International Journal of Forecasting 41 (2025) 702–715

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Partisan bias, attribute substitution, and the benefits of an indirect format for eliciting forecasts and judgments of trend

David A. Comerford [a,*], Jack B. Soll [b]

[a] Behavioural Science Centre, Economics Division, University of Stirling, FK9 4LA, UK
[b] Fuqua School of Business, Duke University, NC 27708-0120, USA

ARTICLE INFO

Keywords:
Attribute substitution
Judgment of trends
Comparison processes
Ideology
Partisan bias

ABSTRACT

A majority of Americans reported the economy to be worsening when objective indicators showed it to be recovering. We show that this is symptomatic of attribute substitution—people answer a taxing question as though asked a related easy-to-answer question. An implication of attribute substitution is that forecasts will vary across a direct format, which asks whether the economy will be better in 12 months, versus an indirect format, which asks respondents to rate both current conditions and the conditions they expect for 12 months' time. We compare these formats in three studies and over 2,000 respondents. Relative to the direct format, the indirect format delivers trends that show greater consensus across Republicans and Democrats; are less equivocal about the course of the US economy; and are more realistic about the magnitude of change in opinion poll data.

> Ronald Reagan used to ask the question, "Are you better off than you were four years ago?" In this case, are you better off than you were in six? And the answer is, the country is definitely better off than we were when I came into office (Obama, 2014).

Upon hearing President Obama's claim, many Americans were incredulous. According to a contemporaneous *Fox News* poll, 59% deemed as mostly false the statement "By almost every measure, the American economy and American workers are better off now than they were in 2008" (Blake, 2014). How can it be that many people thought the economy was getting worse, when the opposite was true?

In their poll, Fox News was asking what we call a *direct format question*, in which respondents are asked directly, in a single question, whether conditions at time 2 are better or worse than conditions at time 1. Direct format questions are widely used in forecasting and decision making. For instance, the Michigan Survey of Consumers uses this kind of question to elicit the beliefs of ordinary people about changes in the US economy. Results from these surveys are highly publicized and consequential. For instance, stock exchange trading volumes increase on release of each monthly wave of data from the Michigan Survey (Hiban, 2013).

Kahneman and Frederick (2002) coined the term *attribute substitution* for the psychological process by which people intuitively and unwittingly replace a difficult-to-answer question with an easier-to-answer one. In this paper, we apply this idea to understand judgments of change over time. We propose that people often answer direct format questions about trends in x by answering a simpler question, such as "Is x good or bad now?" Many people have a ready answer to this and, when x concerns

* Corresponding author.
E-mail addresses: david.comerford@stir.ac.uk (D.A. Comerford), jsoll@duke.edu (J.B. Soll).

https://doi.org/10.1016/j.ijforecast.2024.11.005

the economy, that answer can often be predicted by their political affiliation. *Partisan bias* in this realm is well documented: Opponents of the political party in power tend to express more negative economic trends than supporters (Bartels, 2002; Conover, Feldman, & Knight, 1987). While it is especially prevalent in the United States—the focus of our studies—partisan bias is also observed in other polities such as the United Kingdom (Wagner, Tarlov, & Vivyan, 2014) and Indonesia (Budi & Pamungkas, 2020).

The extant literature offers several explanations for partisan differences in judgments of trend. First, people may experience "separate realities" that lead them to draw different conclusions (Kull, Ramsay, Subias, & Lewis, 2004). Many individuals circulate within "echo chambers" of selectively presented facts and conforming perspectives (Flaxman, Goel, & Rao, 2016). Even when aware of bias in the sampling of opinions, people do not fully correct for the fact that their observations are biased by the slice of the world presented to them (Koehler, 2014). Second, people may derive satisfaction from casting their own political party in a favorable light and disparaging the accomplishments of their political opponents. This may outweigh scruples they have about misrepresenting their true beliefs. Recent studies show that partisan differences are attenuated when people are paid for correct answers to factual questions about politics and the economy, which has been interpreted as providing evidence for insincere reporting (Bullock, Gerber, Hill, & Huber, 2015; Prior & Sood, 2015). Third, motivated reasoning may lead people to seek out, recall, evaluate, and interpret information to reach a desired conclusion (Kunda, 1990). This type of reasoning can happen outside of awareness and often entails self-deception, which distinguishes it from outright dishonesty (Dunning, 2015). Motivated reasoning is especially well documented in the political domain (e.g., Jost & Amodio, 2012) and can even cause people to develop and maintain beliefs that strain plausibility (Dunning, 2015).

Here we propose attribute substitution as an explanation for why people, especially opponents of the party in power, report the economy to be in decline when it is in fact strengthening. Substitution is likely when a response to a question different from the one being asked readily comes to mind. Consider the question, "Are you better off than you were *x* years ago?" A logical response to this question requires three steps: (1) evaluate the target (today's conditions); (2) evaluate the referent (conditions *x* years ago); (3) compare the two and report the trend accordingly. For many people, a thought or association will immediately come to mind when the economy is mentioned (e.g., *things are bad now, I don't make enough money*). Attribute substitution can cause bias when people neglect to consider whether their initial thought delivers an answer that fits the question being asked. This process is consistent with a dual-process model of cognition, in which an intuitive system provides an initial response and a reflective system only sometimes intervenes to check or correct that response (Kahneman & Frederick, 2002; Meyer & Frederick, 2023). Thus, even if conditions are poor but improving, a quick response combined with a lack of reflection can cause some people to respond that conditions are getting worse.

We are not the first to suggest attribute substitution as a source of bias in judgments of economic and social trends. When asked to what extent alcohol abuse, teenage loitering, or any other social ill is a problem, people's answers track their global affective evaluation of the state of society (van der Bles, Postmes, & Meijer, 2015). Similarly, after the 2008 financial crisis, individuals who had a more negative emotional reaction tended to perceive a greater amount of risk (Burns, Peters, & Slovic, 2012). The authors explain these results as a manifestation of the affect heuristic, a form of attribute substitution that evaluates a target with reference to the feelings that the target evokes.

There is also an influential model of expectations formation in the finance literature that implicates attribute substitution as driving a range of forecasting biases that are commonly observed in investment decisions, such as extrapolation bias and risk neglect (Bordalo, Gennaioli, & Shleifer, 2018). The diagnostic expectations model posits that decision makers answer a question that asks for a forecast of the economy as though they have been asked "What news have you heard about the economy recently?" As a result, "diagnostic expectations overweight future outcomes that become more likely in light of incoming data" (Bordalo et al., 2018, p. 199).

Though we agree that substitutions of this type lead to bias in forecasts, our focus here is different. Bordalo et al. (2018) posit recent news as the cue that forecasters draw on when formulating their forecasts. They then proceed to work out the implications for market volatility. We are more agnostic about the cues to which people attend. Some people may rely on the recent news they have heard to inform their forecasts; others may let their feelings guide their forecasts; some may call on their attitudes toward the economy; some may use an amalgam of all these cues. We theorize that from attending mainly to information and attitudes about conditions in the moment and treating this as a sufficient substitute for the three-step calculation of change described above, partisan bias will emerge. We expect this to happen because partisanship is likely to motivate people to use and interpret cues in a way that favors their preferred answer. Crucially, our reasoning suggests that using indirect questions that ask separately about conditions at different points in time can attenuate partisan bias.

1.1. Indirect versus direct elicitation of trends

We opened this paper by discussing a *direct format question* that asks, "has *x* improved or worsened since a year ago?" and noted that a logical response to this question requires the respondent to engage in three steps. Now consider an alternative elicitation approach. Instead of asking whether *x* has improved or worsened since a year ago, the indirect format elicits separate ratings of *x* today and *x* a year ago. A trend judgment can then be inferred by comparing the two ratings. The indirect format explicitly guides the respondent through the necessary steps, whereas the direct format does not.

