Journal of Behavioral Decision Making

WILEY

RESEARCH ARTICLE OPEN ACCESS

Never … Often? Comparisons That Shape People's Likert-Type Ratings of Behavior Frequencies

Jeremy D. Strueder[1] | Jane E. Miller[2] | Isaac T. Petersen[1] | Paul D. Windschitl[1]

[1]Department of Psychological and Brain Sciences, University of Iowa, Iowa City, Iowa, USA | [2]Climate Change Research Network, Vanderbilt Law School, Vanderbilt University, Nashville, Tennessee, USA

Correspondence: Jeremy D. Strueder (jeremy-strueder@uiowa.edu)

Received: 10 September 2025 | **Revised:** 20 January 2026 | **Accepted:** 13 April 2026

Keywords: behavioral frequency judgments | health behavior | Likert-type scales | social comparison | vague verbal quantifiers

ABSTRACT

Responses to Likert-type behavioral frequency (LBF) questions often do not consistently map onto objective numerical estimates. Prior research suggests that social and other comparisons may underlie this divergence, but the relative influence of different comparison standards—and the cognitive processes supporting them—remains unclear. Across two studies, we examined how comparisons to peers, averages, experts, past selves, and conceptually irrelevant standards shape LBF responses for common health behaviors (e.g., hand washing and flossing). Participants provided LBF judgments, absolute frequency estimates, and comparative judgments for each behavior. Study 1 showed that direct comparisons predicted LBF judgments above and beyond participants' own absolute frequency estimates, with comparisons to experts and average others being especially influential. Even when controlling for shared methodological variance, all comparison types explained unique variance in LBF responses. Study 2 replicated this pattern of results. Moreover, additional analyses in Study 2 suggest that participants were not making precise, pairwise comparisons between numeric estimates, but were instead relying on more abstract, gist-like impressions of how their behavior compared to others. Together, these findings underscore the importance of considering the comparative and interpretive nature of self-report measures, particularly in contexts where behavioral frequency carries social, normative, or evaluative meaning.

Likert-type behavioral frequency (LBF) questions—such as "How often do you feel overwhelmed at work?" with response options like *never*, *sometimes*, and *always*—are a staple in psychological research (e.g., Likert 1932; Schwarz and Oyserman 2001; Tourangeau et al. 2000; see Jebb et al. 2021, for review). Moreover, they are used to measure behavior frequency in a wide range of applied contexts, such as in CDC health-related telephone surveys, consumer behavior surveys (e.g., McKinsey & Company), and health surveys conducted by the World Health Organization (e.g., the Health Behavior in School-aged Children study; see Inchley et al. 2020). A key reason for their popularity is that they are—on the surface—intuitive and easy to use for respondents, especially when compared to more cognitively demanding numeric response formats that ask respondents to recall behavior frequency in terms of absolute frequencies (Al Baghal 2014; Bradburn and Miles 1979; Burton and Blair 1991; Tourangeau et al. 2000). However, the apparent simplicity of LBF judgments masks the complexity of the cognitive processes involved in answering them (Schwarz 1999). Notably, past research shows that LBF judgments do not consistently map onto objective numerical estimates (e.g., Budescu and Wallsten 1985; Conrad et al. 1998;

Jeremy D. Strueder and Jane E. Miller contributed equally to this work.

All work was completed at the Psychological and Brain Sciences Department, University of Iowa, Iowa City, IA. Correspondence concerning this article should be addressed to Jeremy D. Strueder.

Schwarz 1999; Wright et al. 1994; Woltz et al. 2012). Some of this work has also provided insights about a key reason why: Comparisons of various types—such as social comparisons—can influence how internal estimates of behavior frequency are mapped onto LBF response scales (e.g., Robert Pace and Friedlander 1982; Schaeffer 1991; Sudman et al. 1996; Wänke 2002; Wright et al. 1994).

In the present work, we examined these comparative influences more closely. Specifically, although there are multiple types of comparisons that could shape LBF responses—such as those involving peers, average others, past selves, or expert standards—prior research has not systematically evaluated their relative contributions. Our primary goal, therefore, was to assess the relative impact of these different comparison standards in shaping LBF responses. We tested this in studies that asked participants for LBF responses regarding health-related behaviors (e.g., frequency of hand washing). In doing so, we also aimed to learn more about the nature of the comparison processes that influence LBF responses—that is, whether they reflect a more gist-like comparison process or one that involves comparisons of verbatim frequency estimates.

1 | Interpreting Likert-Type Behavioral Frequency Questions: The Role of Context and Comparison

Likert-type behavioral frequency (LBF) questions may appear simple on the surface, but a closer look reveals that they can involve a range of cognitive processes. As outlined by previously established models of survey responding (e.g., Schwarz and Oyserman 2001; Tourangeau et al. 2000), providing a behavioral frequency judgment requires respondents to navigate several stages: comprehending the question, retrieving relevant behaviors from memory, forming a judgment about frequency, and mapping that judgment onto a response scale. Although numeric frequency formats emphasize retrieval (i.e., "How many times did X occur?"), LBF items shift the cognitive emphasis toward judgment and mapping. Whatever representation that a person has or constructs about their frequency of a behavior, LBF questions require that it be mapped onto a 1-X rating scale defined by anchors like "rarely" and "often."

