Journal of Personality and Social Psychology
2013, Vol. 105, No. 5, 740–748

0022-3514/13/$12.00 DOI: 10.1037/a0032662

COMMENT

The Impact Bias Is Alive and Well

Timothy D. Wilson
University of Virginia

Daniel T. Gilbert
Harvard University

A substantial body of research on affective forecasting has found that people often overestimate the affective impact of future events. Levine, Lench, Kaplan, and Safer (2012) argued that whereas people may overestimate the duration of their emotional responses, they do not overestimate the initial intensity of these responses as much as previous research has suggested. We suggest that Levine et al. (a) failed to review or include in their meta-analysis many studies that directly contradict their claim, (b) used a faulty classification scheme, (c) collapsed across conditions that were meant to (and did) produce opposing effects, and (d) miscoded some of the studies they did include. When these errors are corrected, their claim is clearly not supported. Levine et al. also reported the results of 4 studies, which are open to alternative explanations. The impact bias is alive and well.

Keywords: affective forecasting, impact bias, emotion, prediction

People chart their futures by imagining how they will feel as events unfold. Over the last 2 decades, a substantial body of research has examined the accuracy of these affective forecasts and has uncovered a variety of forecasting biases and the mechanisms that produce them (Gilbert & Wilson, 2007; Hahn, Ndiaye, Wilson, & Gilbert, 2011; Hsee & Zhang, 2004; Kahneman & Snell, 1992; Loewenstein, 2007; Mellers & McGraw, 2001; Wilson & Gilbert, 2003). Initial studies suggested that people overestimate how long their emotional reactions to an event will last, and that tendency was termed the *durability bias* (Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998). Soon, it became apparent that people often mispredict their initial emotional reactions to emotional events as well; thus, we changed the term to the *impact bias* to reflect the tendency for people to overestimate the initial impact and/or duration of an emotional event (Gilbert, Driver-Linn, & Wilson, 2002; see also Eastwick, Finkel, Krishnamurti, & Loewenstein, 2008).

Levine, Lench, Kaplan, and Safer (2012) raised questions about these errors and the research that identified them. They argued that people may well overestimate the duration of their affective responses but that overestimates of the initial intensity of those responses are not as large as previous research suggested and are often produced by a procedural artifact. In this article, we evaluate their claims.

The Levine et al. (2012) Argument

The crux of Levine et al.'s (2012) argument is that "people can accurately predict the intensity of their feelings about events" (p. 585), but that previous research has underestimated this accuracy because of the way in which people's forecasts have been measured. Specifically, they argued that when people are asked to predict how they will feel in general in the days or weeks or months that follow a focal event (e.g., an election or romantic breakup), they are prone to interpret this question to mean "How will you feel when you are thinking about the event?" If, after the event occurs, people are asked how happy they are in general, with no reference to the focal event, an apples-and-oranges problem exists: People think they were asked one thing ("How happy I will be about the event"), but researchers ask them something else after the event occurs ("How happy are you in general?").

This artifactual explanation is actually quite similar to one of the mechanisms that we and others have identified as a major source of affective forecasting errors—a mechanism that we have called focalism (Wilson, Wheatley, Meyers, Gilbert, & Axsom, 2000) and that others have called the focusing illusion (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2006; Schkade & Kahneman, 1998). When people are asked to predict how they will feel weeks or months after a focal event occurs, they are faced with a difficult task: They must predict not only how they will feel in the future when they are thinking about the event but also how often they will be thinking about it and how they will feel when the event is *not* focal in their thoughts. Not surprisingly, people cannot fully divorce themselves from the event that is currently in mind

Timothy D. Wilson, Department of Psychology, University of Virginia; Daniel T. Gilbert, Department of Psychology, Harvard University.

The writing of this article was supported by National Science Foundation Grant SES-0951779. We thank Matt Crawford and Derrick Wirtz for providing us with data from their studies that allowed us to compute effect sizes, Nick Sevdalis for providing us with details about his experiments, Jennifer Pattershall for sharing details of her meta-analysis of the literature, and Elizabeth Dunn and Michael Hoerger for their comments on a previous version of the manuscript.

Correspondence concerning this article should be addressed to Timothy D. Wilson, Department of Psychology, 102 Gilmer Hall, University of Virginia, Charlottesville, VA 22904-4400. E-mail: tdw@virginia.edu

and thus tend to overestimate the extent to which they will think about the event in the future and the overall impact that the event will have on their future feelings. In short, even if people understand perfectly well what they are being asked, it is difficult for them to answer that question accurately. For example, college football fans overestimated the impact of an upcoming game on their happiness in part because they overestimated how much they would be thinking about the game after it occurred (Wilson et al., 2000).

Levine et al. (2012) acknowledged that focalism can lead to forecasting errors but suggested that nonetheless the magnitude of the impact bias has been overestimated because forecasters misunderstand the question they are being asked. According to this argument, affective forecasts would be more accurate if the forecasting question matched the experience question. This could happen in two ways: First, if people interpret forecasting questions to mean how they will feel when thinking about the event, then these forecasts should more accurately predict their happiness when they are, in fact, thinking about it. Second, if people understand that the forecasting question is asking about their general level of happiness in the future, even when they are not thinking about the event, then these forecasts should better match their general happiness reports after the event occurs.

