On Normativity and Epistemic Intuitions:
Failure of Replication[1]

Hamid Seyedsayamdost
University of London

Abstract

In one of the earlier influential papers in the field of experimental philosophy titled *Normativity and Epistemic Intuitions* published in 2001, Jonathan M. Weinberg, Shaun Nichols and Stephen Stich reported that respondents answered Gettier type questions differently depending on their ethnic background as well as socioeconomic status. There is currently a debate going on, on the significance of the results of Weinberg et al. (2001) and its implications for philosophical methodology in general and epistemology in specific. Despite the debates, however, to our knowledge, there has not been a replication attempt of the experiments of the original paper. We collected data from four different sources (two on-line and two in-person) to replicate the experiments. Despite several different data sets and in various cases larger sample sizes and hence greater power to detect differences, we failed to detect significant differences between the above-mentioned ethnic and socioeconomic groups. Our results suggest that epistemic intuitions are more robust across ethnic and socioeconomic groups than Weinberg et al. (2001) indicates. Given our data, we believe that the notion of differences in epistemic intuitions among different ethnic and socioeconomic groups that follows from Weinberg et al. (2001) needs to be corrected.

In recent years, there has been a lively debate regarding the use of intuitions in accessing and solving philosophical problems. One line of criticism against the use of intuitions has been that these cannot be taken as reliable evidence since intuitions vary widely depending on whose intuitions are probed and under what circumstances these are elicited. For example, data has been published showing that respondents from different ethnic, socioeconomic and linguistic backgrounds offer different intuitions on various philosophical questions. To mention just a few, see (Machery, Mallon, Nichols, & Stich, 2004; J. M. Weinberg, Nichols, & Stich, 2001). Data demonstrating differing intuitions for a range of philosophical problems has also been suggested among female and male respondents. For a review of these results, see (Buckwalter & Stich, forthcoming).

One of the earlier papers that gave rise to these discussions is Weinberg et al. (2001). In it the authors reported that individuals answered Gettier type[2] questions differently depending on their ethnic background as well as socioeconomic status (as measured by an education proxy). This paper has received widespread attention and at the time of writing it has been cited 354 times according to Google Scholar.[3] There is currently an exchange between philosophers on the significance of these findings and its implications for philosophical methodology. For the current discussion, see (Nagel, 2012, 2013; Stich, 2012). Despite these discussions and despite the influence that this paper has had on an entire field, to our knowledge, there has not been a replication attempt of Weinberg et al. (2001) to date, where researchers present participants with identical scenarios as in the original paper in order to test the robustness of the findings. We collected data through four different sources, two on-line and two in-person where we presented

---

[2] With Gettier type or Gettier style scenarios we broadly refer to cases including 'unwarranted' or 'disputed' knowledge including all the cases presented by Weinberg et al. (2001) and discusses in this paper.
[3] http://scholar.google.com/scholar?cites=2305777674912570473&as_sdt=40000005&sciodt=0,22&hl=en

individuals with scenarios identical in wording to those asked by Weinberg et al. (2001). Our results suggest that responses to Gettier style questions are not significantly different among individuals from different ethnic backgrounds or socioeconomic statuses. Given our data, the findings of Weinberg et al. (2001), which have had wide exposure, do not seem to be robust and the notion of differences in intuitions for these cases among different ethnic and socioeconomic groups will likely have to be corrected.

This paper is divided in two parts. Part One examines ethnic differences and Part Two examines differences based on socioeconomic status. In Section 1 of Part 1 we will provide a description of our four data sources and the methods used in data collection. We do this in the first segment so that in the subsequent sections we can compare our data more readily with that of Weinberg et al. (2001) without having to introduce and explain the methods of data collection every time. In Section 2 of Part 1 we will present our results for East Asian and Western participants and compare these to the outcomes of Weinberg et al. (2001). In the third section of the first part we will present the results for South Asian and Western participants and again compare these to the relevant data from Weinberg et al. (2001). Part Two consists of one section only, where we introduce the experimental methods and provide the results for data on socioeconomic status and epistemic intuitions. Following the second part, we will conclude with a short discussion section. Throughout the rest of this paper we will use the terms and abbreviations Western (W), East Asian (EA) and South Asian or Indian Subcontinent (SC), following the terminology in Weinberg et al. (2001) for consistency.

Part One: Ethnicity and Epistemic Intuitions
Section 1: Methods

Data Set 1
Procedure

For this data set we visited undergraduate classes at the London School of Economics (LSE).

Participation was voluntary although no one refused. After a brief introduction, we handed out a

one-page questionnaire. Each student only saw one scenario. We explained that there were

several different questions and that therefore some would complete the questionnaire faster than

others. We did hand out several different questions, however, only one of them was a Gettier

style scenario. In all, the whole process took about five minutes. In order to determine

participants' ethnic backgrounds, we used the relevant questions from Richard Nisbett's

demographic instrument.

Participants
We mainly visited philosophy classes, however, given the size of the philosophy department, we

also visited some classes in the International Relations department to complement the data.

About 13 percent of the data came from non-philosophy classes. We will provide a breakdown

of the numbers in the Appendix.

This data set consisted of 83 Ws, 42 EAs and 34 SCs for a total of 159 participants.[4]

Data Set 2
Procedure

For our second study we used the resources at the London School of Economics' Behavioural

Research Lab (BRL). The BRL compiles a database of individuals interested in participating in

---

[4] We had some concerns about the proportions of the various ethnic groups and thought that this may not be representative of the student body. However, data on the ethnic backgrounds and origins of students made available by the LSE suggests that our worry was unwarranted. For details, see the following documents: http://www2.lse.ac.uk/intranet/LSEServices/planningAndStatistics/pdf/Context_Statistics.pdf and http://www2.lse.ac.uk/intranet/staff/equalityAndDiversity/docs/Equality-data-reporting/2013-Student-numbers.pdf

studies. Participants then receive email notifications whenever studies are being conducted. Participants received 5 pounds sterling to participate in a 30-minute study that consisted of several different tasks including answering questions from a wide variety of different fields in philosophy. Upon arrival participants were given a brief introduction. Then they were brought to a workstation in a computer lab where they started the survey.

Participants

This data set consisted of 64 Ws, 61 EAs, and 60 SCs for a total of 185 participants.

Data Set 3
Procedure

For the third data set we set up questionnaires on SurveyMonkey (SM) that consisted of six questions, four of which were Gettier type questions (the four that are discussed in this paper) and two were Goedel style questions taken from Machery et al. (2004). Participants sign up with SM and receive links to surveys from time to time. For every survey completed, SM donates $0.50 to a charity of the participant's choice. In addition, participants are entered into a draw for a chance to win a $100 gift card of an online retailer.[5] The first page of the survey was a brief introduction giving some background information. This included, for example, that the survey was for an academic study and the approximate time the study would take. After seeing the six questions, participants filled out a demographic questionnaire and finally there was also a text box to leave comments.

Participants

SM collects demographic information on individuals who sign up to participate. We asked SM to send out invitations to people of White/Caucasian background and individuals of Asian

---

[5] For more information, see https://contribute.surveymonkey.com/how-it-works

background. SM does not classify among different regions of Asia, so we used our own demographic questionnaire to filter out East and South Asian participants. The breakdown for our sample was as follows. Seventy-five were classified as Ws, 36 EAs and 12 SCs for a total of 123. Being in SM's White/Caucasian category did not automatically categorize respondents as Westerners. For example, West Asians who were in SM's White/Caucasian category were not classified as Western. We relied on our own questionnaire to categorize participants, however, we used SM's categorization to narrow down the target audience.

Data Set 4
Procedure

The data for this study was collected through Harvard University's Moral Sense Test (MST) website.[6] Participants visited the MST website without being solicited and took part in the surveys that consisted of several different questions from various fields of philosophy. We did not design this survey specifically to test Gettier type scenarios. In fact only one of the scenarios from Weinberg et al. (2001) was included. However, we were interested in demographic differences and hence collected the necessary data for comparison among ethnic groups. Our surveys were limited to eight questions maximum followed by a demographic questionnaire.

