From Gut Feelings to Considered Forecasts: The Effects of Incentivizing and Unpacking the Forecasting Question

David A. Comerford

Abstract: Expectations are, along with preferences, a core driver of behaviour. The standard way to elicit forecasts in surveys is using a direct question format that asks whether a variable will have increased or decreased between time 1 and time 2. This paper compares direct format forecasts against forecasts from an identical but incentivized condition and from an indirect format condition. The indirect format condition asks respondents to estimate the unemployment rate at time 1 and to then to estimate the unemployment rate at time 2; the direction of change can then be inferred from the two estimates. In each of three studies, with cumulatively over 1,000 respondents, accuracy incentives increased the likelihood of correct forecasts. Study 2 provides evidence for mechanism: incentives reduce the role of affect on direct format forecasts. Studies 1 and 3 show that the indirect format yields forecasts that are more similar to incentivized forecasts than to unincentivized direct format forecasts. Study 3 also demonstrates that incentivized and indirect forecasts better explain car purchase behavior than unincentivized direct format forecasts. I conclude that the indirect format offers promise as a means to debias forecasts.

Keywords: Subjective Expectations, Forecasting, Incentives, Attribute Substitution.

JEL codes: C80 Data Collection and Data Estimation Methodology; D84 Expectations

---

1 Economics Division and Behavioural Science Centre, University of Stirling, FK9 4LA. This work has profited greatly from presentations at the SJDM conference 2016, BDRM conference 2018 and from conversations with Jack Soll, Rick Larrick, Liam Delaney and Richard Taffler.
From Gut Feelings to Considered Forecasts: The Effects of Incentivizing and Unpacking the Forecasting Question

1. Introduction

Expectations are, along with preferences, a core driver of behaviour. To the extent that we wish to predict behaviour, it is important that we accurately measure expectations (Manski, 2004; Delavande, Giné and McKenzie, 2011). A further advantage of accurate expectations data is that they can contain private information that, when aggregated, forecasts the macroeconomy. The expectations data contained in consumer sentiment surveys have been shown to forecast consumer spending after controlling for contemporaneous macroeconomic indicators (Ludvigson, 2004). Reflecting this forecasting power, subscribers pay for early access to the data from the Michigan Survey of Consumer Sentiment and trading volumes increase within milliseconds of that data being released (Javers, 2013). Yet, survey measures of expectations also demonstrate systematic errors. Forecast errors vary reliably across the business cycle in line with an extrapolation bias; expectations are generally overoptimistic in booms and too pessimistic in busts (Souleles, 2004; Barberis, Greenwood, Jin, and Shleifer. 2015). Additionally, consumer sentiment data consistently manifests partisan bias; supporters of a given political party are more optimistic in both their forecasts and backcasts of the economy when their preferred party holds office (Conover, Feldman and Knight, 1987; Bartels, 2002).

The current research seeks to improve the forecasting power of survey measures of expectations. The standard way to elicit forecasts in surveys is using a direct question format that asks whether a variable will have increased or decreased between time 1 and time 2. This paper compares forecasts of unemployment elicited by this direct format against forecasts from an identical but
incentivized condition and from an indirect format condition, which unpacks the forecasting question into its constituent questions.

Across three studies, I demonstrate that accuracy incentives systematically alter forecasts and that incentivized forecasts better explain behavior than do unincentivized forecasts. One contribution of the current research then is to demonstrate that the standard question format that is currently used to elicit forecasts elicits biased measures of expectation.

Much of what follows is concerned with explaining the bias that is moderated by incentives. There are two candidates: deliberate misreporting and cognitive processing. Whereas deliberate misreporting will trouble only survey responses, bias induced by cognitive processing can contaminate incentive-compatible behavior. The second contribution of the paper is that I present process evidence that in the absence of incentives, respondents engage a heuristic termed attribute substitution. Attribute substitution occurs when respondents substitute an easy to answer question for the cognitively taxing question they have been asked (Kahneman and Frederick, 2002). In this case, forecasters answer the question “will unemployment be higher or lower at t2 than at t1?” as though asked “do I feel good or bad about the economy right now?”