It is novel to apply the indirect format to trend judgments, but it has been applied successfully in the domain

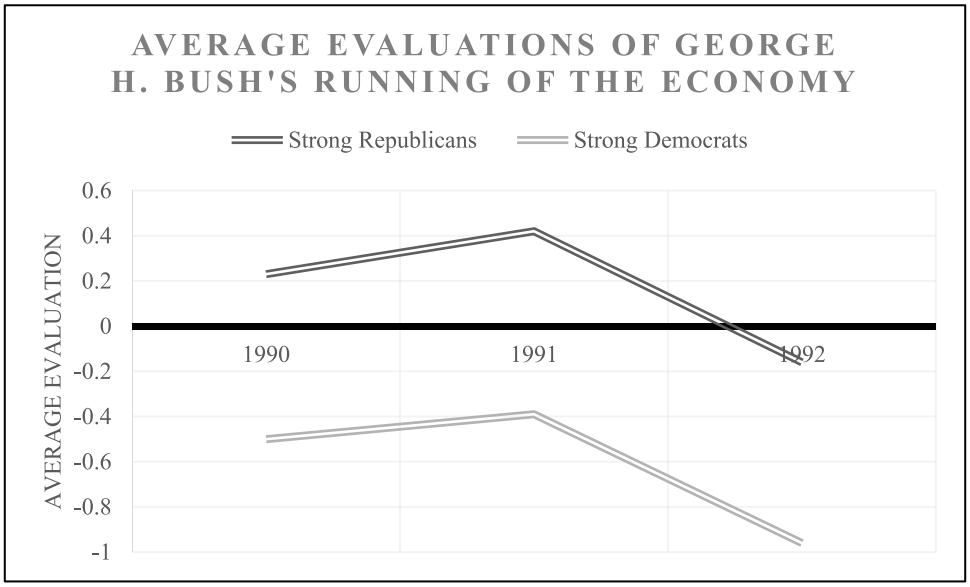

Fig. 1. Across partisans, levels of approval differ but trends in approval over time are similar.
Source: Adapted from Figure 1 of (Bartels, 2002).

of interpersonal comparisons. Chambers and Windschitl (2004) review the cognitive sources of the better-than-average effect and consider the magnitude of that bias under a direct versus an indirect elicitation format. Studies consistently show that better-than-average effects are larger if a respondent is asked in a direct format than if asked in an indirect format. One set of studies is especially informative for understanding bias in judgments of trend over time. In a series of experiments, Klar and Giladi (1997) and Giladi and Klar (2002) manipulated the absolute quality of sets of items and then asked participants to make comparisons within those sets. They found that participants who were comparing within a set of superior items were biased toward judging any randomly selected item as better than the average of that set whereas those comparing within a set of inferior items were biased toward judging any item as worse than average. This pattern of results suggests a general error: When people are asked to compare the performance of a specific target relative to some specific referent, they answer as though asked about the absolute performance of the target. This conflation of absolute and relative evaluations predicts that when the economy is weak but slowly recovering from a crisis, people will incorrectly report the economy to be worsening.

There is every reason to suspect that the two judgments that make up the indirect format will be biased also. A key advantage of the indirect format is that any bias that is common to both ratings will cancel out when trend is inferred. To see why this is especially useful, see Fig. 1 below, taken from Bartels (2002), p. 126). The figure depicts opinion poll results on George H. W. Bush's running of the US economy at three points in time. Whenever people were asked to evaluate Bush's running of the economy, Democrats were more negative than Republicans, consistent with partisan bias. The change in approval from one period to the next, however, is strikingly similar across Democrats and Republicans. If we focus on perceptions of change rather than on overall level, members of both parties are providing the same message.

There is another reason that the indirect format is expected to produce less biased comparisons of x and y than the direct format, which is that trends elicited by the indirect format should be less vulnerable to insincere response (e.g., partisan cheerleading) than those elicited by the direct format. This is because the indirect format is less transparent in eliciting a comparison or a measure of trend. Taken together, this leads to our first hypothesis.

H1: The indirect format will manifest smaller partisan differences than the direct format.

A useful regularity we can harness to test our argument is that anyone who is partisan biased in judgments of trends in the economy is also likely to be biased in their judgments of the absolute level of the economy. Thus, under a Democratic president, Democrat partisans will tend to report inflated answers to questions about both the trend in the economy and about the level of the economy at each point in time, and Republicans will tend to report deflated answers.

Judgments of trend are also likely to be affected by reality, in addition to impressions and attitudes. There are two reasons for this. First, more thoughtful respondents might actually consider conditions at both time periods when asked by the direct method. Second, even if not explicit, perceptions of trend could still serve as a cue to direct judgments (e.g., having read about trends in the news) in a way that does not get filtered through attitudes. Suppose that we ask the same person to respond to both the direct method (e.g., a question that asks how the economy has changed in the past year) and

the indirect method (e.g., two questions that ask about the state of the economy one year ago and today). If people are sensitive to the actual trend, the trend from the indirect method (as measured by the difference between the two responses) should be predictive of the same person's response to the direct question. Moreover, if direct responses are also driven by attribute substitution, then the level of the responses from the indirect method (as measured by the average of the two indirect responses) should be predictive as well.

> H2: If the same person answers by both the direct and indirect method, then the direct trend will be predicted (positively) by both indirect trend and indirect level. Moreover, there will be an effect of indirect level even after the effect of indirect trend is accounted for.

1.2. Overview of studies

The studies presented below were conducted in 2012 and 2020. Study 1 was the first, in which we randomly assigned respondents to either the direct or indirect format. We found large differences in the economic trends reported by respondents depending on which format they saw. Studies 2 and 3 were designed to probe the robustness of that difference and the mechanisms driving it. Consequently, elements of the design of the studies vary. Studies 2a and 2b, which were conducted in 2020, ask about trends in opinion polls. Study 3, conducted in 2012, asks about economic trends but uses different questions and response scales than Study 1 does. Also, Studies 2a and 2b required respondents to report the magnitude of change in a variable over time and so these studies, unlike the others, did not allow participants to indicate that there had been no change over time. The nature of the attention filter (Oppenheimer, Meyvis, & Davidenko, 2009) also differed across studies, and consequently the studies likely differed in terms of the threshold level of attention that respondents needed to remain in the sample. Across all these variations, each of Studies 1, 2, and 3 shows that partisan differences were larger when answered by the direct format than when answered by the indirect format, in line with H1 (which was typically supported).

Two contributions that Study 2 makes relative to Study 1 are that it uses a within-participant design and that it asks about an objectively verifiable outcome, the magnitude of change in presidential candidates' opinion poll ratings. The within-participant design of Study 2 allows us to test H2. To expand on this, suppose that the indirect ratings comprise a time 1 and a time 2 rating (t1 and t2). The respondent provides assessments of both t1 and t2, and also D, which is a direct rating of the trend. Typically, t1 and t2 are elicited contemporaneously within the survey, and D is elicited elsewhere in the survey to minimize any carryover effect. Attribute substitution holds that the measures D, t1, and t2 are imbued with a common bias. That is, when one is asked about the economy, one's feelings and attitudes affect answers to all of these questions, even though the questions are about substantively different things.

The difference between the indirect assessments, t2 – t1, provides an implicit measure of judged trend. In principle, this implicit trend judgment should perfectly predict the direct trend judgment. That is, for each respondent it ought to be the case that t2 −t1 = D (perhaps with added noise, but without a systematic bias). This principle would be violated if attribute substitution played a role, however. In that case, we expect an additional effect on D of the level of their indirect assessments, as captured by the mean of t1 and t2, which we refer to as the indirect level. Thus, a regression of D on both the indirect trend and the indirect level provides a test of H2. We expect a positive influence of both, where the additional effect of indirect level isolates the impact of attribute substitution on direct trend judgments.