Participants

The total number of participants in this data set equaled 239 of whom 198 were Ws, 25 EAs and 16 SCs.

Not all data sets included all of the Gettier type scenarios from the original paper. In the results sections below we will simply present the data sets that included the relevant scenarios.

---

[6] http://moral.wjh.harvard.edu/index2.html

Section 2: East Asians and Westerners

Individualistic Truetemp: W and EA

The first result that Weinberg et al. present is for the Individualistic Truetemp Case. Below is

the wording of the scenario as taken from Weinberg et al. (2001).

> One day Charles is suddenly knocked out by a falling rock, and his brain becomes re–
> wired so that he is always absolutely right whenever he estimates the temperature where
> he is. Charles is completely unaware that his brain has been altered in this way. A few
> weeks later, this brain re–wiring leads him to believe that it is 71 degrees in his room.
> Apart from his estimation, he has no other reasons to think that it is 71 degrees. In fact, it
> is at that time 71 degrees in his room.

Following the scenario, participants saw the below question together with the two answer

choices.

> Does Charles really know that it was 71 degrees in the room, or does he only believe it?
>
> REALLY KNOWS   ONLY BELIEVES

Weinberg et al. conducted a Fisher's Exact test for 214 participants (25 East Asian, 189 Western)

which yielded a p-value of 0.020114. For all of the results discussed in this paper, we will

present charts of the percentages of responses in the body of the paper and leave the number

breakdowns for the Appendix. The chart for the Individualistic Truetemp case as produced by

Weinberg and colleagues is presented below in Figure 1a and for the breakdown of the data see
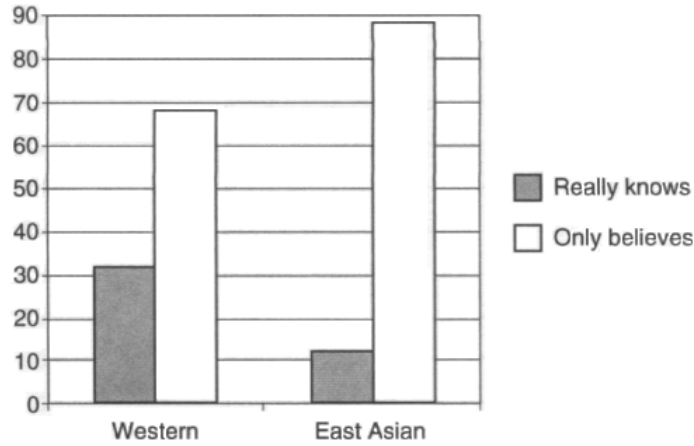
Appendix A.

Figure 1a: Individualistic Truetemp – Percentage of
Responses from Western and East Asian Participants –
Weinberg et al. (2001) Data[7]

The data we collected in Data Set 3 yielded a different picture and in fact the proportion of

responses were almost identical for Ws and EAs (we did not collect data for this scenario in our

first and fourth data sets). We carried out a Chi-Squared[8] test comparing East Asian and

Western responses for our sample of 111 participants (36 East Asian, 75 Western) which yielded

$\chi^2 = 0.029$, p = 0.866 (p-exact = 1.000). See Figure 1b below for percentages of responses and

Appendix A for the number breakdown.

Data Set 2 also did not produce a significant difference between Ws and EAs. We tested N = 59

(30 East Asians, 29 Western) using Chi-Squared and attained $\chi^2 = 0.508$, p = 0.476 (p-exact =

0.532). See Figure 1c for percentages and Appendix A for number of responses.

---

[7] Chart taken from Weinberg et al. (2008).
[8] We will report the outcomes of Chi-Squared tests whenever none of the cells had an expected count of less than
five and provide the values for Fisher's Exact in parenthesis in order to maintain compatibility with Weinberg et
al. (2001). Whenever at least one of the expected counts was less than five, we will report Fisher's Exact only.
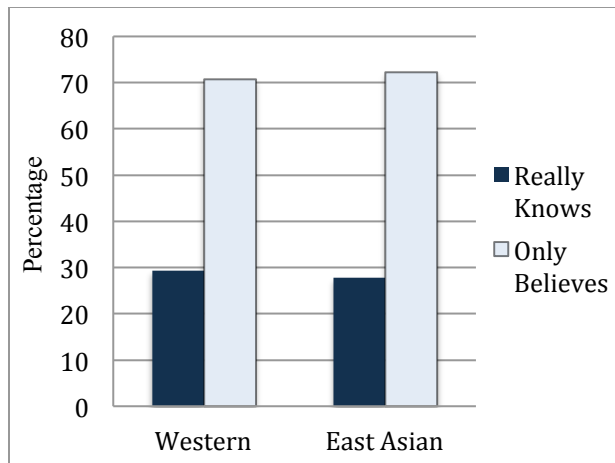
Figure 1b: Individualistic Truetemp – Percentage of Responses from Western and East Asian Participants – Data Set 3
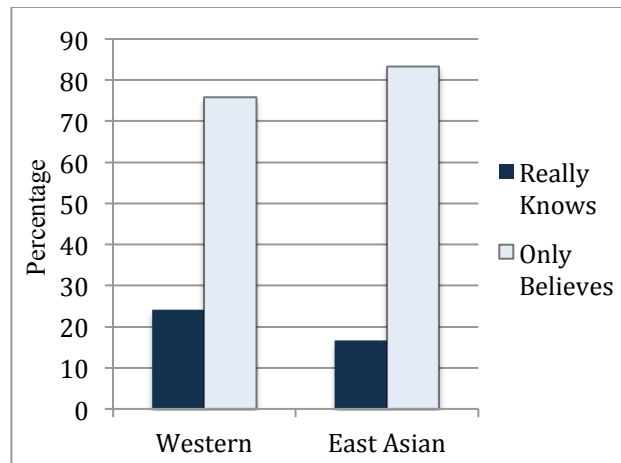
Figure 1c: Individualistic Truetemp – Percentage of Responses from Western and East Asian Participants – Data Set 2

In addition to the Individualistic Truetemp case, Weinberg et al. (2001) also collected data on two variations described as Elders and Community Wide Truetemp scenarios. We did not collect data on these scenarios as Weinberg and colleagues themselves report no significant differences here.

Gettier Car Case: W and EA

The next scenario that Weinberg et al. present is the Gettier Car case for Ws and EAs. The scenario reads as follows.

> Bob has a friend, Jill, who has driven a Buick for many years. Bob therefore thinks that Jill drives an American car. He is not aware, however, that her Buick has recently been stolen, and he is also not aware that Jill has replaced it with a Pontiac, which is a different kind of American car. Does Bob really know that Jill drives an American car, or does he only believe it?

REALLY KNOWS   ONLY BELIEVES

We used the same wording in all of our surveys except for in Data Set 4 where we replaced the names of the cars from Buick and Pontiac to Toyota and Honda, respectively and also changed the origin of the cars from 'American' to 'Japanese', accordingly.

Weinberg et al (2001) collected data for 89 participants (23 East Asians, 66 Western) and report a p-exact value of 0.006414. See Figure 2a below for percentages of responses and Appendix B for the numbers.
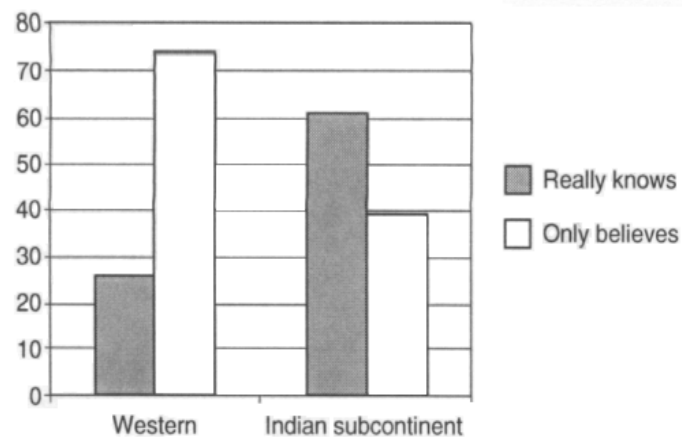

Figure 2a: Gettier Car Case – Percentage of Responses from
Western and East Asian Participants – Weinberg et al. (2001)
Data[9]

Two of the surveys in which we collected data for this scenario yielded a very different picture; in both instances the percentage of East Asians who chose 'Really Knows' was lower than their Western counterparts.