The insight that incentives shift the process of expectation formation from a fast-and-frugal System 1 approach to a more deliberative System 2 approach (terminology taken from Kahneman, 2011) has immediately practicable implications for survey measures of expectations. It implies that forecasts will contain greater signal and less noise if elicited in a manner that obviates the need for respondents to engage in attribute substitution. Attribute substitution arises because the question that has been asked is complex. Consider the question “will unemployment be higher or lower at t2 than at t1?” Comerford and Soll (in prep) term this a direct format
question because it directly asks for a judgment of trend. A normative response to a direct format question requires respondents to engage in three distinct steps: 1. estimate unemployment at $t_1$, 2. estimate unemployment at $t_2$, 3. compare the estimates. In Studies 1 and 3 I compare forecasts elicited by this complex direct format with those elicited from the logically equivalent, but less complex indirect format. It explicitly asks respondents to estimate unemployment at $t_1$ and then, in a separate question, asks them to estimate unemployment at $t_2$. The third contribution of the current research is to demonstrate that the indirect format has similar effects on forecasts as do accuracy incentives. I conclude that the indirect format has potential to improve forecasts derived from surveys.

2. Study 1

2.1. Methods

I chose to elicit forecasts of trends reported in the Bureau of Labor Statistics' unemployment measure. I did so because that measure is unidimensional and objectively verifiable. Whereas experts might plausibly disagree on whether the economy has improved or worsened, there can be no dispute regarding the question of whether the Bureau of Labor Statistics' data indicates an increase or a decrease in unemployment. A further advantage of taking unemployment as the focal outcome is that a layperson could reasonably be expected to apprehend trends in unemployment. In the course of everyday life people are exposed to job postings, new hires at work and to acquaintances who have recently found work. Unemployment fell from 5.7 percent to 5 percent over the period respondents were asked about and so these real, tangible consequences of falling unemployment would have been there for respondents to observe. Finally, unemployment has the advantage that it lends itself to forecasting by laypeople. It is
likely that some respondents had private information that is diagnostic for forecasting e.g. news that their firm would be investing in a new facility; sites in their neighborhood being cleared for construction etc.

Between November 26th and December 3rd 2015, 301 respondents were recruited through Amazon Mechanical Turk. Of these, 101 were assigned to the direct format, 101 were assigned to an incentive-compatible direct format condition and 99 were assigned to the indirect format condition. The characteristics of respondents are summarized in Table 1. No characteristics differed significantly across survey conditions.

Table 1: Characteristics of respondents by survey condition

<table>
<thead>
<tr>
<th></th>
<th>Std. Direct, n = 101</th>
<th>Incentivized, n = 101</th>
<th>Indirect, n = 99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age</td>
<td>37</td>
<td>37</td>
<td>35</td>
</tr>
<tr>
<td>% Female</td>
<td>52</td>
<td>57</td>
<td>56</td>
</tr>
<tr>
<td>% Republican</td>
<td>27</td>
<td>28</td>
<td>19</td>
</tr>
<tr>
<td>% Democrat</td>
<td>50</td>
<td>51</td>
<td>59</td>
</tr>
</tbody>
</table>

The forecasting task was introduced as follows:

On December 4th, unemployment figures will be released up to November 30th which will allow determination of the change in unemployment over the course of the year to date.

The direct format question was worded as follows:

Do you think the percentage of the US labor force that is unemployed has increased, decreased, or stayed the same since January 1 2015?

The Indirect format was worded as follows:
On January 1st 2015, what percentage of the US labor force do you estimate was unemployed? (just enter a number without a percentage sign)

On November 30th 2015, what percentage of the US labor force do you estimate will be unemployed? (just enter a number without a percentage sign)

The incentivized procedure was identical to the direct format procedure but it prefaced the question with the following text:

If your answer to the following question correctly predicts the trend over 2015 that is revealed on December 4th, you will receive double the usual payment for your response to this survey. That bonus payment will be made on December 5th.