Levine et al. (2012) reported the results of four experiments and a meta-analysis of the affective forecasting literature to support their claims. They concluded that in those studies that take pains to eliminate the procedural artifact, the impact bias is reduced or even eliminated. Most of our comments will focus on Levine et al.'s literature review and meta-analysis because we believe that these present the broadest challenge to the affective forecasting literature. We then comment on the Levine et al. studies. To forecast our conclusion, we certainly do not claim that every forecasting study ever done contained perfect measures. It is possible that some participants in some studies did not understand fully what they were being asked. However, we believe that there is substantial reason to doubt that a methodological artifact has led to widespread inflation of affective forecasting errors. There are many studies that directly contradict Levine et al.'s claims, and their own studies are open to alternative explanations. Most importantly, we believe that the Levine et al. meta-analysis is seriously flawed and that the conclusions they drew from it are unwarranted.

Problems With the Literature Review

Levine et al.'s (2012) review did not include several studies that are inconsistent with their claims. In one study (Dolan & Metcalfe, 2010), for example, researchers asked participants to forecast how they would feel a week after a soccer championship game. They made the meaning of this question crystal clear: "It is very important to note that we are asking you how you will feel overall seven days after the final, rather than when you are specifically thinking about the result" (Dolan & Metcalfe, 2010, p. 734). The researchers then measured participants' affective states a week after the match. The data revealed a strong impact bias: Fans of the winning team overestimated how good they would feel 7 days after the game (Hedges's $g = .84$), and fans of the losing team overestimated how bad they would feel (Hedges's $g = 1.09$). (Generally, an effect size of 0.2 is considered to be small, an effect size of 0.5 is a medium effect, and an effect size of 0.8 is considered to be large.) Even though the forecasting question clearly matched what participants were later asked to report, a large impact bias was found.

Levine et al. (2012) also omitted a study in which affective forecasts were assessed on a nonverbal measure (Wilson, Wheatley, Kurtz, Dunn, & Gilbert, 2004, Study 3). Before finding out whether they had been chosen for a date, participants learned that they were part of a pilot program to ensure that participants left psychology studies in the same mood as the one in which they arrived. They were offered an herbal drug that was said to improve moods and were asked to select the dosage that they thought they would need if they were ultimately not selected for the date. On average, participants selected 1.29 mg of the drug. Another group of participants selected their dosage only *after* learning that they had not gotten the date. These participants selected .64 mg of the drug—an amount that was significantly lower than the dosage selected by the first group ($g = .67$). In other words, "forecasters" in this study selected more of a mood-repair drug than "experiencers" actually ended up needing, suggesting that the forecasters had overestimated how bad they would feel. This impact bias was clearly not due to confusion about the meaning of a question because on this measure there *was* no question.

Levine et al. (2012) included a study by Wilson, Meyers, and Gilbert (2003, Study 1) in their meta-analysis but failed to mention that its results directly contradict their claims. Wilson et al. asked participants to predict what their average level of happiness would be 1, 2, 3, and 4 days after the 2000 presidential election was settled, which happened on December 13, 2000, when Al Gore conceded the race to George W. Bush. On December 14, the researchers e-mailed participants and asked them to report their average level of happiness that day. On the face of it, this might seem to be an example of the problem with which Levine et al. were concerned because participants may have thought that the researchers were asking them to predict how happy they would be when they were thinking about the election, but they may not have been thinking about the election when they were e-mailed the next day. Indeed, in their meta-analysis, Levine et al. coded this study as having just such a mismatch. However, when participants in this study were emailed on December 14, the researcher clearly reminded them that he or she was conducting a study of the U.S. presidential election—and reminded them that Al Gore had conceded the night before. Thus, participants were thinking about election when they reported their postelection affect, and yet a strong impact bias emerged: Bush supporters were not as happy as they had predicted and Gore supporters were not as unhappy as they had predicted (gs = .99 and .98, respectively). In short, this study did not have the mismatch between forecasts and experiences that Levine et al. claimed is critical for producing the impact bias, and it nonetheless produced it.

Problems With the Meta-Analysis

A meta-analysis is only useful if (a) it includes all relevant studies or at least does not exclude studies that tend to show a particular pattern of results, (b) uses a meaningful classification scheme to divide studies into those that should and should not demonstrate an effect, (c) does not collapse across conditions that are meant to produce opposing effects, and (d) correctly codes and

classifies the studies it does include. Levine et al.'s (2012) meta-analysis falls short on all of these counts.

Problems With Study Inclusion

The Levine et al. (2012) meta-analysis failed to include a substantial literature on the overestimation of affective states such as fear, anxiety, and pain (for reviews, see Rachman, 1994; Rachman & Arntz, 1991). Some of these studies were done under the very conditions that Levine et al. claimed should minimize or eradicate the impact bias, and yet they produced it. For example, in one study (Rachman, Lopatka, & Levitt, 1988), participants who had been diagnosed with panic disorder underwent exposure therapy in which they completed exercises that they expected to make them afraid. Before each exercise, they predicted the maximum amount of fear they would experience, and right after each exercise, they reported the maximum amount of fear they had actually experienced. In this and many similar studies, participants overestimated the amount of fear they would experience, contrary to what Levine et al.'s argument would predict. It could be argued, of course, that studies of clinical samples should not be included in a review of the literature because they may be atypical in many respects. However, the overestimation of fear and pain has been observed in nonclinical samples as well (e.g., McMillan & Rachman, 1988). Other studies missing from the Levine et al. meta-analysis include Dolan and Metcalfe (2010); Quoidbach and Dunn (2010); Oishi, Whitchurch, Miao, Kurtz, and Park, (2009); Wang, Novemsky, and Dhar (2009); Yuan and Kring (2009); Botti and Iyengar (2004); Wilson, Meyers, and Gilbert (2001, Study 1); and Wilson et al. (2003, Study 2).