Data Set 1, which was the closest in procedure to the original study consisted of 125 (42 East Asian, 83 Western) participants and a Chi-Squared test between EA and W yielded $\chi2 = 2.557$, p = 0.110 (p-exact = 0.143). See Figure 2b for the chart, Appendix B for the number breakdown and results including data only collected in philosophy classes.

In Data Set 3 the sample totalled 111 (36 East Asians, 75 Western) and a Chi-Squared test

---

[9] Chart taken from Weinberg et al. (2008).

between EA and W yielded $\chi^2 = 0.003$, p = 0.958 (p-exact = 1.000). See Figure 2c for the chart

depicting the percentages of responses and Appendix B for the numbers.

Our sample from Data Set 4 included 223 individuals (25 East Asians, 198 Western) and a

Fisher's Exact test between EAs and Ws yielded p = 0.775 (one cell with excepted count < 5).

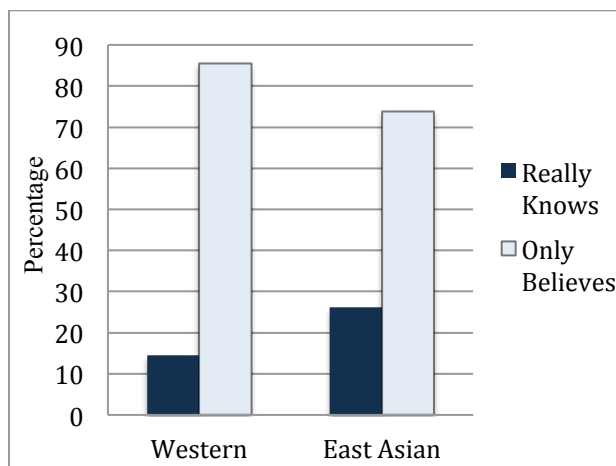See Figure 2d below for a chart depicting percentage of answer choices and Appendix B for the

numbers.

Figure 2b: Gettier Car Case – Percentage of Responses from Western and East Asian Participants – Data Set 1
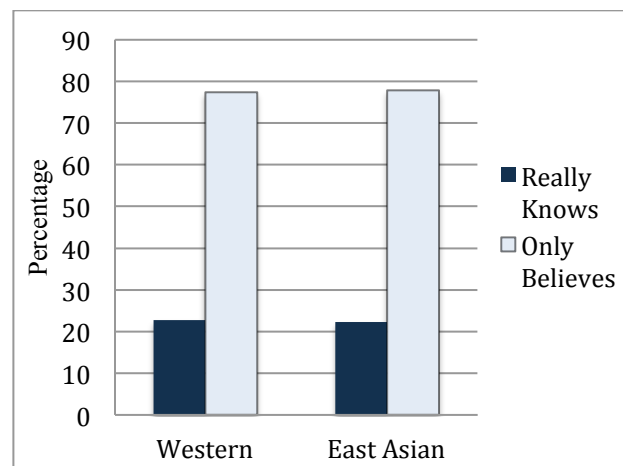
Figure 2c: Gettier Car Case – Percentage of Responses from Western and East Asian Participants – Data Set 3
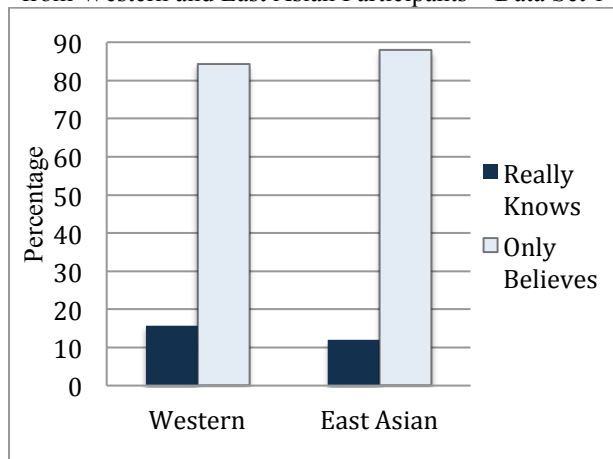
Figure 2d: Gettier Car Case – Percentage of Responses from Western and East Asian Participants – Data Set 4

There are two further scenarios that Weinberg et al. examined. These are termed the Conspiracy Case and the Zebra Case. For the exact wording of these scenarios, see Appendix J. Weinberg et al. did not detect a difference on these cases between EAs and Ws and neither did we. We will present a summary of the outcomes below but will not go further into detail, as there was no disparity between the original and replication studies.

Conspiracy Case (Data Set 3)
N = 111 (36 East Asian, 75 Western); χ^2 = 0.194, p = 0.660 (p-exact = 0.800)

Conspiracy Case (Data Set 2)
N = 66 (31 East Asian, 35 Western); χ^2 = 0.326, p = 0.567 (p-exact = 0.713)

Zebra Case (Data Set 3):
N = 111(36 East Asian, 75 Western); χ^2 = 1.124, p = 0.289 (p-exact = 0.346)

Section 3: South Asians and Westerners

Individualistic Truetemp: SC and W

Weinberg et al. do not report the results for the Individualistic Truetemp case for South Asians and Westerners. We are not sure whether this is because there was no significant difference or whether the necessary data was not available. We assume the latter to be the case as the authors reported negative results for East Asian and Western participants for other scenarios. Again, we did not detect a difference between South Asians and Westerners for this scenario.

We tested a sample of 55 individuals (26 South Asian, 29 Western) and conducted a Chi-Squared test comparing the two ethnic groups, which yielded χ^2 = 0.009, p = 0.926 (p-exact = 1.000). See Figure 3a for the percentages of answers chosen and Appendix C for the numbers.
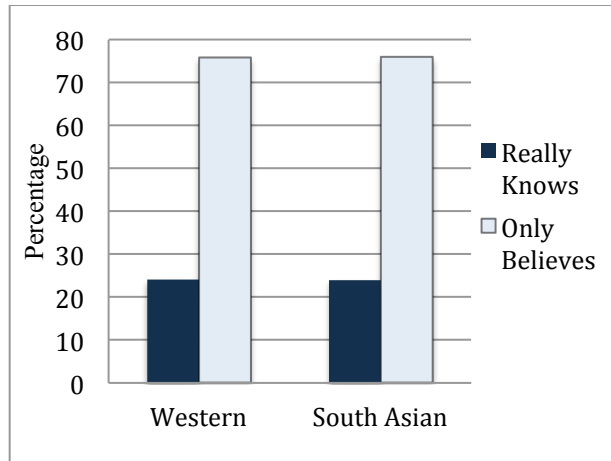
Figure 3a: Individualistic Truetemp – Percentage of
Responses from Western and South Asian Participants –
Data Set 2

Gettier Car Case: SC and W

The outcome of the Gettier Car case Weinberg et al. present for SCs and Ws is similar to the

sample of EAs and Ws. In both cases a larger number of non-Western participants respond that

Bob really knows that Jil drives an American car, whereas this relationship is reversed for

Western participants. Western individuals, as opposed to non-Westerners, according to the

original paper predominantly choose the 'Only believes' answer choice.