2.2. Results

Figure 1 depicts, for each survey condition, the percentage respondents indicating that employment would increase, decrease and remain unchanged. It documents that respondents in the direct format condition were most likely to expect increasing unemployment and least likely to expect decreasing unemployment. To analyze these results statistically, I constructed a binary dependent variable that is coded for whether the forecast correctly predicted the true direction of change in unemployment.

I first analyse the effect of the incentive by comparing incentivized against standard direct format forecasts. A binary logistic regression finds that the incentivized condition was more likely to deliver a correct forecast than the standard direct format condition (Table 2, model 1: \(n = 202\), \(OR = 3.24, z = 3.93, p < .001\)).
I next compare the standard direct format forecasts and indirect format forecasts. A binary logistic regression finds that the indirect format condition was more likely to deliver a correct forecast than the standard direct format condition (Table 2, model 2: \( n = 200, \text{OR} = 2.09, z = 2.55, p = .011 \)).

Lastly I compare the incentivized and indirect format conditions and find that the indirect format condition was no less likely than the incentivized condition to deliver the correct forecast (Table 2, model 3: \( n = 200, \text{OR} = .64, z = 1.45, p = .146 \)).

**Figure 1: Percentage Forecasting Increased, Decreased or Unchanged Unemployment for the US during 2015 (Study 1)**

![Figure 1: Percentage Forecasting Increased, Decreased or Unchanged Unemployment for the US during 2015 (Study 1)](image)

### 2.3. Mechanisms

This section considers the question of whether cheerleading or heuristics explain the difference across the standard direct format and the incentivized procedure. At the time of this study, there
was a Democrat in the White House. If respondents engage in deliberate cheerleading then Democrats would be expected to deliver forecasts that make the governing Democrat administration look good i.e. to forecast a fall in unemployment. A fall in unemployment happens to be the true trend also and so accuracy incentives would be expected to have a null effect or possibly even a negative effect on the forecasts of cheerleading Democrats. If cheerleading is the only source of bias then there is no circumstance in which accuracy incentives would cause Democrats to become more positive i.e. more correct.

For this analysis, I exclude the indirect format condition from this analysis and restrict our sample to just Republicans \((n = 54)\) and Democrats \((n = 102)\). In the absence of cheerleading, accuracy incentives would cause an increase the proportion of Democrats correctly indicating a decrease in unemployment. This is what is found (model 5: \(n = 102, \ OR = 3.18, \ z = 2.61, \ p = .009\)).

Further evidence against the deliberate partisan cheerleading story is that the effect of incentives is virtually identical for Republicans as for Democrats (model 4: \(n = 55, \ OR = 3.09, \ z = 2.01, \ p = .045\)).

### Table 2: Logistic Regression Models of Correctly Forecasting Falling Unemployment

<table>
<thead>
<tr>
<th></th>
<th>Std. Direct vs. Incentivized</th>
<th>Std. Direct vs. Indirect</th>
<th>Incentivized vs. Indirect</th>
<th>Std. Direct vs. pooled others Republicans only</th>
<th>Std. Direct vs. pooled others Democrats only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incentivized</td>
<td>1.177</td>
<td>1.155</td>
<td>1.128</td>
<td>1.155 (0.443)**</td>
<td>1.128 (0.562)*</td>
</tr>
<tr>
<td></td>
<td>(0.299)**</td>
<td>(0.443)**</td>
<td>(0.562)*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indirect format</td>
<td>0.735</td>
<td>-0.442</td>
<td>-0.219</td>
<td>-0.219 (0.289)*</td>
<td>0.160 (0.284)</td>
</tr>
<tr>
<td></td>
<td>(0.289)*</td>
<td>(0.304)</td>
<td>(0.200)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.219</td>
<td>0.958</td>
<td>-0.693</td>
<td>0.160</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.200)</td>
<td>(0.222)**</td>
<td>(0.408)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>202</td>
<td>200</td>
<td>55</td>
<td>102</td>
<td></td>
</tr>
</tbody>
</table>