Problems With the Classification Scheme

Levine et al. (2012) coded two features of the studies they did include: (a) whether participants were asked *specific questions* or *general questions* about their affective responses to a focal event and (b) whether these questions were asked *immediately* after the event (when people were presumably still thinking about the event) or after a *delay* (when people were presumably not still thinking about the event). Levine et al. suggested that participants should be most accurate when the question was specific and when it was asked immediately after the event occurred because these are the conditions in which the measure presumably matches the participants' interpretation of the affective forecasting question (i.e., "How will you feel about the event when you are thinking about it?"). In support of their claim, they reported that the average effect size for the impact bias in the seven studies that met these criteria was close to zero and nonsignificant ($g = -.01$).

Unfortunately, delay of measurement is a poor proxy for the likelihood that participants are currently thinking about a focal event. If participants are asked how they feel the instant an event concludes, then it is probably safe to assume that they are still thinking about it—but what if they are asked 60 s later? Strangely, Levine et al. (2012) classified as "delayed measurement" any study that did not ask participants questions about their affective states the very instant the event concluded, which meant that studies in which these questions were delayed by 2 min were classified in the same way as studies in which these questions were delayed by years. Surely, people are more likely to be thinking about an election or a romantic breakup 2 min after it happened than 2 years after it happened—and this is especially likely to be true if in those 2 minutes they were continuously answering *other* questions about the event.

To illustrate the problem with this coding scheme, consider a version of the "dating game" study mentioned earlier (Wilson et al., 2004, Study 1). In this study, the researchers manipulated whether participants were asked how they felt either immediately after learning that they had not been chosen for the date or a few minutes later. (This was done by counterbalancing the order of the affect measure and a recall measure.) Levine et al. (2012) did not include this study in their meta-analysis, but if they had, the immediate condition would have been coded as one in which the impact bias *should not* occur because people were still thinking about the event, and the delayed condition would have been coded as one in which the impact bias *should* occur because people were presumably no longer thinking about the event. In fact, there was a large impact bias in both conditions of similar magnitude ($g = 1.39$ when affect was answered first, $g = 1.59$ when affect was answered second and thus delayed). This study illustrates two problems with the Levine et al. meta-analysis: First, a strong impact bias was found under the precise conditions that they argued should eliminate it (when affect was measured immediately after an event). Second, it shows that it is nonsensical to code as delayed conditions in which affect was measured a few minutes later, during which time people were asked other questions about the event and thus were surely still thinking about it.

Problems With Collapsing Across Conditions and Ignoring Moderator Variables

Research on affective forecasting has progressed rapidly in the past decade, and researchers have identified several variables that moderate, eliminate, or even reverse the impact bias. As Lepper, Henderlong, and Gingras (1999) argued, collapsing across conditions in such cases can seriously distort the results of a meta-analysis. They noted that the likelihood of reaching erroneous conclusions is "exponentially exacerbated by some authors' willingness to collapse across diametrically opposed effects" (Lepper et al., 1999, p. 672). This is especially problematic when the researchers conducting the meta-analysis are attempting to demonstrate a null effect (e.g., that the impact bias does not exist). It is deeply misleading to average across a control and experimental condition that produced opposite effects and then conclude that no effect exists.

Yet this is precisely what Levine et al. (2012) did in their meta-analysis of the affective forecasting literature. Consider a study by Dunn, Biesanz, Human, and Finn (2007) in which participants expected to have a one-on-one conversation either with their romantic partner or with an opposite-sex stranger. Participants in the forecaster condition predicted how they would feel right before this conversation, and participants in the experiencer condition reported how they actually felt right before the conversation. Dunn et al. found the standard impact bias in people's predictions about how they would feel before interacting with their partner: Forecasters overestimated how positively they would feel. Dunn et al. predicted and found the opposite of the impact bias when people predicted how they would feel before interacting with the stranger, that is, in this case, forecasters *underestimated* how

positively they would feel. Why? As the authors predicted, before conversing with the stranger, participants put on a happy face, which actually put them in a better mood than they anticipated. Thus, Dunn et al. found an interesting exception to the impact bias, a case in which it actually reverses. Both this reversal (in the stranger condition) and the impact bias (in the romantic partner condition) were significant and of large magnitude ($g = -.98$ and 1.04, respectively). How did Levine et al. code this study? They averaged across the stranger and romantic partner conditions and reported a Hedges's g of .03, implying that participants made highly accurate forecasts. This is obviously untrue. Participants in different conditions made inaccurate forecasts in different directions, exactly as the researchers had predicted.

As another example, Crawford, McConnell, Lewis, and Sherman (2002) investigated the conditions under which people experience regret after picking the wrong team in a sports wager. Participants read about a football game and learned that they would win $5 if they correctly picked the winning team. Before choosing, another participant weighed in with some advice, telling participants that they should definitely pick one of the teams. After deciding whether or not to take this person's advice and after making their choice, all participants learned that the team they picked had lost the game. Participants predicted that they would feel more regret if they rejected good advice than if they accepted bad advice, but exactly the opposite occurred. In other words, participants overestimated how bad they would feel if they rejected good advice ($g = .49$) and underestimated how bad they would feel if they accepted bad advice, ($g = -.45$). When coding this study, Levine et al. (2012) averaged across people's decisions about whether to follow or reject the other participant's advice, computing an overall Hedges's g of $-.07$, which implies that participants made highly accurate forecasts. Again, this is obviously untrue. Participants in different conditions made inaccurate forecasts in different directions, exactly as the researchers had predicted.