Range-frequency theory offers one perspective on this mapping of an internal representation onto a 1-X rating scale (Parducci 1965). The theory is a formal, domain-general model for ratings. It posits that a rating can be modeled as a compromise between two comparative components—how the target magnitude compares to the lowest and highest magnitudes in a contextually relevant set and how the target magnitude ranks among all the magnitudes in the set. For the present domain, the target magnitudes being asked about in LBF questions are frequencies, so the two comparative subcomponents would likewise involve other frequencies. In short, range-frequency theory suggests that a perceived distribution of other frequency values would be crucial in shaping what response is selected on the scale of an LBF question. More recent process-level accounts of judgment, such as the Decision by Sampling (DbS) theory (Stewart et al. 2006), similarly treat judgments as emerging from comparisons against a set of accessible reference values. Notably, however, both range-frequency theory and DbS are domain-general and agnostic about where these values are from; comparison values could be from various sources (e.g., thoughts about peers, thoughts about the self in the past, observations of strangers or media characters, and memories of expert recommendations). Moreover, the salience and differential impacts of various types of comparative referents are not determined by either theory; only an empirical approach, such as the one in this paper, can speak to this.

It is also notable that verbal terms that serve as endpoint anchors in LBF questions are typically vague (e.g., Brun and Teigen 1988; Budescu and Wallsten 1985). Words like "rarely" and "often" lack fixed meanings and are interpreted in relation to contextual and personal factors (e.g., Borgers et al. 2003; Robert Pace and Friedlander 1982; Schaeffer 1991; Wänke 2002). For example, research shows that people tend to interpret the same verbal quantifier differently depending on the behavior in question (Bradburn and Miles 1979)—for instance, "often" might correspond to a low absolute frequency when referring to the number of earthquakes a Californian experiences in a year, but a large absolute frequency when referring to the number of times that guns are fired in western movies (Pepper and Prytulak 1974). Interpretations of verbal quantifiers also vary across demographic groups: age, education, cultural background, and socioeconomic status can all shape how respondents construe verbal quantifiers (Borgers et al. 2003; Krumpal et al. 2008; Schaeffer 1991; Schneider and Stone 2016; Wright et al. 1994). This variability does not necessarily undermine the utility of LBF questions, but it highlights that their meaning is constructed dynamically and may reflect both subjective context as well as objective behavior.

A study by Wänke (2002) provides an interesting example of how salient comparison targets can change the way respondents map internal representations of behavior frequency onto LBF response scales. Students who were asked to report how frequently they go to the movies on an LBF scale provided—on average—different responses depending on whether they were told that the study was investigating the leisure activities of city residents (a comparison group with a lower base rate of movie attendance) versus fellow university students (a comparison group with a higher base rate of movie attendance). Notably, differences in reference groups did not affect absolute numeric reports, suggesting that it is the process of mapping internal estimates of absolute frequency onto LBF scales that invites comparisons. These findings are consistent with both range-frequency theory, which suggests that people use available comparison values to position their own behavior within a subjective distribution when mapping behavior onto LBF scales, and with the idea that vague verbal labels on LBF scales are interpreted flexibly in relation to contextually salient referents.

2 | Which Comparisons Have an Influence?

Despite broad agreement that comparison processes are relevant, very little is known about the relative contributions of these different points of comparison. Particularly in health contexts, it could be the case that expert recommendations serve as the primary reference point for how respondents interpret verbal quantifiers in LBF questions. Expert recommendations have been shown to impact behavior (Cummings et al. 1979; Larson and Story 2009; Michie et al. 2012; Prochaska et al. 2006), and it is reasonable to suspect that respondents may use their knowledge of expert benchmarks—if available—to inform where their behavior frequency falls along an LBF scale. Alternatively, social comparison theory (Festinger 1954) suggests that in the absence of objective standards, individuals often evaluate themselves relative to others in their social environment. As such, it is also possible that respondents primarily draw on social comparisons—including those to close peers (Alicke et al. 1995) or generalized population norms (Suls and Wheeler 2000)—to inform how frequently they engage in a behavior (Cole and Korkmaz 2013). Work on the *local dominance effect* suggests that comparisons to close peers such as friends or family could have a greater influence on judgments of relative behavioral frequency than global comparisons to the average person or a distant social group (Alicke et al. 2010; Bruchmann and Evans 2013; Buckingham and Alicke 2002; Klein 1997, 2003; Mussweiler and Rüter 2003; Zell and Alicke 2009, 2010). However, this has not been tested in a unified empirical framework and thus remains an open question.

In addition to the question of whether some types of comparisons matter more than others in shaping responses to LBF questions, there is also an outstanding question about the cognitive basis of these comparative influences. One possibility is that the relative nature of LBF questions prompts respondents to engage in pairwise comparisons of absolute representations of frequency—explicitly contrasting their own behavior with one or several recalled standards (e.g., "I floss more than my roommate but less than what my dentist recommends"). Alternatively, another possibility is that the influence of comparative processes on responses to LBF questions is grounded in a more gist-like representation of how one's behavior stacks up relative to others or to general norms. Gist-based representations of relative frequency in LBF judgments could reflect simple categorical distinctions (e.g., "more than most") that integrate various sources of comparison information without retrieval of explicit comparison standards (e.g., expert recommendations about an absolute frequency). The theoretical distinction between people making pairwise comparisons of absolute representations and people using gist representations of relative frequency is similar in nature to *fuzzy-trace theory* (e.g., Brainerd and Reyna 2002, 2004; Reyna and Brainerd 1995), which proposes that people encode and rely on both verbatim (precise) and gist (intuitive) mental representations, often favoring gist representations when making judgments under uncertainty or low-effort conditions. Currently, it is not clear which of these two cognitive processes—either making pairwise comparisons of absolute/verbatim frequencies or relying on gist representations of comparative frequency—is primarily responsible for the influence of comparisons on LBF judgments.