Notes: All models passed the proportional odds assumption.* \(p<0.05\); ** \(p<0.01\)
The alternative to the cheerleading story is that the effect of incentives was to motivate
respondents to expend greater effort on forecasting than they would have in the absence of
incentives. I would then expect to see that incentivized respondents spent longer considering
their response than did respondents to the standard direct format condition. To analyze response
time I trim the distribution, excluding the most rapid five percent of respondents and the slowest
five percent. I also correct for skew in the response time data through a Box-Cox procedure. A
bivariate OLS regression of the resultant variable finds that respondents in the incentive
condition took significantly longer to respond than did those in the standard direct format
condition \((n = 187, t = 2.90, p = .004)\).

I asked respondents to report the direction of change in unemployment from a point in the past
(January 2015) to a point in the future (December 2015). It is possible that an effect of incentives
was to prompt our online respondents to search the web for official statistics on unemployment
in January 2015. If this was the case, then incentives clearly prompted retrieval of more
diagnostic information than did the standard question but we would be worried about the
generalizability of these results – typically economists want to elicit expectations over outcomes
for which published data is unavailable. I cannot observe whether our respondents in the direct
format condition searched for official statistics. I can however observe whether this was true of
respondents in the indirect format condition. Recall that the indirect format asked respondents to
report the unemployment rate in January 2015. If respondents searched for that data online then
they would have found a figure of 5.7\% (Bureau of Labor Statistics, 2015). With randomization,
there is no reason to expect that respondents to the indirect format condition of the survey were
any more motivated to search the web for statistics than were those in the direct format
condition. Still, it may be that the indirect format made more obvious to respondents that this would be an appropriate strategy and so prompted search for that information. If online searching does explain the discrepancy across the indirect format and the standard direct format then I should see that discrepancy disappear when I drop from the sample those respondents who parroted official statistics. In fact, there was only one respondent who reported that the unemployment rate was 5.7% in January 2015. Dropping that response from the sample leaves the direct v. indirect format results described in model 2 of Figure 2 substantively unchanged \((n = 199, \text{OR} = 2.05, z = 2.49, p = .013)\). I conclude that the difference across the standard direct format procedure and the other two procedures is not explained by respondents searching the web for official statistics.

Having observed that the directional effect of incentives is not consistent with the prediction offered by cheerleading, my preferred explanation for the effect of incentives is that they shift respondents from System 1 processing to a System 2 judgment process. This accounts for the similarity of incentivized forecasts and indirect format forecasts; the indirect format imposes on respondents the normative procedure for making forecasts. Even if people are poorly-informed regarding the topic at hand, the procedural guidance offered by the indirect format might lead them towards a sensible forecast. The indirect format delivers a measure of how well informed respondents were on the subject of unemployment: I can compare their reports of the unemployment rate in January against the 5.7% figure reported by the Bureau of Labor Statistics. The mean estimate for unemployment was 14.6 percent. The median estimate was 10 percent. Only a quarter of respondents offered an estimate that was within 2 percentage points of the official statistic. Even if I drop that most informed quarter of the indirect format sample from the data, I still find that the remaining respondents were significantly more likely than respondents to
the unincentivized direct format to correctly forecast decreasing unemployment ($n = 175$, OR = 1.93, $z = 2.12$, $p = .034$).

3. **Study 2**

The results of Study 1 are consistent with the theory that respondents to the standard direct format question engaged in System 1 processing when forming their forecasts, whereas incentivized respondents engaged in more deliberative System 2 processing. Study 2 tests for use of a specific heuristic in direct forecasts: the prediction is that when asked whether unemployment will go up or down, respondents answer with reference to their feelings about the current economy (Comerford and Soll, 2019). I hypothesize that incentives reduce the weight that forecasts place on feelings about the current economy.