Levine et al. (2012) acknowledged that collapsing across moderator variables is a potential problem but argued that "the approximate size of the effects, and the inferences drawn from the analyses, do not change if these effect sizes (13) are omitted" (p. 595). This does not appear to be the case. Consider the specific-delayed category, in which participants answered specific questions about how they felt about the focal event some time after it occurred. Levine et al. reported that a total of seven studies had an average effect size of $g = .37$, which did not reach significance ($z = 1.79$, $p = .07$). But two of the seven studies tested theoretically driven conditions designed to moderate or reverse the impact bias.[1] When these two studies are removed from the meta-analysis, the average effect size of the remaining five studies is significant ($g = .53$, 95% CI [0.17, 0.89], $z = 2.86$, $p = .004$).[2]

Problems With Coding the Direction of Effects

The category in the Levine et al. (2012) meta-analysis that provides the strongest test of their hypothesis is the specific-immediate category, which includes those studies in which participants were asked specific questions about their reactions to a focal event immediately after that event occurred. According to Levine et al., this is the category in which participants' forecasts should be most accurate, and consistent with their argument, the average effect size of the impact bias in studies in this category was close to zero ($g = -.01$). We believe this is incorrect.

The seven studies (from five published articles) that fell into this category are listed in Table 1, along with the effect sizes that Levine et al. (2012) computed. Table 1 also includes the effect sizes that we computed for each study. In most cases, our calculations differ from the calculations of Levine et al. Some of these differences are minor. For example, Levine et al. computed all effect sizes as if forecasters and experiencers were separate groups of participants, despite the fact that four of the seven studies featured within-participant designs in which the forecasters and the experiencers were the same people. In those cases, we computed effect sizes using paired-group statistics, either directly from the data in published reports (Buehler & McFarland, 2001), from data provided by the researchers (Crawford et al., 2002; Wirtz, Kruger, Scollon, & Diener, 2003), or by making an assumption about the correlation between forecasted and experienced affect (Sevdalis, Harvey, & Bell, 2009). In the latter case, we used a value from a recent meta-analysis that averaged the correlation across several studies ($r = .28$; Mathieu & Gosling, 2012). When this was the only difference between our coding and Levine et al.'s, the resulting gs were similar and actually somewhat smaller in our analysis than theirs (see, e.g., the Buehler & McFarland, 2001, study).

Other differences were not minor. For four studies (three by Sevdalis et al., 2009; one by Wirtz et al., 2003), our effect-size estimates were in the opposite direction to those of Levine et al. (2012). The Appendix includes a detailed explanation of our calculations, but in brief, they differ because Levine et al. coded the results of these four studies as reversals of the impact bias when, we believe, they are clear demonstrations of it. (It is worth noting that the first authors of these articles agree with our codings of their studies; N. Sevdalis, personal communication, August 31, 2012; D. Wirtz, personal communication, October 8, 2012.)

[1] The two studies in this category that included moderator variables were Hsee and Zhang (2004), Study 3, and Koo, Algoe, Wilson, and Gilbert (2008), Study 4. The purpose of Hsee and Zhang's Study 3 was to examine conditions under which people would overpredict their emotional reactions (when participants considered two chocolate bars that differed only in size) and conditions under which they would not (when they considered their enjoyment of a pleasant and unpleasant recall task that differed qualitatively). As predicted, participants forecasted that they would be happier with the larger chocolate bar than they actually were (an impact bias) but forecasted accurately the degree to which they would enjoy the pleasant recall task more than the negative one. For some reason, Levine et al. (2012) included only the latter condition in their meta-analysis, namely, the one that hypothesized and found the absence of an impact bias. Similarly, the point of the Koo et al. studies was to show that under certain conditions people underestimate the affective benefits of counterfactual reasoning about positive events, leading to the opposite of the impact bias. Given that these studies tested theoretically driven cases in which the impact bias should be reversed or eliminated, we believe it was inappropriate to include them in a meta-analysis designed to show that the impact bias does not exist.

[2] We used the program Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2005) for all calculations, using the same assumptions as Levine et al. (2012). For example, we used random-effects models and used study, rather than subgroup, as the unit of analysis.

Table 1
Effect Sizes in Specific-Immediate Studies From Levine, Lench, Kaplan, and Safer (2012)

Study	Event	Subgroup	Hedges's *g*	
			Levine, Lench, Kaplan, & Safer (2012)	Wilson & Gilbert
Negative events				
Buehler & McFarland (2001), Study 2	Lower-than-expected grade	Failure	0.33	0.30
Crawford, McConnell, Lewis, & Sherman (2002)	Lost a wager	Did not follow advice	−0.07[a]	0.49
	Lost a wager	Followed advice	−0.07[a]	−0.45
Sanna & Schwarz (2004)	Poor grade on exam	Failure	0.40	0.39
Sevdalis, Harvey, & Bell (2009)				
Study 1	Unfavorable trade	Buyers	−0.78	0.70
Study 2	Unfavorable trade	Buyers	−0.91	0.83
Study 3	Unfavorable trade	Buyers	−1.40	1.35
Wirtz, Kruger, Scollon, & Diener (2003)	Spring break	Negative affect	−0.75	1.34
Positive events				
Buehler & McFarland (2001), Study 2	Higher-than-expected grade	Success	0.32	0.28
Sanna & Schwarz (2004)	Good grade on exam	Success	0.51	0.51
Sevdalis, Harvey, & Bell (2009)				
Study 1	Favorable trade	Owners	0.12	0.11
Study 2	Favorable trade	Owners	0.17	0.16
Study 3	Favorable trade	Owners	0.24	0.23
Wirtz, Kruger, Scollon, & Diener (2003)	Spring break	Positive affect, subjective well-being	0.75	1.33
Average effect size			−0.01	0.50

[a] Levine et al. (2012) collapsed across the subgroups of this study and computed one overall effect size.