3 | The Current Research

The present work sheds light on the two interrelated issues just reviewed—one about the relative contributions of different points of comparison on people's LBF responses, and the other about the gist versus verbatim mechanisms underlying these contributions.

The main goal driving our first study was to examine the relative influence of different types of comparison standards on LBF responses. We examined this in the context of health-related behaviors (e.g., flossing teeth, washing hands, and eating breakfast). Participants first responded to an LBF question for each behavior. The LBF scales we used included verbal labels for the scale endpoints (e.g., "almost never" to "extremely often"). We then solicited their absolute numeric estimates of the frequency for each behavior, and we administered a series of five direct-comparative questions—each asking how often they performed the behavior relative to a different specified standard (e.g., close peers and expert recommendation). Multilevel modeling was used to examine the extent to which people's beliefs about the five possible comparison standards for a given behavior were predictive—beyond their absolute numeric estimates—of their Likert response about their own behavior frequency (their LBF).

Study 2 followed up on key results from Study 1, using almost the same paradigm, but with additional measures that helped to examine the extent to which the influence of comparisons could be attributed to people making pairwise comparisons of various absolute estimates/standards, or whether they are grounded in more gist-like representations.

The studies were preregistered and approved by the University of Iowa Institutional Review Board (#201201727). The materials, data, code, and preregistrations are available on OSF: https://osf.io/c8abu

4 | Study 1 (1A and 1B)

Study 1 was originally run as two separate studies, with 1b being a replication of 1a. Because Studies 1a and 1b were identical except for whether participants were tested in a lab or remotely, we report the studies together and will collectively refer to them as Study 1. Parts 1a and 1b were preregistered separately but identically. See the above OSF link for access to preregistrations, materials, and data.

The primary focus of Study 1 was on identifying which types of comparisons most strongly predict LBF judgments above and beyond absolute numeric reports of behavior frequency. There were five types of comparisons that we queried. Four of these were selected as theoretically meaningful standards that people might use when constructing their LBF responses: comparisons to close peers, population norms (i.e., the average person of the same age and gender), expert recommendations, and one's past self. We hypothesized that these four relevant direct-comparison judgments would shape LBF responses. Specifically, we predicted that, even when controlling for self-absolute frequency estimates of behavior, people's responses to the direct-comparison questions would significantly predict their answers on the LBF scale. We did not have specific hypotheses about which of the four main comparisons would be most predictive, though this was a main question of interest.

The fifth direct-comparison question asked participants to compare their behavior to a non-meaningful and very distant social standard: the average European. The "average European" distant social comparison was included not as a theoretically meaningful standard, but as a check on shared-method variance (Podsakoff et al. 2012). Like the other comparison questions, it was measured on a 7-point Likert-type scale, but we did not expect it to be a strong predictor of LBF responses. Rather, we included it because any shared-method variance that might make the other comparison questions predictive of LBF judgments (because all are answered on 7-point scales) should also be shared by the compared-to-average-European question. Therefore, controlling for the compared-to-average-European responses in analyses predicting LBF judgments should be useful for neutralizing the impact of shared-method variance.

4.1 | Method

4.1.1 | Participants

Our preregistered sample-size targets of 256 for Studies 1a and 1b were based on guidelines for correlational research designs suggesting that correlations become statistically stable and have reduced estimation error when $N \approx 250$ (Schönbrodt and Perugini 2013). The final sample for Study 1a consisted of 275 undergraduate students (212 females, 52 males, 11 unreported, $M_{age} = 18.92$, $SD = 1.65$) who completed the survey in a laboratory setting. The final sample for Study 1b consisted of 180 undergraduate students (112 females, 66 males, two unreported, $M_{age} = 19.05$, $SD = 1.23$) who completed the survey online. All participants received credit for a research requirement in their introductory psychology course. The sample of 180 for Study 1b differs from our preregistered sample size of 256 because the number of participants available was reduced due to restricting participants who had completed Study 1a from participating in Study 1b. All in all, this leads to a total sample size of 455 participants (324 females, 118 males, 13 unreported, $M_{age} = 18.99$, $SD = 1.44$).

4.1.2 | Behavior Types and Procedure

There were eight types of behaviors for which participants answered questions. All were health-related: washing hands, eating red meat, flossing teeth, eating breakfast, drinking alcoholic beverages, eating fish, washing one's face, and drinking caffeinated beverages.

All participants began by completing a self-report Likert-type behavioral frequency (self-LBF) question for each behavior in the order listed above. See the measures section for more information about this and other types of measures (see also Table 1 for a summary). Next, participants were randomly assigned to one of two counterbalancing conditions, which varied the order in which they completed the remaining items. In one condition, participants first provided absolute estimates of how frequently they performed each behavior, followed by five blocks of direct comparison items (with each behavior in a given block).[1] In the other condition, the order was reversed, with participants completing the direct comparison items before the absolute frequency estimates. The order in which the five blocks of direct comparison items were encountered was randomized.

4.1.3 | Measures

4.1.3.1 | Self-LBF Questions (The Outcome Measures). The main outcome measures were self-LBF questions that solicited responses on a 1–7 scale that had verbal labels on the endpoints. For example, for the question about eating red meat, participants were asked, "How often do you eat red meat?" from 1 (*Very infrequently*) to 7 (*Very frequently*). The wording of the questions, except for where the behavior was specified, was the same from behavior to behavior. Labels on the scales appeared only on the endpoints (under 1

TABLE 1 | Types of Measures used in Studies 1 and 2.