3.1. **Methods**

I again elicit forecasts of trends reported in the Bureau of Labor Statistics’ unemployment measure. Between September 5th and 6th 2018, the survey was posted on Amazon Mechanical Turk and was described as eliciting “views on life in the US”. The opening questions asked respondents for their gender, age, country of origin and, as a final question, an attention filter that told respondents to click both “sometimes” and “never” to proceed to the survey (Oppenheimer, Meyvis and Davidenko, 2013). 444 participants were recruited. Of these, 19 were rejected for reporting that their country of residence was somewhere other than the United States and a further 8 failed the attention filter. 209 were assigned to the direct format and 208 were assigned to an incentive-compatible direct format condition.

The question wording was identical across conditions:
Do you think the Bureau of Labor Statistics data will show that the unemployment rate has increased, decreased, or stayed the same between January 1st 2018 and August 31st 2018?

As in Study 1, the incentivized procedure differed only insofar as it included the following text:

If your answer to the following question correctly predicts the trend over 2018 that is revealed on September 7th, you will receive double the usual payment for your response to this survey. That bonus payment will be made on September 8th.

<table>
<thead>
<tr>
<th>Table 3: Characteristics of respondents by survey condition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Std. Direct Format, n = 209</strong></td>
</tr>
<tr>
<td>Mean Age</td>
</tr>
<tr>
<td>% Female</td>
</tr>
<tr>
<td>% Republican</td>
</tr>
<tr>
<td>% Democrat</td>
</tr>
</tbody>
</table>

* p<0.05; ** p<0.01

After forecasting, respondents were asked: “How do you feel about the US economy right now?” and answers were made on a seven-point scale from “very negative”, through “neutral” to “very positive”. On the same screen, participants were asked whether they identify as Republican, Democrat or other.

The characteristics of respondents are summarized in Table 3. There were more Democrats in the control condition than the direct format condition and so I include in each of our analyses a control for being Democrat.

3.2. Results

The Bureau of Labor Statistics unemployment rate decreased slightly over the target period from 4.1% to 3.9% (Bureau of Labor Statistics, 2018). The incentivized condition was more likely to
forecast this trend than the standard direct format condition. Figure 2 depicts the forecasts made by respondents and Table 4 sets out the results of regression models of correctly forecasting decreasing unemployment.

**Figure 2: Percentage Forecasting Increased, Decreased or Unchanged Unemployment between Jan and August 2018, Study 2**

![Bar chart showing percentage forecasting increased, decreased, or unchanged unemployment between Jan and Aug 2018.]

The first column of Table 4 shows that, controlling for whether one identifies as Democrat, forecasts in the direct format condition were significantly more likely to be correct than those in the standard direct format condition (OR = 1.74, z = 2.79, p = .005).

Model 2 restricts the sample to Republicans (n = 140) and finds that the effect of incentives was virtually zero among this group. Model 3 performs an analogous analysis on Democrats (n = 186) and finds that Democrats made more pessimistic forecasts under a Republican administration in the absence of accuracy incentives than they did when incentivized (OR = .53,
This result is consistent with accuracy incentives crowding out the motivation to disparage the economy under a Republican administration.

Table 4: Logistic Regression of Correctly Forecasting Decreasing Unemployment

<table>
<thead>
<tr>
<th></th>
<th>Main effect</th>
<th>Republicans only</th>
<th>Democrats only</th>
<th>Affect*condition Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Std. Direct Format condition</td>
<td>-0.556</td>
<td>-0.046</td>
<td>-0.638</td>
<td>-0.517*</td>
</tr>
<tr>
<td></td>
<td>(0.199)**</td>
<td>(0.334)</td>
<td>(0.298)*</td>
<td>(0.225)</td>
</tr>
<tr>
<td>Democrat</td>
<td>-0.357</td>
<td></td>
<td>-0.124</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.200)</td>
<td></td>
<td>(0.213)</td>
<td></td>
</tr>
<tr>
<td>CurrentFeeling</td>
<td></td>
<td></td>
<td>0.160</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.095)</td>
<td></td>
</tr>
<tr>
<td>CurrentFeeling* Std. Direct</td>
<td></td>
<td></td>
<td>0.229a</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.136)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.522</td>
<td>0.208</td>
<td>0.417</td>
<td>-0.263</td>
</tr>
<tr>
<td></td>
<td>(0.166)**</td>
<td>(0.216)</td>
<td>(0.239)</td>
<td>(0.452)</td>
</tr>
<tr>
<td>N</td>
<td>417</td>
<td>140</td>
<td>186</td>
<td>416</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
<td>0.07</td>
<td></td>
</tr>
</tbody>
</table>