In specific, Weinberg et al. tested a sample of 89 individuals (23 South Asian, 66 Western) and

report p-exact equal to 0.002407. See Figure 4a below for a visual presentation of the

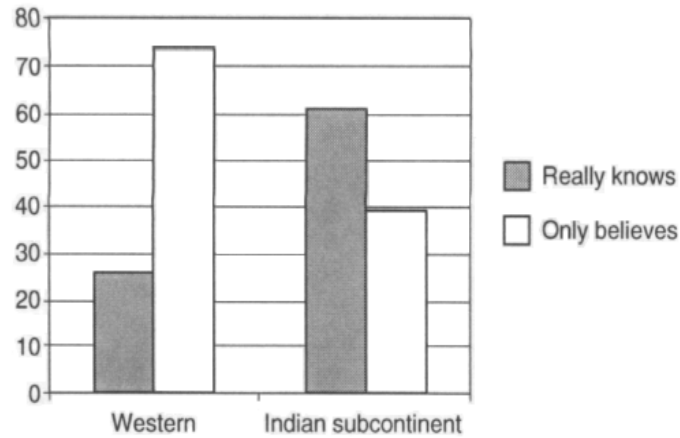percentages and Appendix D for the number breakdown.

Figure 4a: Gettier Car Case – Percentage of Responses from
Western and South Asian Participants – Weinberg et al.
(2001) Data[10]

Data Set 1, where data was collected in classrooms yielded a very different picture. The

percentages of South Asian and Western participants were almost identical. The statistical test

was as follows: N = 117 (34 South Asian, 83 Western); p-exact = 1.000 (one cell with expected

count < 5). See Figure 4b for the percentages, Appendix D for the number breakdown and for

data collected in philosophy classes only.

We had another sample from Data Set 4 and although the sample of South Asians was small, we

will present the results here because there was actually a statistical difference between SCs and

Ws. This data consisted of 214 individuals (16 SCs and 198 Ws). A Fisher's Exact test yielded

p = 0.038 (one cell with expected count < 5). The percentages are depicted in Figure 4c and the

number breakdown in Appendix D. This outcome may not be very meaningful because of the

small sample size of SCs and the large overall sample size. Nevertheless, we include this

outcome for completeness.

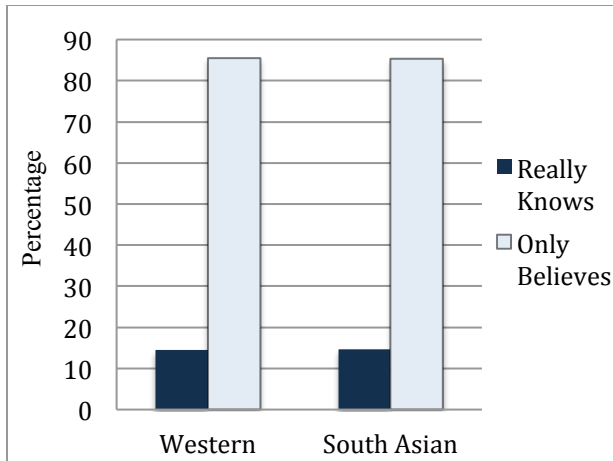[10] Chart taken from Weinberg et al. (2008).

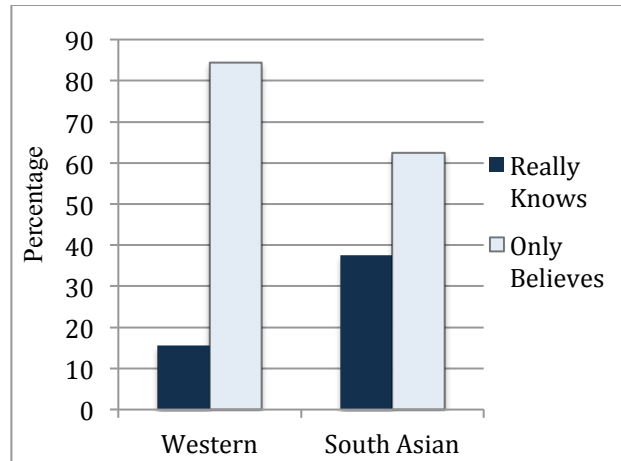Figure 4b: Gettier Car Case – Percentage of Responses from Western and South Asian Participants – Data Set 1

Figure 4c: Gettier Car Case – Percentage of Responses from Western and South Asian Participants – Data Set 4

Conspiracy Case SC and W:

Although Weinberg et al. did not detect a significant difference between East Asians and Westerners for the Conspiracy Case (see Appendix J for the text of this scenario), the authors report a difference on this scenario for South Asians and Westerners.

Weinberg et al. collected a sample of 89 individuals (25 South Asian, 66 Western) and conducted a Fisher's Exact test for which they report a p-value of 0.025014. See Figure 5a for a comparison of the percentages and Appendix E for the numbers.
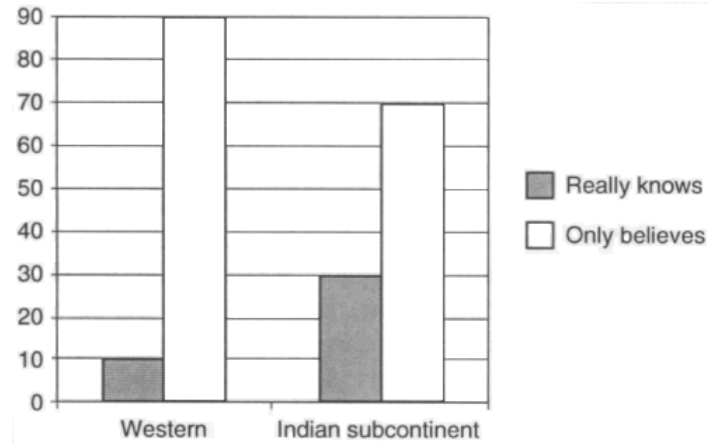
Figure 5a: Conspiracy Case – Percentage of Responses from
Western and South Asian Participants - Weinberg et al. (2001)
Data[11]

For this scenario we had a sample of 69 individuals (34 South Asian, 35 Western) from Data Set 2. We conducted a Fisher's Exact test as two cells had expected counts of less than 5 and the test yielded a p-value of 1.000. For a comparison of the proportion of responses, see Figure 5b below and for the actual numbers, again see Appendix E.
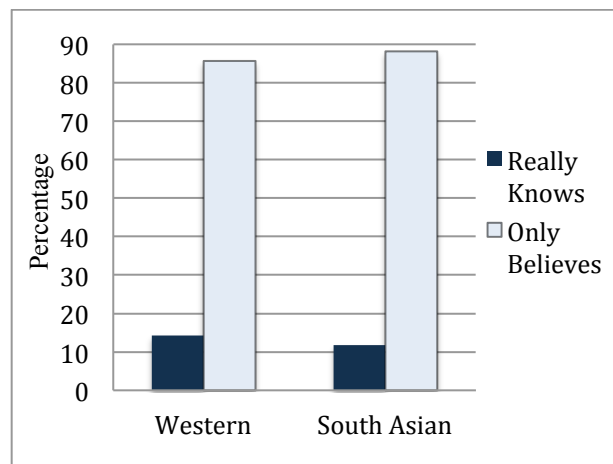

Figure 5b: Conspiracy Case – Percentage of Responses
from Western and South Asian Participants – Data Set 2

Part 2: Socioeconomic Status and Epistemic Intuitions

---

[11] Chart taken from Weinberg et al. (2008).

In their section on socioeconomic backgrounds Weinberg et al. (2001) conclude that

socioeconomic status (SES) has a "major impact on subjects' epistemic intuitions" (p. 453). In

the previous part of this paper we attempted to replicate the results regarding differences among

ethnic groups. Our replication attempts failed. Given the failure of replication for the ethnicity

section of Weinberg et al. (2001), we collected additional data and attempted to replicate the

results on participants from different socioeconomic backgrounds. Here, again, we could not

replicate the results.