We further test both hypotheses in Study 3, which is a conceptual replication. Like Study 1, it asks respondents to forecast the economy. Like Study 2, it is a within-participant experiment and so offers an opportunity to apply the statistical model from Study 2 to new data.

For each study, our criterion for choosing our sample size and all exclusions of data are disclosed in the *Participants* subsection. While Studies 1 and 3 were conducted before preregistration became commonplace, Studies 2a and 2b were both preregistered. All survey questions asked of participants for all studies are reported in the online supplement. The coding of variables and the model specifications are reported in full in the method and results section of each study.

2. Study 1: Hindcasts and forecasts of a weak but recovering economy

Study 1 was conducted in early 2012, when the US economy was recovering from a deep recession. Official unemployment figures show improvement over the period that our participants were asked to backcast: from 9% unemployed in April 2011 down to 8.2% in April 2012, when our study was conducted. The period our participants were asked to forecast saw improvement also: unemployment had declined to 7.5% by April 2013 (Bureau of Labor Statistics, 2013).

2.1. Method

2.1.1. Participants

Respondents were recruited from Clearvoice (now ROI Rocket), which maintains an online survey panel of US residents. Data were collected between April 11 and 14, 2012. We aimed to recruit 500 respondents, as that is the sample size used to construct the Michigan Index of Consumer Sentiment (Curtin, 2013). A total of 543 respondents consented to take part, of which 181 had previously identified themselves to the panel provider as Democrats and 131 as Republicans (remaining participants reported either a third party or no affiliation). Sixteen failed an attention filter. Nine failed to complete the survey. These were the only exclusions. This leaves a sample of 518 ($M_{age} = 49$; 59% female).

Table 1
Question wordings.

	Michigan wording	Our wording
Direct economy	Now turning to business conditions in the country as a whole—do you think that during the next twelve months, we'll have good times financially or bad times or what? [good times/ uncertain/bad times]	Think about the economy of the United States as a whole. What do you think about the times the country has had financially over the **past year**. Has it gotten better or worse? What do you think about the times the country will have financially over the **next year**. Will it get better or worse?
Indirect economy		Think about the economy of the United States as a whole. Do you think that **one year ago** the country was having good times financially or bad times? Do you think that **these days** the country is having good times financially or bad times? Do you think that **one year from now** the country will be having good times financially or bad times?

Notes: The Michigan question is answered "good times", "uncertain", or "bad times." The direct format question is answered on a nine-point scale that runs from "much worse" = 1, through "same" = 5, to "much better" = 9. Each indirect question is answered on a 9-point scale that runs from "very bad times" to "very good times."

Table 2
Results of binary logistic regressions of Trend on response format from Study 1.

	Proportion implied "better" by format					Proportion implied "worse" by format				
Target	%Dir.	%Ind.	Odds ratio	*z*-score	*p*-value	%Dir.	%Ind.	Odds ratio	*z*-score	*p*-value
Forecasts	39	58	.46	4.38	.001	39	6	9.85	7.94	.001
Hindcasts	24	44	.40	4.71	.001	53	17	5.70	8.31	.001

Notes: z-scores are taken from the following binary logistic regressions: prob($better_i$) = *direct format* $_i$ + $order_i$ where $better_i$ is an indicator variable that takes a value of 1 if respondent *i* implied "better" and a value of zero if they implied "same" or "worse"; and, on right, prob($worse_i$) = *direct format* $_i$ + $order_i$ where $worse_i$ is an indicator variable that takes a value of 1 if respondent i implied "worse" and a value of zero if they implied "same" or "better." *Direct format* $_i$ is an indicator variable coded as 1 if respondent *i* answered by the direct format and 0 if they answered by the indirect format. The control variable $order_i$ is an indicator variable that takes a value of 1 if respondent *i* had been randomly assigned to see the economy question before the household question. $N = 518$.

2.1.2. Design

The study had a 2 (format) × 2 (order) between-subject design. The manipulation of interest is format; each participant reported trends in response to either a direct or indirect format. As a robustness check, we also manipulated whether respondents first judged trends in the economy or first judged trends in household finances.

2.1.3. Procedures

Respondents were assigned to either the direct or indirect format for all five targets they were asked to judge: the US economy; own household finances; own household's spending; a neighbor's financial condition; and conditions at a neighboring workplace. Each target was asked about on its own screen, where respondents made a hindcast of the direction of trend over the past year and a forecast over the coming year.

Because the US economy and own household finances are what is asked about in consumer sentiment surveys, questions about these targets appeared first and their order was counterbalanced across participants. We control for this order manipulation in each of the analyses below. Questions on the economy were worded similarly to those used in the Michigan Consumer Sentiment index and are presented in Table 1. Question wordings for the remaining targets are reported in the online supplement.

Our dependent variable is *trend* and takes one of three values: worse, same, or better. Indirect format trends are coded as "better" if the rating for the later period is more positive than the rating for the earlier period. They are coded as "worse" if the later rating is more negative than the earlier rating, and as "same" if both ratings are identical. Direct format trends are coded as "better" if the response was more positive than "the same" and are coded as "worse" if the response was more negative than "the same".

2.2. Results

In what follows, we report results concerning the US economy. For reasons of clarity and concision, we relegate to the online supplement tables of results concerning the other targets—household finances, household spending, neighbor's financial situation, and conditions at a neighboring workplace.

2.2.1. Indirect versus direct trends

Participants were more pessimistic about trends if beliefs were elicited by the direct format than by the indirect format, as is evident from the results summarized in Table 2. Only 39% of participants indicated that the economy would improve by the direct format, compared to 58% by the indirect format. This difference was significant ($z = 4.38$, $p = .001$) in a logistic regression that controls for task order. Table 2 shows that the results are similar when the dependent variable is coded for indicating whether or not the economy would worsen, and also when hindcasts are evaluated. Finally, the greater pessimism by the direct

Table 3
Mean trend (left) and results of OLS regression of mean trend (right) from Study 1.

Target	Direct Rep $n = 66$	Direct Dem $n = 95$	Indirect Rep $n = 65$	Indirect Dem $n = 86$	Direct Δ	Indirect Δ	t-score on party × format inter-action	p-value party × format inter-action
Forecast	−0.44	0.41	0.49	0.65	0.85	0.16	4.23	.001
Hindcast	−0.65	0.09	0.09	0.45	0.74	0.36	2.18	.030

Notes: The t-scores and p-values refer to the interaction of direct format and party affiliation from OLS regressions of *Judged Trend*. The independent variables in these regressions are dummies for order, whether or not the participant identifies as Republican, and whether or not they responded via the direct format. The models also include the crucial interaction between format and Republican, the t-score and p-value of which is reported in the Table. The dependent variables is coded as 1, 0, or –1 to indicate whether the participant indicated that the economy would improve, stay the same, or worsen, respectively. Non-partisans were removed from this analysis, leaving a sample size of $N = 312$.

format remains statistically significant when the interaction between task order and format is also included. The online supplement shows that for all five targets, respondents were significantly more likely to indicate that things had worsened and that they would worsen by the direct format than by the indirect format (Tables S1 and S2).

The indirect format delivered trends that were less equivocal than the direct format delivered. Looking to the top row of Table 2, we see that respondents to the direct format were equally split: 39% reported the economy would get better in the future, 39% reported it would get worse, and 22% (not shown) answered "same". There was greater consensus that the economy would get better among respondents to the indirect format: responses from 58% of participants implied better, 36% gave the same assessment for the two time periods, and just 6% forecast implied that the economy would worsen. This pattern of results is consistent with the conjecture that the indirect format extracts greater signal from respondents than does the direct format. We return to this point in discussing these results.

2.2.2. H1: The indirect format reduces partisan differences

We now turn to a test of our primary hypothesis, which is that the indirect format reduces partisan differences in trend judgments (H1). For this analysis, we limited the sample to those respondents identified by the panel provider as Republicans or Democrats ($N = 312$). We coded judged trend as 1, 0, or −1 according to whether participants indicated improvement, no change, or worsening of the economy, respectively. This coding has the advantage of rendering trends across the two formats comparable.