* $p<0.05$; a $<0.05$ in one-sided preregistered test

The fourth column of Table 4 tests whether incentives reduce the weight placed on feelings about the current economy in forecasting. This is a preregistered test in which I predicted that feelings would positively interact with the standard direct format condition to predict forecasts. The interaction variable is statistically significant in a binary logistic regression of correctly forecasting decreasing unemployment (reported in model 5 of Table 4, $z = 1.69$, one-sided $p = .046$). This indicates that participants in the standard direct format condition relied more heavily on their feelings about the current economy when forecasting than did participants in the direct format condition, supporting my hypothesis.
The fact that the coefficient on the standard direct format condition remains statistically significant after controlling for this interaction implies that the effect of incentives is explained by some additional mechanism, over and above the affect heuristic. Deliberate cheerleading looks to be a likely explanation, given the significant effect of incentives on Democrats.

4. Study 3

Broadly speaking, there are two criteria by which forecasts can be assessed: coherence and correspondence (Ganzach, 2009). The coherence criterion requires a forecast to demonstrate internal consistency. Study 1 demonstrated that standard direct format forecasts are incoherent because they are not consistent with the forecasts returned by a logically equivalent indirect format. Study 2 suggested that standard direct format forecasts are incoherent because they place excess weight on feelings about the current economy.

A second, independent test of the validity of a forecast assesses whether it corresponds to some valid external standard. In Studies 1 and 2 incentivized forecasts were significantly more likely to correspond with the true direction of change in unemployment than were standard direct format forecasts. These results are consistent with the hypothesis that incentivized forecasts are more accurate than standard direct format forecasts, but it is weak evidence because it is drawn from a sample of just two timepoints. In Study 3, I test whether forecasts correspond with incentive-compatible behavior. I asked whether respondents had recently bought a car. All else being equal, respondents whose considered opinion was that the labor market was worsening should be less likely to have bought a car than those whose considered opinion was that the labor market was improving. This intuition is reflected in the fact that the intention to purchase durables is a standard question in consumer sentiment surveys worldwide (e.g. the Michigan
Survey of Consumers, GfK’s Major Purchase Index for the UK). The correspondence criterion in the current study is that a method of eliciting forecasts is more valid than an alternative if it more strongly predicts car purchase than that alternative. I hypothesize that incentivized and indirect format forecasts will more strongly predict car purchase than standard direct format forecasts.

4.1. Methods

The methods and materials are precisely the same as those used in Study 2 but with the addition of an indirect format condition and the addition of the question “Did you get a new car in 2018?” (answered yes or no) at the close of the survey.

Between December 3rd and 6th 2018, the survey was posted on Amazon Mechanical Turk and was described as eliciting “views on life in the US”. The opening questions asked respondents for their gender, age, country of origin and, as a final question, an attention filter that told respondents to click both “sometimes” and “never” to proceed to the survey (Oppenheimer, Meyvis and Davidenko, 2013). 544 participants were recruited. Of these, 475 completed all necessary data for the analyses.

The indirect format questions read as follows:

On January 1st 2018, what percentage of the US labor force do you estimate was unemployed? (just enter a number without a percentage sign)

On November 30th 2018, what percentage of the US labor force do you estimate was unemployed? (just enter a number without a percentage sign)

The standard direct and incentivized questions were worded as follows:
Do you think the Bureau of Labor Statistics data will show that the unemployment rate has increased, decreased, or stayed the same between January 1st 2018 and November 30th 2018?