As seen in Table 1, our codings produce very different meta-analytic results. Specifically, we find an average $g = .50$, 95% CI [0.16, 0.84], $z = 2.87$, $p = .004$, compared to Levine et al.'s (2012) estimate of $g = -.01$. The average g increases if we eliminate the Crawford et al. (2002) study that predicted and found results in the opposite direction in one condition ($g = .59$, 95% CI [0.26, 0.93], $z = 3.47$, $p = .001$).

Another approach to meta-analytic coding controversies is simply to eliminate the controversial studies from the meta-analysis. Doing so in this case would leave five studies in the specific-immediate category: Buehler and McFarland (2001, Study 2), Sanna and Schwarz (2004), and Sevdalis et al. (2009, "owner" conditions). The average effect size of these studies is significant (Hedges's $g = .35$, 95% CI [0.20, 0.51], $z = 4.41$, $p < .001$). We should reiterate that these studies represent the strongest test of Levine et al.'s (2012) hypothesis, namely, that people will make accurate affective forecasts when they are asked specific questions about how they will feel about an emotional event and when their feelings are measured right after that event. Clearly, people do not. Averaged across studies, people instead show an impact bias.

Levine et al. (2012) Studies

In addition to their meta-analysis, Levine et al. (2012) reported the results of four studies in support of their artifactual explanation. Here, we briefly review the results of these studies and some alternative interpretations of them.

Levine et al. (2012) Study 1

Participants predicted how they would feel in the days after the 2008 presidential election if Barack Obama won. Half answered a question about their general happiness ("In general, how happy will you feel?"), whereas half answered a specific question ("How happy will you feel about Barack Obama being elected president?"). Then, 1–3 days after the election, participants answered a general happiness question or a specific one about emotional reaction to the outcome of the election. The first finding of interest is that participants made similar forecasts on the general and specific questions: Obama supporters made equally positive predictions on the general and specific questions, and McCain supporters made equally negative predictions on the general and specific questions. But an impact bias was found only for predictions on the general questions. On these questions, Obama supporters overestimated how happy they would be, and McCain supporters overestimated how unhappy they would be. Participants made accurate predictions on the specific experience questions: Both Obama and McCain supporters' predictions closely matched their emotional reports about the election outcome.

As Levine et al. (2012) noted, there are two interpretations of this finding. One is focalism: Forecasters may have overestimated how much the election would be focal in their thoughts in the days after it occurred, explaining why they answered the general forecasting question in a similar manner to the specific one. But for many participants, the election was not focal in their thoughts, explaining why people's forecasts overestimated their general happiness.

The fact that people made accurate predictions about their specific happiness is puzzling to us, given that we found different results in a similar election study. As noted earlier, Wilson et al. (2003, Study 1) asked people to predict what their average level of happiness would be after the 2000 presidential election was settled. When they contacted participants to measure their actual happiness after the election, they explicitly reminded them of the election. According to Levine et al.'s (2012) analysis, participants should

have made accurate forecasts in this study because they interpreted the forecasting question to mean "How will I feel when thinking about the election?" and they were thinking about the election when contacted later. Yet a strong impact bias was found among both Bush and Gore supporters. One difference between our study and Levine et al.'s is that we included only college students who reported that they cared about politics, whereas Levine et al. included college students without preselecting them. Perhaps our more invested participants were more likely to overestimate their affective reactions. Another difference is that we controlled for individual differences in participants' reports of their baseline level of happiness, whereas Levine et al. did not.

Levine et al. (2012) Study 3

College students were asked to predict their general level of happiness the week after they found out how they did on an exam if their grade was lower than expected, as expected, or higher than expected. In another, "clarifying context" condition, participants answered this same question but only after answering two others designed to clarify its meaning. They were asked, "During the week after you find out your grade, how happy will you feel about your grade?" and "During the week after you find out your grade, do you think your grade will affect your overall mood?" When people answered the general happiness question only, they showed a large impact bias: Their general level of happiness, measured 2–7 days after they received their grade, was not nearly as low as they predicted if they did worse than expected and was not nearly as high as predicted if they did better than expected. But when people were asked the clarifying questions first, they made more moderate forecasts on the general happiness question, and the impact bias was reduced. (It remained significant among those who did worse than expected, and although it was not significant among those who did better than expected, there were only nine participants in this condition, and the effect size was medium to large, $g = .62$.)

Levine et al. (2012) interpreted these results to mean that people in the clarifying context condition understood better the meaning of the general forecasting question and that, when they did, the impact bias was reduced. Again, there is an alternative explanation based on focalism. Research has shown that reminding people that their attention will turn to other things after the emotional event occurs defocalizes them and improves the accuracy of their affective forecasts. In those studies, people induced to think about what their daily routine would be like after an important sporting event moderated their forecasts about how happy or unhappy they would be after that event (Wilson et al., 2000). It is possible that Levine et al.'s clarifying context questions did the same thing. That is, asking people, "During the week after you find out your grade, do you think your grade will affect your overall mood?" might well have cued thoughts about what else they would be doing that week, for example, "Hmm, now that I think about it, I'll find out on that Tuesday how I did on my exam in another class, and I'm really looking forward to my date with Bob on Friday." In short, the clarifying context questions might have defocalized participants, similar to the results of Wilson et al. (2000).