Table 3 presents the mean judged trends for both Republicans and Democrats within each of the elicitation formats. For example, the first row shows that for forecasts, Republicans on the average reported a trend of −0.44 when responding to a direct question, compared to 0.41 for Democrats. The difference between these two averages is $\Delta = 0.41 - (-0.44) = 0.85$. This value indicates a large discrepancy between Democrats and Republican when they responded to the direct format. The corresponding result for the indirect format is $\Delta = 0.16$, which suggests much less partisan disagreement.

To test whether this attenuation of disagreement is statistically significant, we ran an OLS regression with trend as the dependent variable (−1 = worse; 0 = same; 1 = better) in which the independent variables are format (1 = direct, 0 = indirect), partisan affiliation (1 = Republican; 0 = Democrat), and their interaction. The model additionally controls for the economy/household ordering manipulation. The key test concerns the party × format interaction, which can be interpreted as capturing the degree to which partisan differences in judgments depend on the elicitation format. As shown in Table 3, this interaction was significant for both forecasts and hindcasts, providing support for H1. Furthermore, Table S1 of the supplement shows that this interaction was directionally consistent with H1 for 9 out of 10 targets,[1] and was statistically significant for 5 of them.

2.2.3. Exploratory analysis: Indirect level and attribute substitution

An exploratory analysis from Study 1 relates to our theory that attribute substitution is a cause of partisan bias. Here again we restrict the sample to Republicans or Democrats. We first look to those who answered by the indirect format ($N = 151$). From their ratings, we can derive measures of both Indirect Trend, (t2 – t1), and Indirect Level, (t2 + t1)/2. Attribute substitution predicts that a respondent consults easy-to-access cues (e.g., attitudes) when forming an evaluation of the economy. If that is so, then the two evaluations that form Indirect Level will each be informed by these same cues. To the extent that the content of those cues is partisan-skewed, we would expect to see an association between partisan affiliation and Indirect Level. Table 4 shows that the correlations between being a Democrat and Indirect Levels are substantial ($r = .50$ for forecasts and $r = .44$ for hindcasts). These results are consistent with the idea that common cues underlie responses to any single question that asks respondents to evaluate the economy. Also consistent with that mechanism, the fourth column of Table 4 shows substantial correlations between Democrat and Direct Trends ($r = .46$ for forecasts and $r = .43$ for hindcasts).

Meanwhile, the first column shows that the correlations between Democrat and Indirect Trends are relatively low ($r = .12$ for forecasts and $r = .28$ for hindcasts). This result is consistent with our claim that the impact of cues such as attitudes and feelings are attenuated when trend is measured indirectly as t2 – t1. The relatively high correlations between Democrat and Indirect Levels demonstrate that a common partisan bias is present in both t1

[1] The exception asked about change in household spending over the past year.

Table 4
Correlations with being democrat from study 1.

	Indirect Trend, t2 − t1 ($n = 151$)	Indirect Level, t2 + t1 ($n = 151$)	Trend vs. Level, Test of difference in r ($n = 151$)	Direct Trend ($n = 161$)	Indirect vs. Direct Trend, Test of difference in r ($n = 312$)
Forecast	r = .12	r = .50	$z = 4.39$, $p < .001$	r = .46	$z = 3.30$, $p < .001$
Hindcast	r = .28	r = .44	$z = 1.70$, p = .091	r = .43	$z = 1.51$, p = .132

Notes: The reported correlation coefficients are Pearson's r. Direct trend is coded as −1 = worse; 0 = same; 1 = better. The trend vs. level test of difference in r is calculated using the test of the difference between two dependent correlations with one variable in common (Lee & Preacher, 2013). The direct vs. indirect test of difference is calculated using the test of the difference between two independent correlation coefficients (Preacher, 2002). The online supplement reports these analyses and test statistics in further detail.

and t2, and the relatively low correlations between Democrat and Indirect Trends demonstrate that this common partisan bias gets subtracted out when calculating the indirect trend.

2.3. Discussion

Study 1 shows that judgments of economic trends were substantially more positive if elicited using an indirect format than if elicited using the standard direct format. Also, the indirect format led to a smaller partisan gap. These patterns of discrepancy across direct and indirect formats were systematic; the supplement shows that they are consistently observed for other targets too. A key insight on why these direct versus indirect discrepancies occur is that, among judgments elicited by the indirect format, partisan affiliation correlated more strongly with the level variable than it did with trend. This result is consistent with the idea that the associations evoked by a target (e.g., feelings, beliefs, attitudes) contaminate the summary evaluations that respondents give when answering a question about the economy. The overall pattern of results suggests that this contamination occurs both in responses to the direct format question that asks whether the economy is getting better and also in response to each of the two questions that form the indirect format. This mechanism corresponds to attribute substitution—in judging whether the economy is getting better or worse, respondents often answer an easier question (e.g., "how do I feel about the economy") rather than judging the economy at each time point and taking the difference.

Additionally, the observed results show patterns that cannot be explained by either of two alternative mechanisms, partisan cheerleading and response noise. Cheerleaders can easily skew their direct format forecasts because their optimism or pessimism translates directly and transparently onto the response scale. The indirect format is less transparent because its forecast is derived from a comparison of two responses. Thus, a would-be cheerleader for the governing administration not only would have to report that things are good now, but also would need to make sure to provide a less positive assessment for last year. For our data, the partisan cheerleading mechanism suggests a specific hypothesis. In 2011, when these data were collected, the Democrat Barack Obama was president. If cheerleading were the only driver of direct vs. indirect discrepancies, then we would see Democrats report "the economy will get better" more frequently than their indirect forecasts imply. In fact, the observed results show the opposite pattern. Democrats were less likely to report "the economy will get better" than to imply that it would get better by the indirect format (see Table 3). This result cannot be reconciled with deliberate cheerleading as the primary mechanism that underlies format differences. This pattern can, however, be explained by the attribute substitution mechanism. The year 2011 was a time when many people, Democrats as well as Republicans, harbored negative feelings, attitudes, and associations toward the economy following the 2008 financial crisis. Though the economy was recovering, negative cues likely came to mind when respondents were asked in the direct format for the direction of change in the economy. Thus, even Democrats were biased toward reporting that the economy would get worse when answering by the direct format.

Having ruled out deliberate cheerleading as a potential confounding explanation for differences in partisan effects across formats, we next consider another: response noise. The indirect format relies on two survey responses, whereas the direct format relies on just one. The indirect format therefore offers twice the opportunity for error to enter a respondent's trend measure as does the direct format. A thought experiment clarifies what would happen if indirect forecasts were all noise and no signal—forecasts would be just as likely to indicate improvement as worsening. Hence, a noisier format returns more equivocal forecasts. The fact that the indirect format was generally less equivocal than the direct format is contrary to the predictions of the noise account.

3. Study 2: A pre-registered within-subject test

Studies 2a and 2b add to the generalizability of our results. Whereas Study 1 asked respondents about trends in the economy, here we asked about trends in public opinion of two prominent politicians: Study 2a asks about Donald Trump (Republican) and Study 2b asks about Joe Biden (Democrat). Another difference in these studies is that we ask participants to estimate the magnitude of change over time in a verifiable metric, candidate approval ratings.

3.1. Methods

3.1.1. Participants

For Study 2a, we recruited US residents over age 18 from the commercial panel provider ROI Rocket between

Table 5
Questions used in studies 2a and 2b.