Measure type	Study	Judgment target	Response format	Example item
Self-LBF	1,2	Self	Likert-type scale	"How often do you eat red meat?"
Self-absolute frequency	1,2	Self	Numeric frequency	"How many times a month do you eat red meat?"
Direct comparison	1,2	Self vs. social group	Likert-type scale	"How often do you eat red meat compared to your close peers?"
	1,2	Self vs. expert standard	Likert-type scale	"How often do you eat red meat compared to what experts recommend?"
Other-absolute frequency	2	Social group	Numeric frequency	"How many times a month does an average person your age and gender eat red meat?"
	2	Expert standard	Numeric frequency	"How many times a month do experts say a person your age and gender should eat red meat?"

Note: Overview of the types measures used in Studies 1 and 2. Likert-type questions asked participants to report frequency using a verbal scale (e.g., "almost never" to "extremely often"), whereas absolute frequency questions required numeric estimates that could be selected from a drop-down menu. "Self" items refer to judgments about the participant's own behavior; "social group" items refer to judgments about/with regard to a social reference group (e.g., close peers and the average person); "expert standard" items refer to judgments about/with regard to an expert standard or recommendation.

and 7). The specific labels on those endpoints differed slightly for the different behaviors. For instance, the flossing teeth labels were (1) *I do not floss my teeth* and (7) *I floss my teeth a lot*. Full details of all questions and response options are available in the survey materials posted on OSF (https://osf.io/c8abu).

4.1.3.2 | Self-Absolute Frequency Questions.

These questions solicited a participant's numeric estimate of their own absolute frequency of a behavior in a specified time period. For example: "How many times a [month] do you [eat red meat]?" This particular example item had a drop-down menu ranging from 0–40 times a month, but the time period and response option range differed slightly from question to question based on the behavior. See survey materials on OSF for item-specific details (link above). Because self-absolute frequency estimates tended to be highly positively skewed, estimates were log transformed for analyses (with a constant of 1 added to accommodate zeros) to reduce skew and better approximate normality.[2]

4.1.3.3 | Direct-Comparison Questions.

In both studies, there were five types of direct comparison measures for each behavior: a global comparison, a local comparison, a comparison against the perceived expert standards, a prior-self comparison, and a comparison against a distant social group (Europeans). Each of these direct comparisons was answered on a scale from −3 to +3 with labels at the endpoints and midpoint—for example, −3 (*I eat significantly less than* …), 0 (*I eat about the same amount as* …), to +3 (*I eat significantly more than* …). *Direct global comparison* questions (for red meat) asked, "Think about the average person your age and gender. How often do you eat red meat compared to the average person?" *Direct local comparison* questions asked, "Think about your close peers (people you see or hang out with often). How often do you eat red meat compared to your close peers?" *Direct expert comparison* questions asked, "Think about the amount of red meat experts recommend eating. How often do you eat red meat compared to what experts recommend?" *Direct prior-self comparison* questions asked, "Think about the amount of red meat you ate at this time a year ago. How often do you eat red meat now compared to a year ago?" *Direct distant social* group questions asked, "Think about the average European person. How often do you eat red meat compared to the average European?"

5 | Results

5.1 | Preliminary Analyses

To examine basic associations, we computed zero-order correlations—separately for each behavior—among the main variables: the self-LBF judgments, self-absolute frequency estimates (transformed as described above), and five direct comparisons. We then computed averages of these correlations across the eight behaviors—using Fisher z-transformations, which were then back-transformed after averaging. Table 2 displays these average correlations. All predictor variables were moderately to strongly positively correlated with LBF estimates, suggesting that participants who gave higher self-LBF estimates also tended to report higher self-absolute frequency estimates and higher direct comparison estimates.

We also examined intercorrelations among the five direct-comparison variables to assess the potential for multicollinearity in regression models. Correlations ranged from $r = 0.232$ to 0.752 across behaviors, with the strongest associations observed between local and global comparisons. Variance inflation factors (VIFs) from a multilevel model including all five comparisons ranged from 1.13 to 3.32 (average VIF = 2.27), which is at or below the conservative cutoff of 3 (Garson 2012; O'brien 2007).

5.2 | Main Analyses

We will now report a series of mixed-effects analyses intended to comprehensively examine the influence of comparison-based reasoning on LBF judgments. Given that the primary outcome of interest was item-level self-LBF judgments (i.e., one self-LBF judgment per participant per behavior), we employed a multilevel modeling approach to account for the nested structure of the data—specifically, repeated judgments nested within participants and behaviors. This approach allowed us to model variability across the eight target behaviors while appropriately accounting for individual differences. All models were cross-classified multilevel models estimated using the lmer() function from the lme4 package in R (Bates et al. 2015) with random intercepts specified for both participant and behavior.[3] To facilitate interpretation and comparability

TABLE 2 | Average correlations between LBF estimates, self-absolute frequency estimates, and direct comparisons in Study 1.