Levine et al. (2012) Studies 2 and 4

Levine et al. (2012) suggested that focalism was not the correct explanation of the results of their Studies 1 and 3, rather that participants misunderstood the forecasting question about general happiness to mean how they would feel when thinking about the focal events. In Studies 2 and 4, they attempted to test this possibility directly by having people answer general or specific forecasting questions and then asking them, in a multiple-choice format, how they interpreted those questions. A high percentage of participants, Levine et al. reported, interpreted the general question to mean something specific. In Study 2, for example, participants were asked to imagine that the candidate they supported won the 2012 presidential election. Eighty-one percent of the people who were asked the general question ("In general, how happy will you feel?") interpreted this to mean "I thought the question was asking approximately how happy I will be that the candidate I support won the election" rather than "I thought the question was asking approximately how happy my overall mood will be." Study 4 found similar results. When college students answered forecasting questions about their performance on an upcoming exam, around 60% of participants interpreted a general forecasting question to have a specific meaning.

This result may simply reflect the fact that, as we have argued, people find it hard to disentangle the meaning of forecasting questions not because they are worded incorrectly but because people find it difficult to separate in their minds happiness in general from happiness about the event while they are thinking about the event. Levine et al. (2012) went on to report, however, that participants who selected the specific interpretation of the general forecasting questions predicted that the election would have more impact on them than did participants who selected the general interpretation. That is, the people who interpreted the question as asking about their general level of happiness made more moderate forecasts than those who misinterpreted it as asking about how they would feel when thinking about the election, suggesting that question misinterpretation may have inflated assessments of forecasting errors in some studies.

We suspect that readers will find these results to be the strongest in support of Levine et al.'s (2012) alternative hypothesis. A substantial percentage of people interpreted the general forecasting question to be asking about specific reactions to the focal event, and those who did made more extreme affective forecasts. It is important to note, however, that these are correlational findings and that there is an alternative explanation for them: People who believed that the events would have the biggest impact on them may have been more likely to focus on that event and believe the forecasting question was asking about it. That is, rather than a specific interpretation of an event producing a more extreme forecast, it may be that an extreme forecast produces a more specific interpretation of the question.

Summary

In the past 2 decades, dozens of studies by numerous investigators have examined the conditions under which people do and do not make accurate forecasts of their affective responses to future events (Gilbert & Wilson, 2007; Hahn, Ndiaye, Wilson, & Gilbert, 2011; Hsee & Zhang, 2004; Loewenstein, 2007; Mellers & McGraw, 2001; Wilson & Gilbert, 2003). These studies have

shown that forecasts are prone to an impact bias. Levine et al.'s (2012) review and meta-analysis of the literature led them to claim that the impact bias was in large part the result of a procedural artifact. A careful examination of Levine et al.'s article reveals that they (a) failed to review or include in their meta-analysis many studies that directly contradict their claim, (b) used a faulty classification scheme, (c) collapsed across conditions that were meant to (and did) produce opposing effects, and (d) miscoded some of the studies they did include. When these errors are corrected, their claim is clearly not supported. Further, the studies that Levine et al. reported are open to alternative explanations.

References

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). Comprehensive Meta-Analysis (Version 2) [Computer software]. Englewood, NJ: Biostat.

Botti, S., & Iyengar, S. S. (2004). The psychological pleasure and pain of choosing: When people prefer choosing at the cost of subsequent outcome satisfaction. *Journal of Personality and Social Psychology, 87,* 312–326. doi:10.1037/0022-3514.87.3.312

Buehler, R., & McFarland, C. (2001). Intensity bias in affective forecasting: The role of temporal focus. *Personality and Social Psychology Bulletin, 27,* 1480–1493. doi:10.1177/01461672012711009

Crawford, M. T., McConnell, A. R., Lewis, A. C., & Sherman, S. J. (2002). Reactance, compliance, and anticipated regret. *Journal of Experimental Social Psychology, 38,* 56–63. doi:10.1006/jesp.2001.1481

Dolan, P., & Metcalfe, R. (2010). "Oops . . . I did it again": Repeated focusing effects in reports of happiness. *Journal of Economic Psychology, 31,* 732–737. doi:10.1016/j.joep.2010.05.008

Dunn, E. W., Biesanz, J. C., Human, L. J., & Finn, S. (2007). Misunderstanding the affective consequences of everyday social interactions: The hidden benefits of putting one's best face forward. *Journal of Personality and Social Psychology, 92,* 990–1005. doi:10.1037/0022-3514.92.6.990

Eastwick, P. W., Finkel, E. J., Krishnamurti, T., & Loewenstein, G. (2008). Mispredicting distress following romantic breakup: Revealing the time course of the affective forecasting error. *Journal of Experimental Social Psychology, 44,* 800–807. doi:10.1016/j.jesp.2007.07.001

Gilbert, D. T., Driver-Linn, E., & Wilson, T. D. (2002). The trouble with Vronsky: Impact bias in the forecasting of future affective states. In L. F. Barrett & P. Salovey (Eds.), *The wisdom in feeling: Psychological processes in emotional intelligence* (pp. 114–143). New York, NY: Guilford Press.