Format	Questions	Notes
2a Direct introductory question	On April 10th the most recent data was released on President Trump's approval rating. In that most recent data, is the proportion of US residents who answered that they approve "of the way President Trump is handling his job as president" higher or lower than it was this time last year, in April 2018?	Response scale: "higher—More approve of Trump's presidency now than did a year ago" Or "lower – Fewer approve of Trump's presidency now than did a year ago."
Direct estimate of change	By how many points do you think President Trump's approval rating has gone up? i.e. Out of every 100 US residents surveyed, *how many more* answered "approve" in the most recent data than answered "approve" in April 2018?	Question wording varied depending on answer to the direct format question. Respondents who answered "lower" were asked by how many points the rating went down.
2a Indirect	On April 10th the most recent data was released on President Trump's approval rating. In that most recent data, how many out of every 100 US residents answered that they approve "of the way President Trump is handling his job as president"?	
	This time last year, April 2018, how many out of every 100 US residents answered that they approved "of the way President Trump is handling his job as president"?	If respondents entered the same number as they entered for "today" then they received an instruction to provide a different answer.
Direct introductory question	Gallup have been measuring favorability ratings of Joe Biden using the same question since 2008. Today is Joe Biden's favorability rating higher or lower than it was three months ago?	Response scale: "higher – more Americans have a favorable opinion of Joe Biden today than three months ago" or "lower – fewer Americans have a favorable opinion of Joe Biden today than three months ago.
Direct estimate of change	By how many points did Joe Biden's favorability rating go up? i.e. Out of every 100 Americans surveyed, *how many more* would answer that they have a favorable opinion of Joe Biden today than 3 months ago?	Question wording varied depending on answer to the direct format question. Respondents who answered "lower" were asked by how many points the rating went down.
Indirect	Today, how many out of every 100 Americans would answer that they have a favorable opinion of Joe Biden.	
	Three months ago, how many out of every 100 Americans answered that they had a favorable opinion of Joe Biden.	If respondents entered the same number as they entered for "today", then they received an instruction to provide a different answer.

May 3 and 7, 2019. We recruited only respondents who had previously identified themselves to the panel provider as Democrats or Republicans. We aimed to recruit 250 Republicans and 250 Democrats. Republicans answered in much larger numbers than the panel provider expected, so ultimately our sample comprises 1177 respondents (249 Democrats and 928 Republicans; $M_{age} = 54$; 57% female). For Study 2b we recruited US residents over age 18 in precisely the same way. Data were collected on May 24, 2019. Study 2b included an attention filter, which 476 people failed,[2] and so its final sample comprises 507 respondents (253 Democrats and 254 Republicans; $M_{age} = 57$; 64% female).

3.1.2. Design

Study 2a is a 2 × 2 × 2 mixed design. The within condition is direct format vs. indirect format. The between conditions are party affiliation (Democrat vs. Republican) and order (direct format first vs. indirect format first). Study 2b is the same, except that we dropped the order manipulation; all respondents answered by the direct format before answering by the indirect format.

3.1.3. Procedure

Each study asked respondents to estimate the change in polling for a given politician over a given period. To help respondents make these judgments, each survey opened by giving respondents important background information on how the polls are conducted and by quizzing them on their comprehension of that material.

Study 2a asked our respondents to estimate the scale of change in Trump's approval ratings over the past year. We showed respondents the question used in Gallup's surveys; explained that an approval rating is the percentage of respondents who report that they approve of the President; and showed the highest and lowest approval ratings ever recorded since polling began in the late 1950s. Respondents were then presented with a one-item comprehension check that asked them to calculate the difference in approval rating between the highest and lowest ratings ever recorded (see online supplement for the quiz). Those who answered incorrectly were asked the comprehension question once again. All respondents then proceeded to the remainder of the survey.

Study 2b asked respondents to estimate the scale of change in Joe Biden's favorability ratings over the previous 3 months by both the direct and indirect formats (always in that order). The survey opened by showing respondents the question used in Gallup's surveys and explained that a favorability rating is the percentage of respondents who report holding a favorable opinion of a named person. Respondents were then presented with

[2] The attention filter read as follows: "Will you read each question before answering? To indicate that you will, tick only the middle two options listed below: I will answer all questions; I will read all questions; I will only answer questions I have read; I will not answer all questions".

a three-item comprehension check that asked them the highest and lowest levels of favorability that could possibly be recorded and to calculate the largest change in favorability from one survey to the next that could possibly be recorded (see online supplement for the quiz). All respondents then proceeded to the remainder of the survey. Table 5 presents the questions that delivered the dependent variables.

To discourage respondents from searching online for the true values, we reminded respondents each time we elicited change measures that we were interested in their estimates and that they would see the true values at the close of the survey.

After eliciting trends, we asked respondents for their age, gender, and party affiliation and to answer "How do you feel about Donald Trump as president?/Joe Biden" on a seven-point scale from "very negative" to "very positive".

We closed the survey by informing respondents of the true change in approval ratings. Donald Trump's approval ratings had risen four percentage points between April 2018 and April 2019. Joe Biden's favorability ratings had fallen three percentage points over the period asked about, February 2019 to May 2019.

3.2. Results

In Study 2a, two respondents estimated favorability ratings for Donald Trump that exceeded 100%; we constrained these responses to be 100%. The results reported below do not change substantively if we drop these two respondents from the sample. In Study 2a, we varied the order of the direct and indirect formats; OLS regressions on format order show that this manipulation had no effect on indirect estimates of change (t = .50, p = .616) or on direct estimates of change (t = .80, p = .426), and so in what follows we pool the data across order conditions.

3.2.1. H1: Partisan differences

Fig. 2 shows a systematic partisan effect in estimates of change and shows that the partisan difference was larger when change was elicited by the direct format than when it was elicited by the indirect format, which supports Hypothesis 1. OLS regressions of direct change as a function of indirect format and partisan affiliation show that a given respondent's direct format change measure is significantly predicted by their party affiliation even after controlling for their indirect format change measure (Trump: $b = 11.54$, $t(1170) = 6.31$, $p < .001$; Biden: $b = -11.16$, $t(504) = 4.97$, $p < .001$).

3.2.2. H2: Indirect level predicts discrepancies between the direct and indirect format

Table 6 describes the pattern of discrepancy between a respondent's direct format estimate of change and that same respondent's indirect format estimate of change. Each regression describes the results of an OLS model of the direct format estimate of change that controls for indirect format estimate. Normatively, it should be the case that the direct format estimate of change is only predicted by the indirect format estimate of change because the direct and indirect measures are logically equivalent—both are the same respondent's estimate of the change in a political leader's approval over some given time period. In violation of that normative standard, Models 2 and 4 report the results of our preregistered tests of Hypothesis 2.[3] They demonstrate that, over and above indirect estimate of change, Indirect Level also predicts the respondent's direct format estimate. In both cases, the addition of Indirect Level leads to a gain in model fit (R^2 increases from .10 to .19 for Trump judgments and increases from .40 to .44 for Biden judgments). This suggests that the same cues that inform evaluations of absolute performance are also used to inform evaluations of change in performance over time, which is the key prediction of attribute substitution.

Further evidence that indirect level and direct estimates are informed by similar cues is that indirect level and direct estimates are correlated with one another (Trump: $r = 0.31$; Biden: $r = 0.63$). In line with our theory that the redundant cues introduced by attribute substitution are subtracted out of indirect estimates, indirect estimates are not positively correlated with indirect level (Trump: $r = 0.02$; Biden: $r = -0.09$).

3.2.3. Exploratory analysis—Accuracy of forecasts

To gauge the accuracy of forecasts, we calculate the square of the difference between each respondent's estimate of change and the actual percentage point change. Averaging this across the sample gives the mean squared error (MSE). To facilitate interpretation, we report here the square roots of the MSEs (Trump $_{\text{indirect}} = 10.45$ vs. Trump $_{\text{direct}} = 18.37$; Biden $_{\text{indirect}} = 16.27$ vs. Biden $_{\text{direct}} = 27.72$). So, when estimating the change in Trump's approval rating, on average respondents were 10 percentage points off when answering by the indirect format but were 18 percentage points off when answering by the direct format. To test whether these differences in mean squared errors across the direct and indirect formats are reliably different from one another, we first put each through a Box–Cox transformation to correct for skew and kurtosis in the raw data. We then ran these normalized variants through paired t-tests. The resultant t-tests show that the indirect format has significantly smaller mean squared errors than does the direct format (Trump: $t(1130) = 12.28$, $p < .001$; Biden: $t(501) = 2.57$, $p = .011$).