		Average correlations					
Measure	**Range**	**1**	**2**	**3**	**4**	**5**	**6**
1. Self-LBF judgment	1 to 7	—					
2. Self-absolute frequency	0 to 4.33	0.759	—				
3. Expert comparison	−3 to 3	0.694	0.650	—			
4. Global comparison	−3 to 3	0.719	0.667	0.702	—		
5. Local comparison	−3 to 3	0.660	0.611	0.626	0.752	—	
6. Prior-self comparison	−3 to 3	0.300	0.287	0.285	0.298	0.266	—
7. Distant comparison	−3 to 3	0.616	0.560	0.624	0.651	0.581	0.232

Note: All of the average correlations are significant at $p < 0.001$. Correlations were averaged across the eight health behaviors using Fisher z-transformation and back-transformed for reporting. As described in the Measures section, the self-absolute frequency estimates were log transformed before going into the analyses.

across predictors and outcomes, all continuous variables—including the outcome variable (self-LBF judgments), the log-transformed absolute-frequency estimates, and the direct comparisons—were standardized prior to analysis. Absolute frequency estimates were also log transformed prior to standardization to correct for positive skew; a constant of 1 was added before transformation to accommodate zero values. To assess whether direct comparisons accounted for additional variance in LBF judgments beyond what was explained by absolute frequency estimates, we compared nested models using likelihood ratio tests. We also examined changes in marginal R^2 (variance explained by fixed effects) to quantify the incremental predictive value of each set of predictors. Marginal R^2 values were obtained using the r.squaredGLMM() function from the MuMIn package (Nakagawa and Schielzeth 2013; Nakagawa et al. 2017).

5.2.1 | Do Comparisons Add Predictive Power Beyond Absolute Frequency Estimates?

We start by examining whether the inclusion of the direct comparisons (global, local, expert, previous self, and distant social group) adds predictive power above and beyond self-absolute frequency estimates for predicting self-LBF judgments. To do this, we first fit a base model with self-absolute frequency estimates as the sole predictor of self-LBF judgments. As expected, self-absolute frequency estimates were a strong predictor of self-LBF judgments, $\beta = 0.78$, $SE = 0.01$, $t(3315.39) = 67.23$, $p < 0.001$, *95% CI* [0.75; 0.80]. For a one standard deviation increase in self-absolute frequency (for a given behavior), there was a 0.78 standard deviation increase in self-LBF judgments. We then fit a model in which the direct-comparison judgments were also included as predictors. Adding the direct comparisons significantly improved model fit, $\chi^2(5) = 1191.17$, $p < 0.001$. The marginal R^2 increased from 0.55 to 0.68 with the inclusion of the four direct comparison measures, indicating that the direct comparisons accounted for an additional 13% of the variance in self-LBF judgments. These results confirm that comparisons do indeed shape responses to LBF questions above and beyond estimates of absolute frequency.

5.2.2 | Which Specific Comparisons Add Predictive Power Beyond Absolute Frequency Estimates?

Having confirmed that direct comparisons inform responses to LBF questions above and beyond absolute estimates of frequency, our next main goal was to examine the relative influence of different types of comparison standards on responses to LBF questions. Figure 1 plots the model coefficients of the model including self-absolute frequency estimates and all direct comparisons as predictors. As expected, self-absolute frequency estimates were still a strong predictor ($\beta = 0.37$, $SE = 0.01$, $t(3311.00) = 25.62$, $p < 0.001$, 95% *CI* [0.34, 0.40]) of LBF judgments. Importantly, all five direct comparisons were also significant predictors of self-LBF judgments. Specifically, aside from self-absolute frequency estimates, participants' self-LBF judgments were most strongly predicted by global comparisons to the average person ($\beta = 0.19$, $SE = 0.02$, $t(3271.49) = 11.97$, $p < 0.001$, 95% *CI* [0.16, 0.22]) and by comparisons to expert benchmarks ($\beta = 0.17$, $SE = 0.01$, $t(3285.66) = 11.47$, $p < 0.001$, 95% *CI* [0.14, 0.19]). Peer comparisons were also a notable predictor ($\beta = 0.10$, $SE = 0.01$, $t(3275.79) = 7.60$, $p < 0.001$, 95% *CI* [0.08, 0.13]). Surprisingly, distant social comparisons were similarly predictive ($\beta = 0.11$, $SE = 0.01$, $t(3328.35) = 8.55$, $p < 0.001$, 95% *CI* [0.08, 0.13]). Lastly, comparisons to the previous self were also significantly predictive of LBF judgments, but to a much lesser extent ($\beta = 0.03$, $SE = 0.01$, $t(3321.81) = 3.67$, $p < 0.001$, 95% *CI* [0.02, 0.05]).[4]

5.2.3 | Does Each Comparison Explain Unique Variance?

In addition to comparing effect sizes across direct comparison predictors, we also ran a series of analyses in which we assessed