At first sight it may not seem too implausible that individuals who go through a similar kind of

educational training, who as a result acquire certain skills and who come from similar

socioeconomic backgrounds would have similar responses to certain scenarios. However, it is

possible that due to the method of data collection used in Weinberg et al. (2001), the authors

were capturing something other than intuitions regarding the scenarios presented to participants.

Weinberg et al. (2001) approached individuals in downtown New Brunswick and offered

McDonald's gift cards in exchange for participation. This kind of setting is not ideal for various

reasons. First, individuals were asked to fill out the questionnaire in a busy environment with

various distractions. Second, when experimenters hand out a survey and wait nearby until

participants complete the questionnaire, participants will inevitably feel a certain obligation to

complete the task as quickly as possible in order not to keep the experimenter waiting. In all,

Weinberg et al. (2001) may have been testing reading comprehension and ability to concentrate

in a busy environment more than anything else.

Apart from these considerations, it is also very well possible that the samples Weinberg and colleagues collected were simply not representative of their respective populations. The samples are relatively small and a replication attempt may be worthwhile in any case.

We would like to point out that after collecting data and informally sharing our results with other researchers, we were informed that Weinberg and colleagues themselves have expressed doubts about their procedures for this section of their paper. However, we could not find a published record for this with the exception of a mention in Beebe (2012). Given the doubts of the authors and yet no correction in a formal manner, we believe that a replication attempt may be useful, regardless of the perceived shortcomings of the original study.

For this part of the paper, we setup two questionnaires on SurveyMonkey (SM) to test the responses of individuals from different socioeconomic backgrounds on the Gettier-type scenarios discussed in Weinberg et al. (2001). Conducting the surveys online may have lessened some of the distractions present in the method Weinberg and colleagues employed.[12]

Methods

Procedure

The SM procedure is the same as described in Part 1 of this paper. Participants sign up with SM and receive links to surveys from time to time. For every survey completed SM donates $0.50 to a charity of the participant's choice. In addition, participants are entered into a draw for a chance to win a $100 gift card of an online retailer.[13]

---

[12] The SM surveys presented here are different from the SM survey of Part One of this paper; that survey was designated Data Set 3.
[13] For more information, see https://contribute.surveymonkey.com/how-it-works

We set up two templates on SM, which we will refer to as Template 1 and Template 2 from here on. The templates were identical with the exception of the order in which the Gettier-style scenarios were presented. Participants first saw a brief introduction stating that we were conducting the questionnaire for an academic research project in the field of philosophy. Next, participants saw the four Gettier-type questions. For the first template the order was Smoking Conspiracy, Zebra, Truetemp and Gettier Car. In the second template the order was Zebra, Gettier Car, Smoking Conspiracy, and Truetemp. After the scenarios, we presented participants with a four-item mood indicator, which was for a different research question. Finally, there was a very short demographic section where we asked about ethnic background and education. SM furthermore provided us data on gender, age range, household income[14] and education. For our data analysis we used data on education that participants submitted in our survey and not the data provided by SM. There was some discrepancy between the two sources, which may be partly explained by the fact that the information is not always up to date with SM and individuals make progress in their educational attainments.

Weinberg et al. (2001) reported significant differences for the Conspiracy and the Zebra Cases (from their paper, it appears that the other two scenarios did not yield a difference, although this is not mentioned explicitly). Hence, we chose the specific orders mentioned above in order to have the Conspiracy Case as the first scenario in Template 1 and the Zebra Case as the first scenario in Template 2.

Participants

---

[14] Data on income was missing for one of the data sets, namely for the low SES data from Template 1.

Weinberg and colleagues used an education proxy to categorize participants as either low or high socioeconomics status. Individuals who indicated that they had never attended college were classified as low SES, whereas participants who indicated that they had taken one or more courses at the college level were classified as high SES.[15]

The survey with the second template was initiated about two weeks after the first survey and we asked SM not to send out invitations to any of the individuals who participated in the first questionnaire. For the first template, we asked SM to restrict participation to individuals who were 24 years of age or older. We were concerned that given the criteria for distinguishing low and high SES by an education proxy we might get many young respondents for the low SES group. After reviewing the data for the first template, we realized that our concern was unfounded and we omitted this requirement for the second template.

For Template 1 our sample consisted of 107 participants (38 low SES, 69 high SES; 54 Male, 52 Female, 1 Missing). For Template 2 our sample consisted of 134 individuals (47 low SES, 87 high SES; 71 Male, 61 Female, 2 Missing).

Scenarios

We used the same wording as in Weinberg et al. (2001) for all the scenarios with the exception of the Car case where we replaced the names of the cars from Buick and Pontiac to Ford and Jeep, respectively, in order to make the scenario more current.[16] The Car and Truetemp cases

---

[15] In order to maintain continuity with the terminology used in Weinberg et al. (20010), we will use the terms low and high SES throughout this paper.

[16] This may not have been a good choice of car brands, as Jeep became the subject of the U.S. presidential campaign, which we were not aware of at the time. There were some campaign ads circulated about Jeep's purchase by Fiat, an Italian company and that production of Jeep vehicles would be outsourced to China. This

can be found in Part 1 of this paper and the Conspiracy and Zebra cases can be found in

Appendix J.

Results

In order to make comparison easier, we will briefly give a summary of the results reported by

Weinberg et al. (2001) before we present the replication outcomes.

Conspiracy Case

Weinberg and colleagues collected 59 responses for the Conspiracy Case of which 24 were

coded as low SES and 35 as high SES. A Fisher's Exact test is reported with a p-value of

0.006778. See Figure 6a below for the chart depicting the percentages of responses as taken

from Weinberg et al. (2001) and for the breakdown of the actual numbers, see Appendix F.

Zebra Case

For the Zebra Case Weinberg et al. collected a sample of 58 individuals of which 24 were

classified as low SES and 34 as high SES. The p-value for Fisher's Exact test reported is

0.038246. See Figure 6b below for a chart representing the percentage of responses and

Appendix F for the numbers.

Truetemp and Gettier Car Cases

As mentioned above, Weinberg et al. do not provide results for the Truetemp and Gettier.

---

topic remained an issue after the elections were over. For further details, see http://www.politifact.com/truth-o-meter/statements/2012/oct/30/mitt-romney/mitt-romney-obama-chrysler-sold-italians-china-ame/ and http://www.politifact.com/truth-o-meter/article/2012/dec/12/lie-year-2012-Romney-Jeeps-China/ .
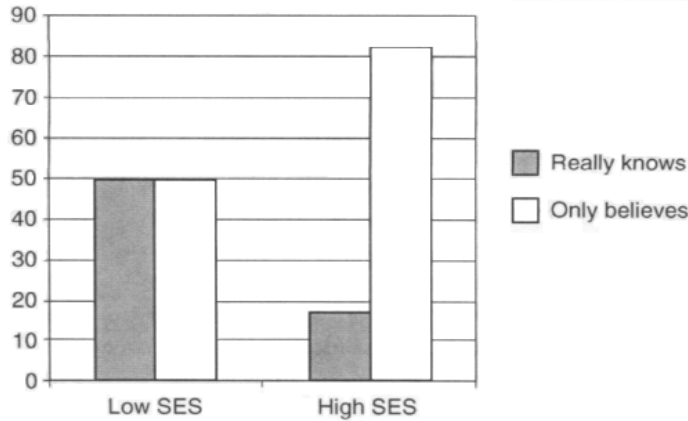
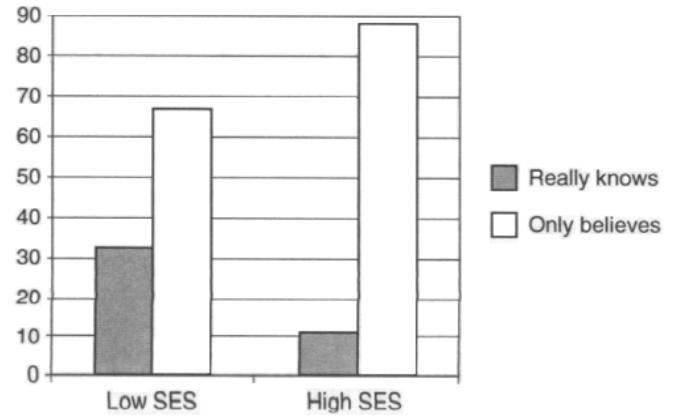Figure 6a: Conspiracy Case – Percentage of Responses from Low and High SES – Weinberg et al.[17]

Figure 6b: Zebra Case – Percentage of Responses from Low and High SES – Weinberg et al.[18]

Results: Template 1

Conspiracy Case

We carried out a Chi-Squared test[19] comparing low and high SES responses for a sample of 107 participants (38 low and 69 high SES) which yielded $\chi^2 = 0.108$, p = 0.743. See Figure 7a below for percentages of responses and Appendix G for the breakdown of the numbers.

