing the suggestion that best completes the pattern. In total you will see 12 patterns. The first two patterns are for training purposes. During the training, you have as much time as you want to make your decision. After that you will have a maximum of 1 minute to solve each of the remaining 10 patterns. When you have made a decision for a pattern, please confirm with the OK button. You will continue with the next pattern and cannot go back. Now you have one minute time to solve the remaining 10 patterns, respectively. Please confirm your choice each time with the OK button. You will receive 10 points for each correct answer. After time has expired you will see the next pattern. Should you not have made any input until then, you receive no points for the pattern.

**Stage 2:** In this part of the experiment you will be asked to make several predictions about the development of a price over time. For this purpose the respective price path over 50 periods is shown in a graph. This price path was artificially generated by a random choice mechanism. The starting price is always 100. In any of the following periods, the price can either remain the same, increase by 1 or 2, or decrease by 1 or 2. Thus in each period, there are 5 options for the price to develop (+2, +1, 0, −1, −2). Each of these options is equally likely. Here you see how such a price path can look like:
Every single price path you will see during the experiment was generated by the same process and was chosen randomly. Your task will be to determine an interval which includes the price in period 100 with a certain probability. This interval consists of an upper bound and a lower bound. Three intervals will be requested:

1. Determine an interval, which includes the true price in period 100 in 9 out of 10 cases (90%).

2. Determine an interval, which includes the true price in period 100 in 7 out of 10 cases (70%).

3. Determine an interval, which includes the true price in period 100 in 1 out of 2 cases (50%).

Your answer will thereby look like this: “I estimate that the true price in period 100 in 9 out of 10 cases will be between ____ (lower bound) and ____ (upper bound).” Consider that a 90% interval has to be larger than for example a 70% interval. You should determine a 50% interval such that, in your opinion, the true price after another 50 periods (in period 100) will be included only in one out of two cases. Assume that the price development after period 50 will be generated by the same random choice mechanism you already know. You can make your decision based on the following information: the type of interval you have to determine (90%, 70% or 50%), the price path until period 50, and the period 50 price level. You will now be asked to make three predictions one after another. Please consider carefully which of the three intervals is requested. For each of the three predictions, the participant who made the most precise predictions earns a bonus of 100 points. Thus you can earn up to 300 points with the following three predictions. Remember: the most precise prediction for e.g. a 90% interval is the one which includes the price in period 100 in exactly 9 out of 10 simulated cases.