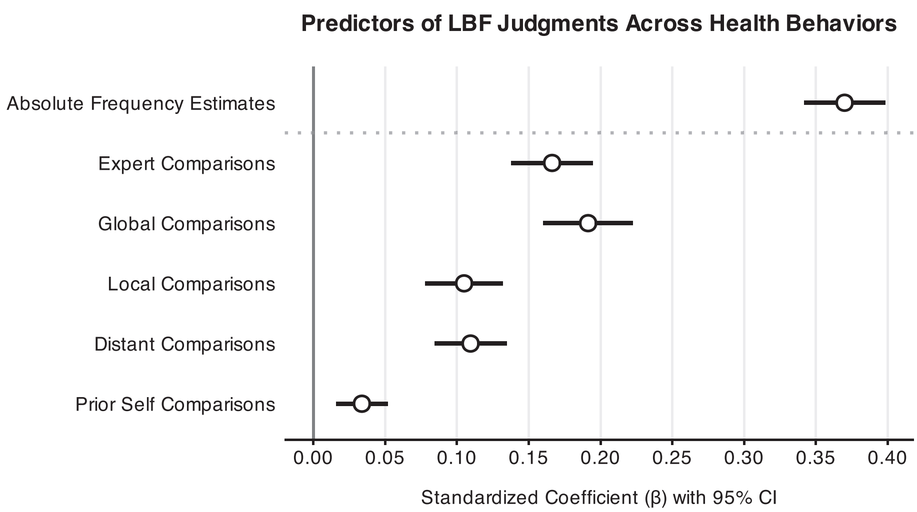

FIGURE 1 | Study 1. Model coefficients of the cross-classified multilevel model including self-absolute frequency estimates and all direct comparisons as predictors. Error bars represent 95% confidence intervals based on profile-likelihood estimation. Dotted horizontal line separates absolute frequency estimates from direct comparisons.

the extent to which each direct comparison measure predicts variance above and beyond self-absolute frequency estimates. Namely, we fit five models, each with self-absolute frequency estimates and one direct comparison measure as predictors. We then compared these five models to the base model with only self-absolute frequency estimates as a predictor to assess how much—if any—additional variance a given direct comparison measure explained. Table 3 summarizes these models.

All five direct comparison models provided a better fit than the base model, in which only self-absolute frequency estimates were included (all p values < 0.001). Within each model, the direct comparison measures were all significant predictors (all p values < 0.001). Looking at changes in marginal R^2 values, we found that global comparisons explained the most additional variance with 9.01%. Comparisons to expert benchmarks and comparisons to peers explained 5.39% and 6.14% of additional variance, respectively. Lastly, comparisons to a distant social group explained 6.53% of additional variance, whereas comparisons to the previous self only explained 0.49% of variance above and beyond self-absolute frequency estimates.

5.2.4 | How Much of the Predictive Power of Direct Comparisons Comes From Conceptual vs. Methodological Contributions?

Contrary to our a priori expectations, previous analyses showed that all types of direct comparisons—including those to a distant social group (the average European)—predicted self-LBF judgments above and beyond self-absolute frequency estimates. One possibility is that this broad pattern of predictive value reflects methodological overlap: Both the self-LBF judgments and the direct comparison measures used similar response formats (e.g., 7-point Likert scales), which could inflate associations due to shared-method variance. To examine whether methodological similarity alone could account for these effects, we conducted a series of follow-up analyses. Specifically, we tested whether each conceptually relevant direct comparison (global, local, expert, and previous self) explained additional variance in self-LBF judgments beyond that explained by responses to the absolute frequency question and to the direct-comparison question about the distant social group. Our reasoning was that if a given comparison measure only predicts LBF judgments due to methodological similarity (e.g., shared scale format), then it should not improve model fit once another methodologically similar predictor is already included. Conversely, if it does explain additional variance, this would indicate a unique conceptual role in shaping LBF judgments. The reason we used the distant social group variable rather than one of the other direct-comparison variables as (essentially) a control variable in this analysis was grounded in our preregistered plans; we had anticipated that there would be no conceptual connection between how people responded to the self-LBF questions and the direct comparison with the average-European question, but there could be shared-method variance.[5]

Table 4 summarizes the results of these analyses. In each case, the inclusion of the additional direct comparison significantly improved model fit (all $ps < 0.001$), and each relevant direct comparison was a significant predictor within its respective model (all $ps < 0.001$). Examining changes in marginal R^2, we found that global comparisons contributed the largest additional increase at 4.51%, followed by the local (2.90%) and expert (2.24%) comparison predictors. Although comparisons to one's previous self also significantly improved model fit, the increase in marginal R^2 was modest (0.31%). Together, these results indicate that the predictive value of these conceptually relevant comparisons cannot be fully explained by methodological similarity to the outcome measure. That said, although each relevant comparison measure explained unique variance in self-LBF judgments beyond both absolute frequency estimates and the distant social group comparison, the magnitude of the additional explained variance was notably smaller than in our earlier analyses, where each comparison was added to a base model that included only frequency estimates. Specifically, marginal R^2 increases in the present analyses ranged from approximately 0.3% to 4.5%, compared to increases as high as 9% in the previous analyses. This attenuation suggests that methodological similarities between the LBF and direct comparison items might contribute to the overall predictive power of direct comparison measures. Still, the fact that significant variance remains accounted for even after controlling for methodological similarities supports the interpretation that relevant comparisons also play a substantive conceptual role in shaping LBF judgments.

5.3 | Interim Discussion of Study 1

The results of Study 1 suggest that comparison-based reasoning meaningfully contributes to LBF judgments above and beyond participants' own absolute estimates of how frequently they perform a behavior. Although absolute frequency estimates were still the strongest predictor of LBF responses, our analyses showed that direct-comparison judgments were also significantly predictive, accounting for an additional 13% of the variance in self-LBF judgments. Notably, we found that many types of comparisons are predictive of self-LBF judgments, though not all to the same degree. Comparisons to the average person and expert benchmarks emerged as particularly influential. Peer comparisons also had a notable but slightly reduced influence, whereas comparisons to one's previous self were only a very weak predictor of self-LBF judgments.