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 75,* 617–638. doi:10.1037/0022-3514.75.3.617

Gilbert, D. T., & Wilson, T. D. (2007, September 7). Prospection: Experiencing the future. *Science, 317,* 1351–1354. doi:10.1126/science.1144161

Hahn, C., Ndiaye, D., Wilson, T. D., & Gilbert, D. T. (2011). Affective forecasting. *Oxford Bibliographies Online: Psychology*. Retrieved from http://www.oxfordbibliographies.com/view/document/obo-9780199828340/obo-9780199828340-0003.xml

Hsee, C. K., & Zhang, J. (2004). Distinction bias: Misprediction and mischoice due to joint evaluation. *Journal of Personality and Social Psychology, 86,* 680–695. doi:10.1037/0022-3514.86.5.680

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006, June 30). Would you be happier if you were richer? A focusing illusion. *Science, 312,* 1908–1910. doi:10.1126/science.1129688

Kahneman, D., & Snell, J. S. (1992). Predicting a changing taste: Do people know what they will like? *Journal of Behavioral Decision Making, 5,* 187–200. doi:10.1002/bdm.3960050304

Koo, M., Algoe, S. B., Wilson, T. D., & Gilbert, D. T. (2008). It's a wonderful life: Mentally subtracting positive events improves people's affective states, contrary to their affective forecasts. *Journal of Personality and Social Psychology, 95,* 1217–1224. doi:10.1037/a0013316

Lepper, M. R., Henderlong, J., & Gingras, I. (1999). Understanding the effects of extrinsic rewards on intrinsic motivation—Uses and abuses of meta-analysis: Comment on Deci, Koestner, and Ryan (1999). *Psychological Bulletin, 125,* 669–676. doi:10.1037/0033-2909.125.6.669

Levine, L. J., Lench, H. C., Kaplan, R. L., & Safer, M. (2012). Accuracy and artifact: Reexamining the intensity bias in affective forecasting. *Journal of Personality and Social Psychology, 103,* 584–605. doi:10.1037/a0029544

Loewenstein, G. F. (2007). Affect regulation and affective forecasting. In J. J. Gross (Ed.), *Handbook of emotion regulation* (pp. 180–203). New York, NY: Guilford Press.

Mathieu, M. T., & Gosling, S. D. (2012). The accuracy or inaccuracy of affective forecasts depends on how accuracy is indexed: A meta-analysis of past studies. *Psychological Science, 23,* 161–162. doi:10.1177/0956797611427044

McMillan, T., & Rachman, S. (1988). Fearlessness and courage in novice paratroopers undergoing training. *Personality and Individual Differences, 9,* 373–378. doi:10.1016/0191-8869(88)90100-6

Mellers, B. A., & McGraw, P. A. (2001). Anticipated emotions as guides to choice. *Current Directions in Psychological Science, 10,* 210–214. doi:10.1111/1467-8721.00151

Oishi, S., Whitchurch, E., Miao, F. F., Kurtz, J., & Park, J. (2009). "Would I be happier if I moved?" Retirement status and cultural variations in the anticipated and actual levels of happiness. *Journal of Positive Psychology, 4,* 437–446. doi:10.1080/17439760903271033

Quoidbach, J., & Dunn, E. W. (2010). Personality neglect: The unforeseen impact of personal dispositions on emotional life. *Psychological Science, 21,* 1783–1786. doi:10.1177/0956797610388816

Rachman, S. (1994). The overprediction of fear: A review. *Behaviour Research and Therapy, 32,* 683–690. doi:10.1016/0005-7967(94)90025-6

Rachman, S., & Arntz, A. (1991). The overprediction and underprediction of pain. *Clinical Psychology Review, 11,* 339–355. doi:10.1016/0272-7358(91)90112-8

Rachman, S., Lopatka, C., & Levitt, K. (1988). Experimental analyses of panic—II. Panic patients. *Behaviour Research and Therapy, 26,* 33–40. doi:10.1016/0005-7967(88)90031-9

Sanna, L. J., & Schwarz, N. (2004). Integrating temporal biases: The interplay of focal thoughts and accessibility experiences. *Psychological Science, 15,* 474–481. doi:10.1111/j.0956-7976.2004.00704.x

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9,* 340–346. doi:10.1111/1467-9280.00066

Sevdalis, N., Harvey, N., & Bell, A. (2009). Affective equilibria in the endowment effect. *Journal of Behavioral Finance, 10,* 89–100. doi:10.1080/15427560902720545

Wang, J., Novemsky, N., & Dhar, R. (2009). Anticipating adaptation to products. *Journal of Consumer Research, 36,* 149–159. doi:10.1086/597050

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 345–411). San Diego, CA: Academic Press.

Wilson, T. D., Meyers, J., & Gilbert, D. T. (2001). Lessons from the past: Do people learn from experience that emotional reactions are short lived? *Personality and Social Psychology Bulletin, 27,* 1648–1661. doi:10.1177/01461672012712008

Wilson, T. D., Meyers, J., & Gilbert, D. T. (2003). "How happy was I, anyway?" A retrospective impact bias. *Social Cognition, 21,* 421–446. doi:10.1521/soco.21.6.421.28688

Wilson, T. D., Wheatley, T. P., Kurtz, J. L., Dunn, E. W., & Gilbert, D. T. (2004). When to fire: Anticipatory versus postevent reconstrual of uncontrollable events. *Personality and Social Psychology Bulletin, 30,* 340–351. doi:10.1177/0146167203256974

Wilson, T. D., Wheatley, T. P., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology, 78,* 821–836. doi:10.1037/0022-3514.78.5.821

Wirtz, D., Kruger, J., Scollon, C. N., & Diener, E. (2003). What to do on spring break? The role of predicted, on-line, and remembered experience in future choice. *Psychological Science, 14,* 520–524. doi:10.1111/1467-9280.03455

Yuan, J. W., & Kring, A. M. (2009). Dysphoria and the prediction and experience of emotion. *Cognition & Emotion, 23,* 1221–1232. doi:10.1080/02699930802416453

Appendix

Explanation of Different Codings of Sevdalis et al. (2009) and Wirtz et al. (2003) Studies

As seen in Table 1 in the main text, we computed quite different effect sizes than Levine, Lench, Kaplan, and Safer (2012) did for some studies. Here, we explain the rationale for our computations.