Zebra Case

A Chi-Squared test for N = 106 (38 low and 68 high SES) produced $\chi^2 = 0.156$, p = 0.693. See Figure 7b for a graphical depiction and Appendix G for the numbers.

Truetemp

A Chi-Squared test for N = 106 (38 low and 68 high SES) yielded $\chi^2 = 1.947$, p = 0.163. See Figure 7c for the percentages and Appendix G for the numbers.

Gettier Car Case

---

[17] Chart taken from Weinberg et al. (2008).
[18] Chart taken from Weinberg et al. (2008).
[19] For this part of the paper, we will present Chi-Squared tests only as none of the cases had any expected values less than five.

This was the only scenario that produced a significant difference between the two groups. The results are as follows: $N = 106$ (38 low, 68 high SES), $\chi^2 = 6.870$, $p = 0.009$. See Figure 7d for percentages and Appendix G for the number breakdown.
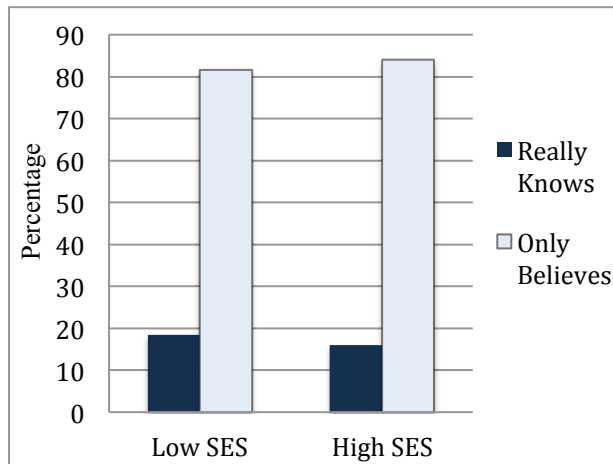

Figure 7a: Conspiracy Case – Percentage of Responses from Low and High SES – Template 1
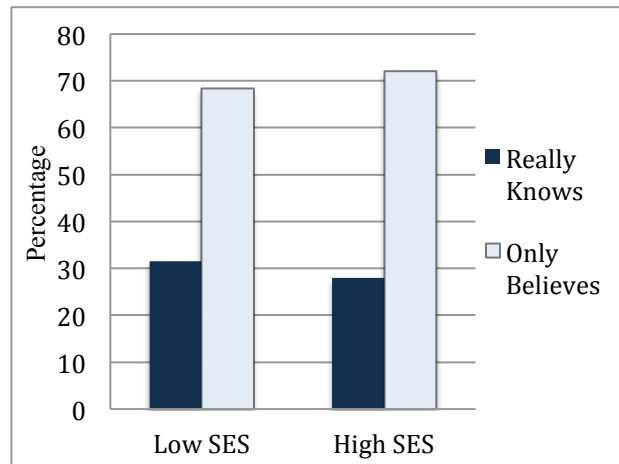

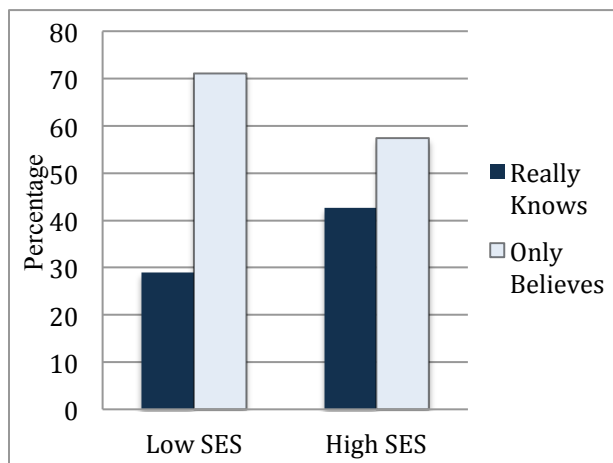Figure 7b: Zebra Case – Percentage of Responses from Low and High SES – Template 1

Figure 7c: Truetemp Case – Percentage of Responses from Low and High SES – Template 1
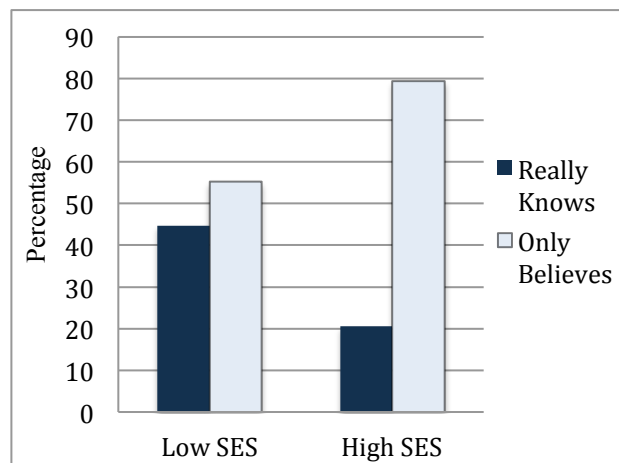

Figure 7d: Gettier Car Case – Percentage of Responses from Low and High SES – Template 1

Results: Template 2

In creating Template 2, we made some changes to the first template. First, we changed the order in which the scenarios were presented. Since, Weinberg et al. in addition to the Conspiracy Case

reported a significant difference for the Zebra scenario, we wanted this case to be placed at the beginning, so we could rule out order effects. Second, given that there was a significant difference for the Gettier Car Case in our first template we wanted to place this scenario further toward the beginning of the survey in order to rule out participation fatigue as the reason for the difference. Below are the results in summary form in the order that the scenarios were presented to participants in Template 2.

Zebra Case

N = 134 (47 low SES, 87 high SES), χ^2 = 0.359, p = 0.549. See Figure 8a and Appendix H for percentages and numbers of responses, respectively.

Gettier Car Case

N = 133 (46 low SES, 87 high SES), χ^2 = 1.913, p = 0.167. See Figure 8b and Appendix H for percentages and numbers of responses, respectively.

Conspiracy Case

N = 132 (45 low SES, 87 high SES), χ^2 = 0.749, p = 0.387. See Figure 8c and Appendix H for percentages and numbers of responses, respectively.