Another a priori goal of ours was to separate the conceptual versus methodological contribution of direct comparisons for predicting LBF judgments. We did this by examining the extent to which responses to comparative questions about a distant social group—in this case, the average European—also predicted self-LBF judgments. To our surprise, we found that these responses explained additional variance beyond self-absolute frequency estimates, suggesting that methodological similarities could indeed play a role in predicting self-LBF judgments. To further probe this, we ran a series of analyses in which we controlled for the conceptually irrelevant but methodologically similar comparison (the European standard) to isolate the conceptual contribution of more meaningful standards. We found that global, expert, local, and—to a very small extent—prior-self comparisons each continued to

TABLE 3 | Analyses examining the added predictive power of each direct comparison measure above absolute frequency estimates in Study 1.

							95% CI for β		Model comparison results		
Model	**Predictors**	β	**SE** β	*df*	*t*	*p*	**Lower-bound**	**Upper-bound**	**Marginal** R^2	Δ**Marginal** R^2	*Sig.* X^2
Base	Self-absolute frequency	0.77	0.01	3315.40	67.23	< 0.001	0.75	0.80	0.554	—	—
Base + expert comparison	Self-absolute frequency	0.53	0.01	3318.59	37.76	< 0.001	0.51	0.56	0.608	0.054	< 0.001
	Expert comparison	0.35	0.01	3332.42	25.61	< 0.001	0.32	0.38			
Base + global comparison	Self-absolute frequency	0.49	0.01	3311.96	35.15	< 0.001	0.46	0.52	0.644	0.090	< 0.001
	Global comparison	0.39	0.01	3328.40	30.46	< 0.001	0.36	0.41			
Base + local comparison	Self-absolute frequency	0.57	0.01	3305.97	42.60	< 0.001	0.54	0.60	0.615	0.061	< 0.001
	Local comparison	0.31	0.01	3320.35	25.28	< 0.001	0.28	0.33			
Base + distant comparison	Self-absolute frequency	0.60	0.01	3324.21	45.79	< 0.001	0.58	0.63	0.619	0.065	< 0.001
	Distant comparison	0.28	0.01	3307.54	23.09	<0.001	0.26	0.31			
Base + previous-self comparison	Self-absolute frequency	0.75	0.01	3316.97	62.38	< 0.001	0.72	0.77	0.559	0.005	< 0.001
	Previous-self comparison	0.09	0.01	3327.17	7.93	<0.001	0.06	0.11			

Note: Standardized predictors. 95% confidence intervals based on profile-likelihood estimation.

TABLE 4 | Analyses examining the added predictive power of each conceptually relevant comparison measure above absolute frequency estimates and distant social group comparisons in Study 1.

							95% CI for β		Comparison to base model		
Model	**Predictors**	β	**SE** β	df	t	p	**Lower-bound**	**Upper-bound**	**Marginal** R^2	**ΔMarginal** R^2	***Sig.*** X^2
Base	Self-absolute frequency	0.60	0.01	3324.21	45.79	< 0.001	0.58	0.63	0.619	—	—
	Distant comparison	0.28	0.01	3307.54	23.09	< 0.001	0.26	0.31			
Base + global comparison	Self-absolute frequency	0.45	0.01	3310.25	31.96	< 0.001	0.42	0.48	0.664	0.045	< 0.001
	Distant comparison	0.16	0.01	3330.70	12.34	< 0.001	0.13	0.18			
	Global comparison	0.31	0.01	3293.75	22.53	< 0.001	0.28	0.34			
Base + local comparison	Self-absolute frequency	0.50	0.01	3310.23	36.22	< 0.001	0.47	0.52	0.648	0.029	< 0.001
	Distant comparison	0.20	0.01	3319.46	15.76	< 0.001	0.17	0.22			
	Local comparison	0.23	0.01	3298.75	18.62	< 0.001	0.21	0.26			
Base + expert comparison	Self-absolute frequency	0.47	0.01	3320.99	33.15	< 0.001	0.44	0.50	0.642	0.022	< 0.001
	Distant comparison	0.19	0.01	3331.08	15.29	< 0.001	0.17	0.22			
	Expert comparison	0.27	0.01	3314.10	18.64	< 0.001	0.24	0.30			
Base + previous-self comparison	Self-absolute frequency	0.58	0.01	3323.91	43.73	< 0.001	0.56	0.61	0.622	0.003	< 0.001
	Distant comparison	0.28	0.01	3312.53	22.64	< 0.001	0.25	0.30			
	Previous-self comparison	0.07	0.01	3328.61	6.68	< 0.001	0.05	0.09			

Note: Standardized predictors. 95% confidence intervals based on profile-likelihood estimation.

explain unique variance in LBF judgments even when controlling for shared methodology, supporting the conclusion that their influence is not reducible to methodological similarities alone. That said, the smaller marginal R^2 increases in these models suggest that methodological overlap may inflate associations to some degree.

6 | Study 2

Study 1 suggests that LBF judgments are influenced by more than just a person's belief about their behavior frequency—they also appear to reflect judgments about relative standing. Importantly, there seems to be a wide range of comparisons that inform LBF judgments. This raises additional questions about the underlying cognitive basis of these comparative effects. One possibility is that participants rely on pairwise comparisons of specific absolute representations of frequency to inform self-LBF judgments. That is, when asked for a self-LBF judgment, people gauge how they respond as a function of thoughts they have about how their own absolute frequency compares to the absolute frequencies of others or for retrieved standards (e.g., I floss almost about 12 times a week, and I think other people floss less than seven times, so I floss pretty often ...). The possibility that people retrieve and compare specific behavioral exemplars is consistent with prior work showing that social comparisons often involve concrete, memory-based contrasts between the self, and others (e.g., Mussweiler and Strack 2000). Alternatively, participants might rely predominantly on gist-based representations that they may have already developed before even being asked, or that they generate on the basis of comparisons without consulting absolute/verbatim frequency representations about comparison standards (e.g., I hardly hear about anyone who flosses as much as I do). Prior work suggests that people often rely on such gist-like, categorical representations when making social judgments, especially when precision is unnecessary or effortful (e.g., Brainerd and Reyna 2002, 2004; Reyna and Brainerd 1995).