3.3. Discussion

The results of Studies 2a and 2b imply that Republicans and Democrats draw on partisan-tainted cues when forming any evaluation of the candidate—be that the candidate's past performance, their current performance, or, crucially, their perceived change in performance over time. Logically, any content that is common to both past and current performance is superfluous to explaining change in performance. Nevertheless, participants' direct judgments of change depended on the overall level of

[3] Preregistration for Study 2a: https://aspredicted.org/C6X_MNN and for Study 2b: https://aspredicted.org/M3P_BCL.

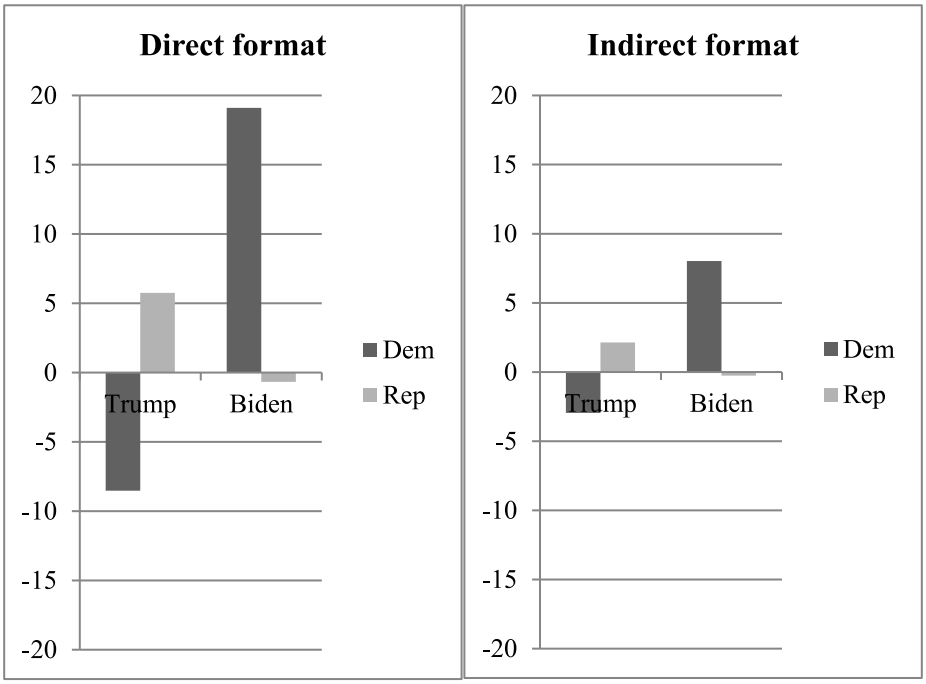

Fig. 2. Estimates by Republicans and Democrats of the change in opinion poll ratings for then-President Trump over the year from March 2019 to March 2020 and for Joe Biden over the three months from February to May 2020.

Table 6
Direct change is predicted by indirect change and indirect level.

	Trump study		Biden study	
	Model 1	Model 2	Model 1	Model 2
Ind. Change	0.582**	0.575**	1.110**	1.144**
	(0.052)	(0.049)	(0.061)	(0.059)
Ind. Level		0.584**		0.417**
		(0.050)		(0.065)
Constant	2.139**	−25.485**	4.927**	−16.083**
	(0.759)	(2.453)	(1.142)	(3.476)
R^2	0.10	0.19	0.40	0.44
N	1,175	1,175	508	508

Notes: Results from OLS regressions of *Direct Change*$_i$, where *Direct Change*$_i$ is respondent *i*'s raw response to the direct format question. *Indirect change*$_i$ is their current estimate minus their past estimate, and *Indirect level*$_i$ is measured as the mean of respondent *i*'s current and past estimates. * $p<0.05$; ** $p<0.01$.

their evaluations. The indirect format avoids this problem—it uses subtraction to cancel out the superfluous content of those cues that inform both past and current evaluations. In this way, the indirect format reduced partisan bias in trend judgments. Indirect judgments were also more accurate in this instance. We build on this result in Study 3, which adds several data points that support our thesis that attribute substitution is a driver of partisan differences in forecasts.

4. Study 3: Within-sample forecasts of the economy over the coming year

Our final study includes elements of Studies 1 and 2, and also additional features, including modifications to the question wording and response scales. In this way, Study 3 provides an opportunity for replication and also tests the robustness of our earlier results to variations in question wording and response scales. Like Study 1, it elicits forecasts of the direction of change in the economy over the coming year. Like Study 2, we employ a within-participant design —each participant contributes both a direct and an indirect forecast.

Study 3 was conducted in June 2012, when objective economic indicators showed that the United States was continuing to recover from recession. Unemployment fell over the time period participants were asked to forecast, from 8.2% in June 2012, when the survey took place, to 7.6% a year later (Bureau of Labor Statistics, 2013).

4.1. Method

4.1.1. Participants

Between June 8 and 10, 2012, we recruited participants from the same panel provider as was used for Study 1. We aimed to recruit 500 participants, a sample of size similar

to that of the Michigan Survey of Consumers. A total of 651 respondents consented, of whom 17 failed the attention filter and did not advance to the main study. Atotal of 610 respondents reported forecasts for the economy by both the direct and indirect formats, and they are the subject of the analysis that follows.

4.1.2. Procedures

Immediately after the attention filter, participants were assigned to one of four conditions that counterbalanced the order of the questions. Some participants first answered the questions about the US economy and then those about household finances, whereas others saw the reverse order. Importantly, some participants answered the indirect format first for each target, whereas others started with the direct forecasts. The two indirect format questions both appeared on the same screen, and on a separate screen than the direct question.

We used more granular response scales than in Study 1 and deviated further from the Michigan Survey's question wordings. We made these choices because this new wave of data collection offered an opportunity to apply question wordings that are more precisely comparable across the direct and indirect formats than the question wordings used in Study 1.

The direct format question about the US economy read as follows:

Would you say that a year from now, business conditions in the United States will be better or worse than they are these days?

[Answered on a 7-point vertical scale ranging from "*much worse a year from now compared with business conditions these days*" (top) through "*same a year from now as business conditions these days" to "much better a year from now compared with business conditions these days*" (bottom).]

The corresponding indirect format questions read as follows, and were presented together on a single screen:

Please enter ratings between 0 and 100 for the questions below, where 100 is the best state of business conditions that you can imagine and 0 is the worst.

How would you rate business conditions in the United States these days?

How would you rate the business conditions that you expect for the United States a year from now?

[Answered by typing a number from 0 to 100 into a textbox.]

Then we elicited indirect and direct forecasts for the other target (own household finances or the US economy). Table S4 in the online supplement reports the complete survey flow, which includes follow up questions on respondents' social and political attitudes.

4.2. Results

4.2.1. Partisan differences

We code forecasts as better, same, or worse using the same criteria as in Study 1. An indirect forecast is coded as worse if the respondent gives a lower rating to business conditions a year in the future than to current business conditions; a direct forecast is coded as worse if the respondent answered by clicking on one of the three responses that used the term "worse". In the statistical results that follow, forecasts are treated as a three-level (−1 = worse, 0 = same, 1 = better) variable and analyzed using OLS regressions. Just as in Study 2, we calculate the indirect level variable for each respondent by taking the mean of their ratings for current and future business conditions.