Sevdalis, Harvey, and Bell (2009) Studies

In three studies, Sevdalis et al. (2009) examined people's reactions to selling or buying a consumer item (e.g., a coffee mug) as part of an investigation of the endowment effect. Participants took part in a large group, and half were randomly assigned to receive the consumer item that was clearly marked as costing £4.00. These "owners" were then randomly paired with one of the other participants ("buyers") and instructed to sell the item to the buyer for an agreed-upon amount of money. Before making the transaction, both owners and buyers predicted how happy they would be selling or buying the item at a range of prices. Then, each pair decided on a selling price, and both buyers and sellers rated how happy they were with the transaction.

A key to understanding the results is that the transactions very much favored owners. In Study 1, for example, owners said in advance that they would be willing to sell the item for a minimum of £3.17, whereas buyers said they were unwilling to pay more than £2.50. Yet buyers ended up paying an average of £4.17—more than the item was worth. It is not entirely clear why owners succeeded in selling the item for such a favorable price. There was pressure to make the trade (indeed, the pairs were required to do so in one of the studies), and for some reason, that pressure worked to owners' advantage, resulting in sales prices that exceeded what buyers had said they were willing to pay in all three studies by a large amount. Regardless of the reasons, it is clear that the outcome was favorable for owners and unfavorable for buyers.

Not surprisingly, owners predicted that they would be quite happy if they sold the item at the high price they eventually received. On a scale that ranged from −3 (unhappy) to 3 (happy), owners predicted a score of 1.57 if they got the price they actually received (averaged across three studies).As it turned out, they were not quite as happy as they had predicted. Right after the transaction, their mean happiness rating was 1.30, resulting in a small impact bias.

Also unsurprisingly, buyers predicted that they would be unhappy if they ended up buying the item for such a high price. They predicted a score of −1.20 if they purchased the item for the price they eventually paid. Yet, right after the transaction, they were not as unhappy as they had predicted: Their mean actual happiness rating was .44. We believe that this is also evidence of an impact bias: Participants predicted that they would be unhappy if an unfavorable event occurred, but once it did, they were not as unhappy as they had predicted. Levine et al. (2012), however, reported that they

> decided to code selling an object as a positive outcome because owners chose to enter negotiations and were successful. Buyers also had a choice of whether to enter negotiations, so we also coded acquiring an object as a positive outcome. (L. J. Levine, personal communication, August 30, 2012).

As a result, they coded the results for buyers as the opposite of the impact bias. As noted, however, participants were required as part of the study to enter negotiations and strongly encouraged to reach an agreement (indeed, they were required to do so in one study).

In case we were interpreting the results incorrectly, we checked with Nick Sevdalis, the first author of the study. He wrote, "I think I can see where Levine et al are coming from," but that

> The slight trouble with this interpretation is that if you look at all the paper figures the buyers' affective curves clearly indicate that the buyers thought buying at the prices they eventually did buy would make them feel worse than the owners. Indeed for these prices their forecasts were below the 0 midpoint of the scale—i.e. negative. From this point of view, the transaction was anticipated to be negative—and hence I cannot see how the outcome for them can be framed as positive on the basis of the data presented in the paper. (N. Sevdalis, personal communication, August 31, 2012)

Thus, in our coding of effect sizes, we considered the affective forecasting errors to be in the direction of the impact bias for both owners and buyers (see Table 1 in the main text).

(*Appendix continues*)

Wirtz, Kruger, Scollon, and Diener (2003) Study

Wirtz et al. (2003) asked college students to predict their affective reactions during their upcoming spring break on three measures: positive affect (e.g., items such as "sociable," "happy"), negative affect (e.g., items such as "irritated," "sad"), and overall subjective experience (e.g., items such as "I expect to enjoy my spring break"). Participants then carried with them during their spring break a personal data assistant, which beeped seven times a day. When it beeped, participants rated their actual positive affect, negative affect, and overall subjective experience on the same scales on which they had made their forecasts. On the measures of positive affect and overall subjective experience, the authors found evidence of an impact bias: Participants' experience was not as positive as they predicted it would be. Participants also overestimated the amount of negative affect they actually experienced. As the authors noted, "Participants overestimated the *intensity* of their spring break experience" (Wirtz et al., 2003, p. 522).

Levine et al. (2012) came up with an average effect size of .00 for this study. When we asked how they did so, they noted two things. First, they disregarded the measure of overall subjective experience because "the subjective experience measure did not assess emotional intensity" (L. J. Levine, personal communication, September 7, 2012). This seems like an odd decision given that that measure included items such as "I expect to enjoy spring break" and that the crux of Levine et al.'s argument is that people make accurate predictions when they are asked to predict how they will feel about an event. Second, Levine et al. coded the results on the negative affect measure as *opposite* to an impact bias, under the assumption that people felt less negative affect than they predicted in reaction to a positive event. We believe that this is a misinterpretation of the results of the Wirtz et al. (2003) study. It is the one of the very few that used experience-sampling technology to measure actual affect during an ongoing event. Even if that event was mostly positive on average, it was bound to include some negative experiences (e.g., traffic jams on the way to the beach, arguments with friends or family). Participants seem to have anticipated this, in that they predicted that they would experience negative emotions during their break to some degree. We agree with Wirtz et al. that the fact that participants overestimated the extent of these negative emotions is a clear instance of an impact bias and have coded it as such.

Received November 21, 2012
Revision received March 11, 2013
Accepted March 14, 2013 ■