Truetemp Case

N = 132 (45 low SES, 87 high SES), χ^2 = 0.165, p = 0.685. See Figure 8d and Appendix H for percentages and numbers of responses, respectively.
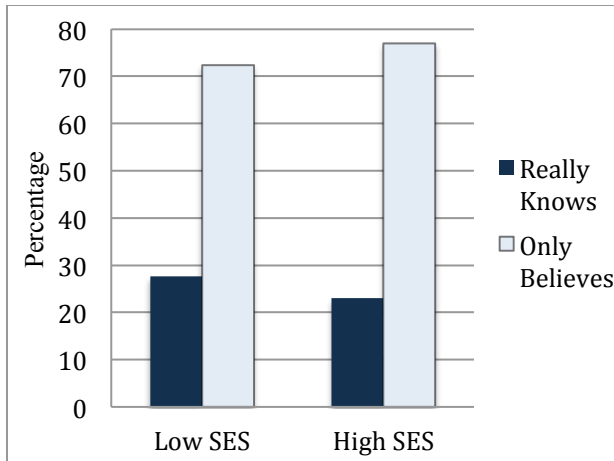
Figure 8a: Zebra Case – Percentage of Responses from Low and High SES – Template 2
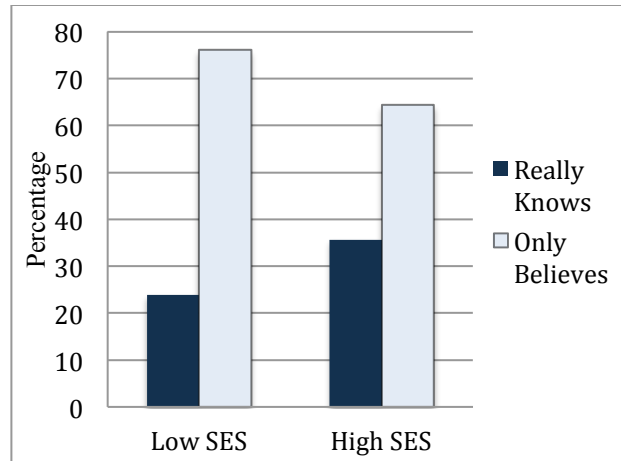

Figure 8b: Gettier Car Case – Percentage of Responses from Low and High SES – Template 2
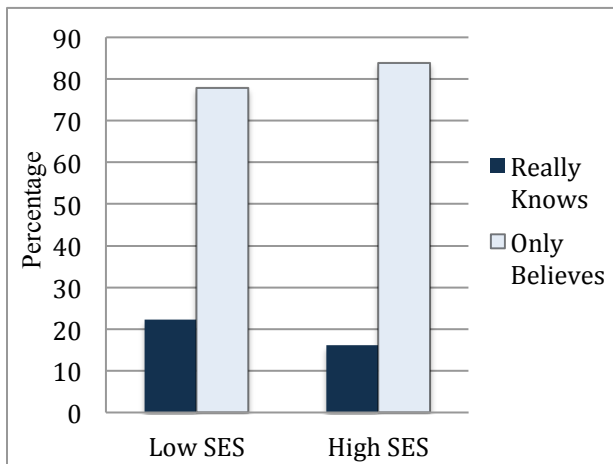

Figure 8c: Conspiracy Case – Percentage of Responses from Low and High SES – Template 2
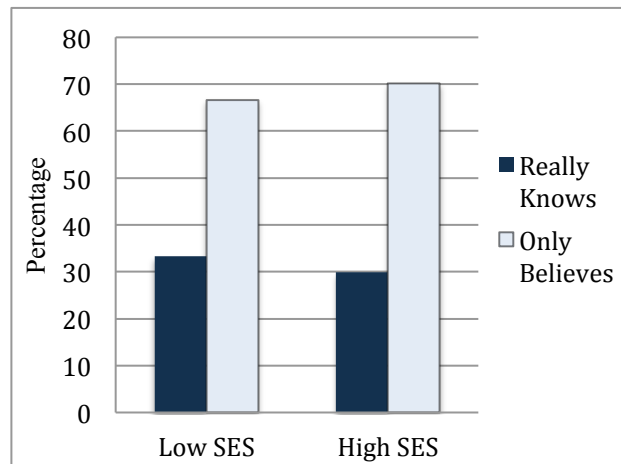
Figure 8d: Truetemp Case – Percentage of Responses from Low and High SES – Template 2

There are few things that may be worth pointing out here. First, the Zebra case again did not yield a significant difference when presented as the first scenario. In fact, this time none of the scenarios yielded a significant difference. The Gettier Car case produced the closest p-value to a significant level, however, this time the direction of the responses was reversed when compared to the first template. That is, this time low SES participants had a lower percentage of 'Really Knows' responses than high SES participants.

Further Analyses

There were various other tests we ran to examine the data. We will not present all the details here, as the details may not be of great interest.

First, we ran an analysis of the combined data from the two templates. Despite the large sample (N = 240), none of the scenarios produced a significant outcome or a p-value close to 0.10. We do not think that this is merely because of cancelling order effects. Rather it seems to be that despite the increased power to detect differences, we still could not find a difference between the two groups. We tested for order effects by comparing the data of the two templates and the only scenario that produced a significant difference was the Gettier Car case.

Next, we wanted to see if there would be a significant difference between the two groups if we made the difference in educational attainment larger. So, for the high SES group we included in our next analysis only participants who had at least completed their Bachelor's degree. Low SES was coded as before. The outcomes for the four scenarios did not change for either one of the templates.

We further ran an analysis excluding participants where the self-reported education level and that provided by SM did not match. None of the outcomes changed. Finally, we ran analyses excluding all participants who fell in the age range 18-29. This made the Gettier Car case for the second template significant (again, in the opposite direction of Template 1) but otherwise all other outcomes remained unchanged.

Discussion and Concluding Remarks

Ethnicity and Epistemic Intuitions

There are two main reasons we see for the different outcomes between the original and the replication study. The first is that the sample sizes for non-Western participants were relatively small in the original study. The second reason for the difference between the original and the replication study may be the difference in data collection. Weinberg and colleagues collected data exclusive in classroom settings at Rutgers University. We collected two of our data sets in-person; one in classroom settings and one in a computer lab. The other two data sets were collected online. There are several reasons why this may make a difference. As Jennifer Nagel has pointed out, the Asian students in the samples of Weinberg et al. (2001) may have had lower levels of motivation or interest (Nagel, 2012). Nagel cites the National Center for Education Statistic, according to whom in 2001 Asian students were more than twice as likely than White students to major in Engineering and Biology (Nagel, 2012). Hence, it is possible that the Asian students captured by Weinberg et al. were more likely to be non-philosophy majors who were taking philosophy classes as electives to fulfill requirements and ultimately less interested. Nagel points out that in the samples where Weinberg et al. detected significant differences, the responses of Asian students were close to 50 percent for each answer choice, which is an indication that the answer choices were selected somewhat randomly.

Unlike Rutgers, the LSE focuses exclusively on the Social Sciences and Humanities and so the pool of students we surveyed in classroom settings may in general have had a different set of interests than the students of the original study and this could have impacted our results. For example, for the Car case highest percentage of 'Really Knows' answers (ca. 25%) for East

Asian participants came from data collected in classrooms. This is considerably lower than the roughly 55% percent reported by Weinberg et al. although still higher than our other samples. Many non-philosophy degree students take philosophy classes at the LSE to fulfill requirements and so the situation may be similar as with the Rutgers samples though less stark because of LSE's focus on Humanities and Social Sciences.

The fact that we could not detect differences for Gettier type scenarios does not imply that individuals from different ethnic backgrounds may not have different intuitions on some other sets of scenarios. As part of our surveys we collected data on some of the questions presented in Machery et al. (2004) on name references and there the data points to a significant difference between Ws and EAs/SCs. Further consideration is needed to pinpoint the elements that make a difference.

Socioeconomic Status and Epistemic Intuitions

With regard to the different outcomes from the original and the replication study on socioeconomic status and epistemic intuitions; again the main reasons for the differences may come down to sample size and method of data collection. We collected data exclusively through online surveys, whereas Weinberg and colleagues collected data in-person. We mentioned in the beginning of Part Two why this may have an impact on responses. However, without having more information, we assume that sample sizes may have been the determining issue here.

Conclusion

The aim of this paper was to test the robustness of the results of Weinberg et al. (2001). Despite collecting data from various sources and attaining larger samples in several of the cases, we

failed to detect differences on epistemic intuitions between participants from different ethnic backgrounds and socioeconomic statuses. With regard to socioeconomic status and epistemic intuitions; we collected data from 241 individuals on four Gettier-style scenarios for which Weinberg et al. (2001) report significant differences between individuals from low and high socioeconomic statuses as classified by an education proxy. We failed to find statistically significant differences. Given this data, we do not believe that socioeconomic status by itself has a major impact on epistemic intuitions for the cases evaluated in this paper. With regard to ethnicity and epistemic intuitions; even though we collected data in several different settings, we could not replicate the results of Weinberg et al. (2001) on differences among individuals from East Asian, South Asian and Western backgrounds. Given this set of data, we do not believe that ethnic background has a significant impact on epistemic intuitions.