As in previous studies, the incentivized procedure was prefaced with text explaining that respondents’ payments would be doubled if they correctly forecast BLS trends.

After forecasting, respondents were asked: “How do you feel about the US economy right now?” and answers were made on a seven-point scale from “very negative”, through “neutral” to “very positive”. On the same screen, participants were asked whether they identify as Republican, Democrat or other. On a final screen respondents were asked if they got a new car in 2018.

### Table 4: Characteristics of respondents by survey condition

<table>
<thead>
<tr>
<th></th>
<th>Std. Direct, n = 159</th>
<th>Incentivized, n = 159</th>
<th>Indirect, n = 157</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Age</td>
<td>38</td>
<td>41**</td>
<td>40</td>
</tr>
<tr>
<td>% Female</td>
<td>70**</td>
<td>64</td>
<td>62</td>
</tr>
<tr>
<td>% Republican</td>
<td>33</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>% Democrat</td>
<td>41</td>
<td>45**</td>
<td>39</td>
</tr>
<tr>
<td>% reporting a new car</td>
<td>14</td>
<td>15</td>
<td>17</td>
</tr>
</tbody>
</table>

* p<0.05; ** p<0.01

The characteristics of respondents are summarized in Table 4.

**4.2 Results**

The first result is that the standard direct format condition again delivered forecasts that were less likely to be correct than did the other two conditions. The Bureau of Labor Statistics records that unemployment had fallen from 4.7 to 3.9 percent over the period January to November 2018. Controlling for age, gender and party affiliation, correct forecasts were made by respondents in the standard direct format condition 56 percent as frequently as by respondents in the incentivized condition (n = 318, OR = .56, z = 2.44, p = .015) and 65 percent as frequently as by respondents in the indirect format condition (n = 316, OR = .65, z = 1.80, p = .072). As in
Study 1, the likelihood of making a correct forecast did not differ significantly across the incentivized and indirect format conditions ($n = 316$, OR = .84, $z = 0.72$, $p = .473$).

The novel test in Study 3 is whether the standard direct forecasts do a worse job of explaining having purchased a car than do forecasts made by the other two elicitation procedures. The headline result is that standard direct forecasts were negatively correlated with having purchased a car, $r = -.13$, whereas forecasts from the other two conditions were positively correlated with having purchased a car: $r = .10$ for the incentivized condition and $r = .03$ for the indirect format condition. The negative correlation of standard direct forecasts with car purchase is explained by the results of Study 2, which showed that current feelings about the economy have excessive influence on standard direct forecasts. In this data there is a negative correlation of current feelings about the economy with car purchase ($r = -.04$).

To formally test whether standard direct forecasts do a worse job of explaining having purchased a car than do forecasts made by the other two elicitation procedures, I interact being in the standard direct format condition with making a positive economic forecast (i.e. forecasting that unemployment had decreased). This interaction does have a significantly negative coefficient in a logistic model of car purchase that controls for the main effects of making a positive economic forecast and being in the standard direct format condition (OR = .31, $z = 2.03$, $p = .043$). In summary, standard direct forecasts correspond less well with incentive-compatible behavior than do forecasts made by the incentivized or indirect procedures.

**Discussion**

The current research demonstrated that reported expectations changed systematically when expectations measures were incentivized for accuracy. It also demonstrated that incentivized
forecasts explain purchasing behaviour better than forecasts from the standard direct format. Lastly it demonstrates that an indirect format elicits forecasts that are more in line with incentivized forecasts than with standard direct format forecasts.