Because each respondent made two forecasts of the economy and it was randomly determined whether respondents were first asked by the indirect or direct format, we can replicate the between-sample analysis in Study 1 by comparing the first forecasts respondents made. Fig. 3 replicates the analysis of partisan differences performed in Study 1 and limits the sample to just those identified by the panel provider as Republican or Democrat. The partisan gap is reduced with the indirect method, which supports Hypothesis 1. We formally tested this by running an OLS regression on the economic forecasts made by partisans (−1 = worse; 0 = same; 1 = better), in which the independent variables are format (1 = direct, 0 = indirect), partisan affiliation (1 = Republican; 0 = Democrat), and their interaction. The model additionally controls for the economy/household ordering manipulation. The coefficient on the party affiliation × format interaction term indicates that the indirect format significantly reduces the partisan gap in economic forecasts relative to the direct format ($b = -0.37$, $t(349) = 2.03$, $p = .043$; see Table S5 in the Appendix).

4.2.2. H2: Indirect level predicts discrepancies between the direct and indirect formats

Study 2 supported our Hypothesis 2 by showing that that the indirect level explains discrepancies between direct and indirect format judgments of trend. When the indirect level is low (e.g., poor polling results, poor economy), people often judge that things will get worse, irrespective of the trend indicated by the two indirect judgments. Here we repeat this test using the within-participant data collected for Study 3 and we use data not merely from partisans but from the sample as a whole ($n = 610$). The model in Table 7 regresses direct forecast (worse = −1, same = 0, better = 1) on indirect forecast (worse = −1, same = 0, better = 1) and on indirect level (scored 0–100). It supports the crucial prediction that a respondent's indirect level variable explains the discrepancy between their direct forecast and their indirect forecast.

Note that there are two reasons that the coefficient on *indirect level* looks small. The first is mechanical—the coefficient on indirect level captures the effect of a one-unit change in a variable that was measured on a 0–100 scale. The second is that, unlike in Study 2, indirect level is positively correlated in these data with indirect forecast ($r = 0.28$). Hence, the coefficient on indirect level is measuring just the residual explanatory power of the level variable, net of what has already been captured in the coefficient on indirect trend. A clean measure of the magnitude of indirect level's contribution to predicting direct forecasts is captured by the gain in fit that results from adding level to the model; R^2 increases from 0.58 in Model 1 to 0.62 in Model 2.

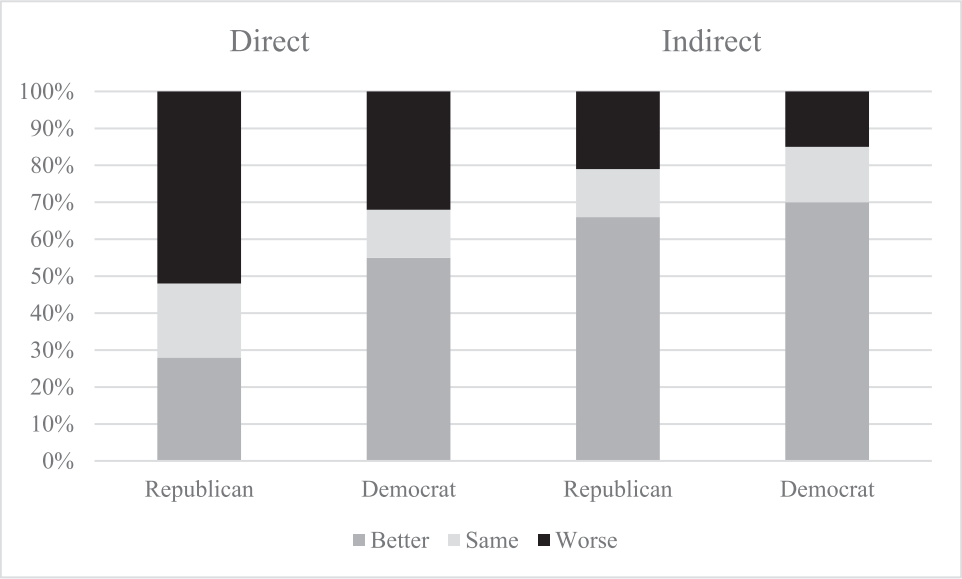

Fig. 3. Economic forecasts made by partisans ($n = 353$) from Study 3.

Table 7
Direct forecast is predicted by indirect forecast and indirect level.

	Model 1	Model 2
Indirect Forecast	0.811**	0.728**
	(0.028)	(0.029)
Indirect Level		0.009**
		(0.001)
Constant	−0.135**	−0.582**
	(0.026)	(0.062)
R^2	0.58	0.62
N	610	610

Notes: Results from OLS regressions of *Direct Forecast*$_i$. Model 1 is specified as *Direct Forecast*$_i$ = *Indirect Forecast*$_i$, where *Direct Forecast*$_i$ is the three-level variable, −1 = worse, 0 = same, 1 = better. *Indirect Forecast*$_i$ is the three-level variable, −1 = worse, 0 = same, 1 = better implied by respondent *i*'s indirect format evaluations and *Indirect level*$_i$ is measured as the mean of respondent *i*'s future and current evaluation. * $p<0.05$; ** $p<0.01$.

Table S7 in the online supplement demonstrates that Hypothesis 2 is also supported for the other target asked about in Study 3, household finances. Additionally, Table S3 demonstrates that Hypothesis 2 is supported even when we control for polynomials of indirect trend; indirect level continues to robustly predict direct forecasts.

4.3. Discussion

Study 3 confirms the patterns of results that were documented in Studies 1 and 2. Just as in Study 1, the indirect format delivered forecasts that are less equivocal and less partisan-tainted than the direct format. Just as in Study 2, discrepancies across a respondent's direct and indirect forecasts were explained by superfluous content, information that was common to that respondent's evaluations of both current and future business conditions (as captured by the indirect level variable).

Further, Study 3 rules out deliberate partisan cheerleading as the mechanism that causes the direct and indirect formats to differ. Whereas cheerleading during Obama's presidency would have caused Democrats to positively bias their responses to the direct question "will business conditions get better?" these results show that Democrats were *less* optimistic when answering by the direct format than when answering by the indirect format.

5. General discussion

Survey questions that use a direct format to ask for judgments of change over time are commonplace and influential. Yet they also show some anomalies—partisan bias is one; another is extrapolation bias, the tendency to make overly optimistic forecasts when things are good and to make overly pessimistic forecasts when things are bad. We have made the case that both phenomena are symptoms of attribute substitution. We theorize that respondents answer the question "has the economy gotten better or worse since a year ago?" as though asked something like "how do I evaluate the economy right now?" Hence, even when all indicators show the economy to be recovering, there is a tendency to answer that the economy has worsened and there is a tendency for partisans to report divergent trends.

Our attribute substitution mechanism predicts that an indirect format, which unpacks the direct format question into its constituent parts, should reduce both partisan bias and extrapolation bias in forecasts. Regarding partisan bias, each of the studies presented here shows that the trends reported by Democrats and Republicans were more similar to one another when elicited by the indirect format than when elicited by the direct format. The results of our studies also rule out that the reduction in partisan differences induced by the indirect format is symptomatic

of noise—whereas a noisier format would deliver more equivocal trend judgments than a less noisy one, in our studies it was the direct format that was more equivocal.

Regarding extrapolation bias, our results echo those of Giladi and Klar, who found that that evaluations of absolute performance can contaminate comparative judgments (1997; 2002). Just as Giladi and Klar (2002) found that participants who were comparing within a set of inferior items were biased toward reporting that any randomly selected item was worse than the average member of its set, so our studies found that respondents were biased toward reporting that the economy had worsened when the economy was weak but recovering.

The indirect format showed greater consensus that the US economy had improved over the past year (Study 1) and also that it would improve over the coming year (Studies 1 and 3). It turns out that the US economy did indeed improve over the periods asked about. Over the year respondents were asked to hindcast in Study 1, unemployment fell from 9% to 8.2%. Similarly, the periods asked about in the forecasting questions from Studies 1 and 3 also saw substantial reductions in the number of people unemployed.