As mentioned in the introduction, Weinberg et al. (2001) has been an influential paper, which has received numerous citations. In discussions with other researchers in the field it often appears that it is an established fact that epistemic intuitions differ among ethnic groups. Our data suggests that this conception needs to be corrected. Despite the importance of the implications of the original paper and despite the debate surrounding the findings of Weinberg et al. (2001) for conducting epistemology as well as philosophy in general, to our knowledge, there has not been a replication attempt of Weinberg et al. (2001) to test the robustness of the reported results. We hope to have provided a useful reference point with this paper.

References

Beebe, J.R., (2012). Experimental Epistemology. In A. Cullison (Eds.), *The Continuum Companion to Epistemology* (pp. 248-270). London; New York: Continuum.

Buckwalter, W., & Stich, S. (forthcoming). Gender and philosophical intuition. Available from http://philpapers.org/rec/BUCGAP

Machery, E., Mallon, R., Nichols, S., & Stich, S. P. (2004). Semantics, cross-cultural style. *Cognition, 92*(3), B1-B12.

Nagel, J. (2012). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research, 85*(3), 495-527.

Nagel, J. (2013). Defending the Evidential Value of Epistemic Intuitions: A Reply to Stich. *Philosophy and Phenomenological Research*, n/a-n/a.

Stich, S. (2012). Do Different Groups Have Different Epistemic Intuitions? A Reply to Jennifer Nagel1. *Philosophy and Phenomenological Research*, no-no.

Weinberg, J. M., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics, 29*(1), 429-460.

Weinberg, J. M., Nichols, S., & Stich, S. P. (2008). Normativity and Epistemic Intuitions. In J. Knobe & S. Nichols (Eds.), *Experimental philosophy* (pp. 17-45). Oxford; New York: Oxford University Press.

Appendix A
Individualistic Truetemp EA and W

Data from Weinberg et al. (2001)

	Really Knows	Only Believes
Western	61	128
East Asian	3	22

Reported p-exact = 0.020114

Data Set 3

	Really Knows	Only Believes
Western	22	53
East Asian	10	26

p = 0.866 (Chi-Squared - all cells with expected count >10)

Data Set 2

	Really Knows	Only Believes
Western	7	22
East Asian	5	25

p = 0.476 (Chi-Squared - all cells with expected count > 5)

Appendix B
Gettier Car Case EA and W

Data from Weinberg et al. (2001)

	Really Knows	Only Believes
Western	17	49
East Asian	13	10

Reported p-exact = 0.006414

Data Set 1

	Really Knows	Only Believes
Western	12	71
East Asian	11	31

p = 0.110 (Chi-Squared - all cells with expected count > 5)

Data Set 1: Gettier Car Case – Philosophy Classes Only:
N = 107 (36 EA, 71 W); $\chi^2 = 0.445$, p = 0.505 (p-exact = 0.601)

Data Set 3

	Really Knows	Only Believes
Western	17	58
East Asian	8	28

p = 0.958 (Chi-Squared - all cells with expected count > 5)

Data Set 4

	Really Knows	Only Believes
Western	31	167
East Asian	3	22

p-exact = 0.775 (one cell with excepted count < 5)

Appendix C
Individualistic Truetemp SC and W

Data Set 2

	Really Knows	Only Believes
Western	7	22
South Asian	6	19

p = 0.991 (Chi-Squared - no cells with expected value < 5).

Appendix D
Gettier Car Case SC and W

Data from Weinberg et al. (2001)

	Really Knows	Only Believes
Western	17	49
South Asian	14	9

Reported p-exact = 0.002407

Data Set 1

	Really Knows	Only Believes
Western	12	71
South Asian	5	29

p-exact = 1.000 (one cell with expected value < 5)

Data Set 1: Gettier Car Case – Philosophy Classes Only:
N = 100 (29 SA, 71 W); $\chi^2 = 0694$, p = 0.405 (p-exact = 0.543)

Data Set 4

	Really Knows	Only Believes
Western	31	167
South Asian	6	10

p-exact = 0.038 (one cell with expected count < 5)

Appendix E
Conspiracy Case SC and W

Data from Weinberg et al. (2001)

	Really Knows	Only Believes
Western	7	59
South Asian	7	16

Reported p-exact = 0.025014

Data Set 2

	Really Knows	Only Believes
Western	5	30
South Asian	4	30

p-exact = 1.000

Appendix F
Data from Weinberg et al. (2001)

Conspiracy Case

	Really Knows	Only Believes
Low SES	12	12
High SES	6	29

Reported p-exact = 0.006778

Zebra Case

	Really Knows	Only Believes
Low SES	8	16
High SES	4	30

Reported p-exact = 0.038246

Appendix G
Data from Template 1

Conspiracy Case

	Really Knows	Only Believes
Low SES	7	31
High SES	11	58

p = 0.743 (Chi-Squared - no cells with expected value < 5)

Zebra Case

	Really Knows	Only Believes
Low SES	12	26
High SES	19	49

p = 0.693 (Chi-Squared - no cells with expected value < 5)

Truetemp Case

	Really Knows	Only Believes
Low SES	11	27
High SES	29	39

p = 0.163 (Chi-Squared - no cells with expected value < 5)

Gettier Car Case

	Really Knows	Only Believes
Low SES	17	21
High SES	14	54

p = 0.009 (Chi-Squared - no cells with expected value < 5)

Appendix H
Data from Template 2

Zebra Case

	Really Knows	Only Believes
Low SES	13	34
High SES	20	67

p = 0.549 (Chi-Squared - no cells with expected value < 5)

Gettier Car Case

	Really Knows	Only Believes
Low SES	11	35
High SES	31	56

p = 0.167 (Chi-Squared - no cells with expected value < 5)

Conspiracy Case

	Really Knows	Only Believes
Low SES	10	35
High SES	14	73

p = 0.387 (Chi-Squared - no cells with expected value < 5)

Truetemp Case

	Really Knows	Only Believes
Low SES	15	30
High SES	26	61

p = 0.685 (Chi-Squared - no cells with expected value < 5)

Appendix J

Conspiracy Case

It's clear that smoking cigarettes increases the likelihood of getting cancer. However, there is now a great deal of evidence that just using nicotine by itself without smoking (for instance, by taking a nicotine pill) does not increase the likelihood of getting cancer. Jim knows about this evidence and as a result, he believes that using nicotine does not increase the likelihood of getting cancer. It is possible that the tobacco companies dishonestly made up and publicized this evidence that using nicotine does not increase the likelihood of cancer, and that the evidence is really false and misleading. Now, the tobacco companies did not actually make up this evidence, but Jim is not aware of this fact.   Does Jim really know that using nicotine doesn't increase the likelihood of getting cancer, or does he only believe it?

REALLY KNOWS   ONLY BELIEVES[20]

Zebra Case

Mike is a young man visiting the zoo with his son, and when they come to the zebra cage, Mike points to the animal and says, "that's a zebra." Mike is right — it is a zebra. However, as the older people in his community know, there are lots of ways that people can be tricked into believing things that aren't true. Indeed, the older people in the community know that it's possible that zoo authorities could cleverly disguise mules to look just like zebras, and people viewing the animals would not be able to tell the difference. If the animal that Mike called a zebra had really been such a cleverly painted mule, Mike still would have thought that it was a zebra. Does Mike really know that the animal is a zebra, or does he only believe that it is?

REALLY KNOWS   ONLY BELIEVES[21]

---

[20] Taken from Weinberg et al. (2001).
[21] Taken from Weinberg et al. (2001).