It will not surprise those working with survey data that incentivizing a survey measure would cause a shift in responses relative to those returned by an incentivized, but otherwise identical, question. Researchers in preference elicitation have for decades expressed concern about strategic response (e.g. Diamond and Hausman, 1994). Strategic response cannot explain the effect of incentives on Democrats in Study 1, however. A contribution of this paper was to identify that incentives had the effect of shifting the process that informs forecasts from heuristic System 1 reasoning to deliberative System 2 reasoning. Study 2 showed that respondents who are asked the cognitively taxing question “will unemployment have increased or decreased between January and September?” answer with reference to how they feel about the current economy. These feelings are likely to include some relevant information for forecasting the economy, but I predicted that incentives would prompt respondents to place greater weight on other relevant information. Consistent with this hypothesis, Study 2 found that incentivized participants relied less on their feelings about the current economy in forming forecasts than did unincentivized participants and so delivered significantly different – and more accurate – forecasts. This result presents process evidence consistent with Schleifer and co-authors’ family of models on expectations formation (Barberis, Greenwood, Lawrence and Shleifer, 2018; Bordalo, Gennaioli and Schleifer, 2018; Greenwood and Shleifer, 2014; Gennaioli, Ma, and Shleifer, 2016).
One straightforward implication of our results is that current survey measures of expectations (e.g. the consumer sentiment index) fail to capture private information that could potentially be included to improve forecasting power.

The implications of this research for understanding individual behavior will depend on the degree to which people internalize their heuristic judgments of expectation. Let’s take one extreme: whenever people are in the process of making a consequential decision regarding the future they form their expectations as though those expectations were incentivized for accuracy. If this is the case then people report one set of expectations in an unincentivized survey but they bring a different set of expectations to bear on their consequential actions. This is important because unincentivized expectations data are used to predict consequential behaviors e.g. migration decisions, human capital investment decisions; for a more thorough discussion see Delavande et al., (2011). The predictions from these models will be biased because they rely on biased expectations data. Notwithstanding the results of Study 3, there is some empirical support that biasing heuristics in judgments of trend do influence behavior. Li, Johnson and Zaval (2011) found that belief in global warming is higher on hot days than on cold days and also found that donations to mitigate climate change were higher on hot days than on cold days. Donating to a charity is a low stakes decision and so future research should consider to what extent attribute substitution informs higher stakes decisions. Between the two extremes considered above lies what I consider to be the most likely scenario: when people are making consequential decisions regarding the future they have an incentive to make the best decision given the available information; yet they have a multitude of different attributes and contingencies to consider and so they may conserve on cognitive effort when processing that information. This intermediate
scenario implies both biased predictions of behavior and also behavior that is suboptimal as a result of having been formed with reference to biased expectations.

I conclude that incentives serve to motivate normative response to the standard expectations measure that is used in survey research. Of course, there are several practical constraints on incentivizing expectations questions. Economists may wish to measure expectations over outcomes that cannot be reliably measured. In these cases it is impossible to determine or reward accuracy. Even where it is possible to determine accuracy, the practicalities of following up with respondents to reward them for their accuracy may compromise anonymity or be prohibitively costly. The indirect format suffers none of these practical limitations and looks to deliver similar effects to incentives (Comerford and Soll, 2019). A promising result from Study 1 in this regard is that a high degree of knowledge does not look to be a prerequisite for retrieving insightful forecasts from the indirect format. Recall what happened when I dropped from the indirect format sample the 25 percent of participants who appeared most knowledgeable regarding unemployment. The remaining 75 percent still outperformed the (on average) more knowledgeable sample of respondents from the standard direct format condition, where “outperformed” means that they delivered forecasts that were more similar to those from the direct format condition. A corollary to this result is that knowledgable respondents can still deliver biased forecasts if they rely on heuristics. Imposing a normative process on forecasting, as the indirect format does, appears to be a promising and generalizable means to deliver less biased measures of expectation. It is noteworthy that surveys to elicit forecasts from professionals and experts often use the indirect format e.g. the Blue Chip Financial Forecasts survey asks respondents to estimate interest rates rather than asking them whether interest rates will rise or fall.
5. References


Borghans, Lex, Huub Meijers, and Bas Ter Weel. 2008. The role of noncognitive skills in explaining cognitive test scores. Economic Inquiry, 46(1), 2-12.