Though it is encouraging that the indirect format delivered trend judgments that better correspond to realized outcomes than the direct format, one limitation of the current research is that we asked about just five outcomes that lend themselves to tests of accuracy—hindcast and forecasts of the US economy and hindcasts of Biden's and Trump's opinion poll ratings. We look forward to future research that asks about a sufficiently large array of outcomes so that statistically meaningful tests for the accuracy of trend judgments can be undertaken. In the meantime, researchers who are wondering whether to use the direct format to elicit judgments of trend might find it informative to refer to Fig. 2. When answering by the direct format in Study 2b, Democrats in our sample estimated that Biden's favorability rating had risen by 19 percentage points. When answering by the indirect format, these very same Democrats made a more reasonable estimate. They were still wrong; Biden's favorability rating had in fact fallen slightly over the 3 months asked about. But it may prove useful that respondents became *less wrong* when answering by the indirect format than they had been when answering by the direct format.

Another limitation of this study is that it gives only a high-level answer to the question of mechanism. Attribute substitution can be thought of as a family of heuristics that encompasses the affect heuristic, the availability heuristic, the diagnostic expectations model of Bordalo et al. (2018), and others. As pointed out in the *Introduction*, attribute substitution occurs when people answer a difficult question as though they had been asked an easier but related question. In the current research, we know which question respondents were asked but we remain agnostic as to which question the respondents answered. Candidates include "how does *x* make me feel?" (the affect heuristic) or "what comes to mind when I think about x?" (the availability heuristic). What was demonstrated here is that the question that respondents asked themselves when formulating their direct format trend judgment overlaps with the question they asked themselves when formulating their judgments of absolute level; hence in Studies 2 and 3 a respondent's indirect level variable explains their direct format judgment.

To the extent that perceptions of trend are consequential, further investigating the judgment processes that underpin perceptions of trend is an important research project. For instance, Mastroianni and Dana (2022) demonstrate that people's support for policies are causally influenced by their perceptions of trend in public opinion regarding those policies. Clarifying the specific form of attribute substitution that distorts trend judgments is a useful exercise because it could help communicate trends more accurately and thereby inform public opinion and voting intentions.

Our research offers insight on ideological bias in perception. There are some objective phenomena about which people of differing ideologies have divergent views. *Is the climate changing? Is voter fraud on the rise? Has the economy improved in the past year?* These are empirical questions. Nevertheless, public opinion is polarized, even though the questions themselves are not inherently political. The current research suggests that attribute substitution might explain why these sorts of issue manifest such large partisan differences.

An important question for future research is whether attribute substitution also occurs in spontaneous judgments of trend and forecasts. For instance, a would-be investor should ask herself by how much the asset under consideration is likely to increase in value. There is evidence from the stock market that is consistent with the reliance on level documented here: when the market is strong, investors are especially likely to overestimate future returns (Cassella & Gulen, 2018). We look forward to future research on this and related questions.

6. Conclusions

Our research demonstrates that direct survey formats for judgments of change over time exhibit anomalies due to attribute substitution, as seen in partisan bias and economic trend misreporting. The indirect format is designed to be more resilient to extraneous partisan influences and, in our limited data, aligns better with realized outcomes. Our findings suggest that the indirect format has potential for improving the accuracy of judgments of trend.

CRediT authorship contribution statement

David A. Comerford: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Jack B. Soll:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2024.11.005.

References

Bartels, L. M. (2002). Beyond the running tally: Partisan bias in political perceptions. *Political Behavior*, *24*(2), 117–150.

Blake, A. (2014). Obama just asked the are you better off question. In *He might not like the answer. washington post, September 29 2014*.

van der Bles, A. M., Postmes, T., & Meijer, R. R. (2015). Understanding collective discontents: A psychological approach to measuring zeitgeist. *PLoS One*, *10*(6), Article e0130100.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2018). Diagnostic expectations and credit cycles. *The Journal of Finance*, *73*(1), 199–227.

Budi, A., & Pamungkas, W. A. (2020). Partisanship in crisis: Public response to Covid-19 pandemic in Indonesia. *Jurnal Ilmu Sosial Dan Ilmu Politik*, *24*(1), 15–32.

Bullock, J. G., Gerber, A. S., Hill, S. J., & Huber, G. A. (2015). Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science*, *10*(4), 519–578.

Bureau of Labor Statistics (2013). Employment situation news release. https://www.bls.gov/news.release/archives/empsit_05032013.pdf.

Burns, W. J., Peters, E., & Slovic, P. (2012). Risk perception and the economic crisis: A longitudinal study of the trajectory of perceived risk. *Risk Analysis: An International Journal*, *32*(4), 659–677.

Cassella, S., & Gulen, H. (2018). Extrapolation bias and the predictability of stock returns by price-scaled variables. *The Review of Financial Studies*, *31*(11), 4345–4397.

Chambers, J. R., & Windschitl, P. D. (2004). Biases in social comparative judgments: the role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin*, *130*, 813–838.

Conover, P. J., Feldman, S., & Knight, K. (1987). The personal and political underpinnings of economic forecasts. *American Journal of Political Science*, *31*(3), 559–583.

Curtin, R. T. (2013). Surveys of consumers. Retrieved from: https://data.sca.isr.umich.edu/fetchdoc.php?docid=24773.

Dunning, D. (2015). Motivated cognition in self and social thought. In *APA Handbook of Personality and Social Psychology, Volume 1: Attitudes and social cognition* (pp. 777–803). American Psychological Association.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(S1), 298–320.

Giladi, E. E., & Klar, Y. (2002). When standards are wide of the mark: Nonselective superiority and inferiority biases in comparative judgments of objects and concepts. *Journal of Experimental Psychology: General*, *131*(4), 538.

Hiban, Y. (2013). Thomson reuters ends early access to key market data. https://money.cnn.com/2013/07/08/investing/homson-reuters-consumer-sentiment/index.html. Accessed November 25, 2021.

Jost, J. T., & Amodio, D. M. (2012). Political ideology as motivated social cognition: Behavioral and neuroscientific evidence. *Motivation and Emotion*, *36*(1), 55–64.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: the psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Klar, Y., & Giladi, E. E. (1997). No one in my group can be below the group's average: A robust positivity bias in favor of anonymous peers. *Journal of Personality and Social Psychology*, *73*(5), 885.

Koehler, D. (2014). The radical online: Individual radicalization processes and the role of the internet. *Journal for Deradicalization*, *1*, 116–134.

Kull, S., Ramsay, C., Subias, S., & Lewis, E. (2004). *The separate realities of Bush and Kerry supporters. The PIPA/Knowledge Networks Poll*.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480.

Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [computer software]. Available from https://quantpsy.org.

Mastroianni, A. M., & Dana, J. (2022). Widespread misperceptions of long-term attitude change. *Proceedings of the National Academy of Sciences*, *119*(11), Article e2107260119.

Meyer, A., & Frederick, S. (2023). The formation and revision of intuitions. *Cognition*, *240*, Article 105380.

Obama, Barack (2014). Retrieved from: http://www.cbsnews.com/news/president-obama-60-minutes/ accessed 1/12/17.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, *45*, 867–872.

Preacher, K. J. (2002). Calculation for the test of the difference between two independent correlation coefficients [computer software]. Available from https://quantpsy.org.

Prior, M., & Sood, K. (2015). You cannot be serious: The impact of accuracy incentives on partisan bias in reports of economic perceptions. *Quarterly Journal of Political Science*, *10*(4), 489–518.

Wagner, M., Tarlov, J., & Vivyan, N. (2014). Partisan bias in opinion formation on episodes of political controversy: Evidence from great britain. *Political Studies*, *62*(1), 136–158.