Study 2 was designed to replicate the findings of Study 1 while also providing additional information about the cognitive bases of the observed comparative effects. The methodological difference from Study 1 was the addition of *other-absolute frequency* questions. These solicited estimates of absolute frequencies about how often other people perform a behavior (e.g., peers, Europeans) or, in the case of the expert benchmark question, how often experts recommend it be performed. The inclusion of this question type allowed us to compare the predictive validity of *direct-comparison* judgments versus *other-absolute frequency* estimates—above and beyond what is accounted for by self-absolute estimates. If the reason why comparative judgments are predictive is that people's LBF judgments are shaped by pairwise comparisons of absolute frequencies, then participants' responses to other-absolute frequency questions should be predictive of self-LBF judgments (similar in strength to how direct-comparative responses are predictive). In contrast, if LBF judgments are more gist-like in nature and pairwise comparisons of absolute estimates are not drivers of LBF judgments, then participants' responses to other-absolute frequency questions would not be as predictive of self-LBF judgments as are the direct-comparative judgments. Lastly, if there was a predictive advantage of the direct-comparison judgments in this study, we could also test whether a shared-method-variance explanation could account for this advantage.

6.1 | Methods

6.1.1 | Participants

Our preregistered sample size of 250 was again based on the work by Schönbrodt and Perugini (2013). The final sample was 234 undergraduate students (186 females, 48 males, $M_{age} = 19.36$, $SD = 1.62$) who completed the survey in a laboratory setting and received credit for a research requirement in their introductory psychology course.

6.1.2 | Procedure

The procedure of Study 2 was identical to that of Study 1, with the exception that the other-absolute frequency estimate items were added into the sequence of questions participants answered. Specifically, every participant again answered the Self-LBF questions first before being randomly assigned to one of two counterbalancing conditions. In one condition, participants first provided self-absolute frequency estimates of how frequently they performed each behavior, followed by absolute frequency estimates for the five comparison groups (*other-absolute frequency*). In the other condition, this order was reversed: participants first completed the other-absolute frequency estimates, then the self-absolute frequency estimates. Finally, participants answered the five blocks of direct comparison items (with each behavior in its own given block) in randomized order.

6.1.3 | Measures

The self-LBF-, self-absolute frequency-, and direct comparison measures in Study 2 were identical to those in Study 1. The only new measures—other-absolute frequency questions—were assessed in the following manner. Akin to the direct comparison measures, there were five types of estimates for each behavior: global, local, expert, previous self, and distant social group. Answers for each estimate type were given via a drop-down menu with a range of occurrences, which differed slightly from behavior to behavior (see survey materials on OSF for a full list of the questions for each behavior). Here, we present the wording primarily for the consumption-of-red-meat behavior. *Global estimate* questions asked, "How many times a month does an average person your age and gender eat red meat?" *Local estimate* questions asked, "Think about your close peers (people you see or hang out with often. How many times a month do they eat red meat?" *Expert estimate* questions said, "Think about how often experts recommend doing this behavior. How many times a month do experts say a person your age and gender should eat red meat?" *Prior-self estimate* questions asked, "Think about how often you ate red meat a year ago. How many times a month did you eat red meat?" *Distant social group estimate* questions said, "Think about the average European person. How many times a month does an average European eat red meat?" As with self-absolute frequency estimates, all types of other-absolute frequency estimates log transformed for analyses to account for positive skew.

6.2 | Results

6.2.1 | Preliminary Analyses

We start by reporting the same set of preliminary analyses as in Study 1. Table 5 summarizes the zero-order correlations among self-LBF judgments, self-absolute frequency estimates, direct comparisons, and other-absolute frequency estimates; averaged across the eight behaviors using Fisher z-transformation. As in Study 1, self-LBF judgments were correlated with self-absolute frequency estimates and with direct-comparison judgments. All types of other-absolute frequency estimates were positively and significantly correlated with self-absolute frequency judgments; likewise, they were also positively and significantly correlated with self-LBF judgments. Intercorrelations among the five other-absolute frequency variables ranged from $r = 0.294$ to 0.519, with the strongest associations observed between local and global measures. VIFs from a multilevel model including all five other-absolute frequency measures ranged from 1.22 to 2.01 (average VIF = 1.60), which did not constitute concern (Garson 2012; O'brien 2007; Tabachnick and Fidell 2007).

TABLE 5 | Average correlations between LBF estimates, self-absolute frequency estimates, direct comparisons, and other-absolute frequency estimates in Study 2.

		Average correlations											
Measure	**Range**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1. Self-LBF judgment	1 to 7	—											
2. Self-absolute frequency	0 to 75	0.754	—										
3. Expert comparison	−3 to 3	0.656	0.645	—									
4. Global comparison	−3 to 3	0.687	0.670	0.706	—								
5. Local comparison	−3 to 3	0.588	0.556	0.616	0.709	—							
6. Prior-self comparison	−3 to 3	0.303	0.279	0.292	0.318	0.342	—						
7. Distant comparison	−3 to 3	0.584	0.588	0.664	0.709	0.603	0.279	—					
8. Expert-absolute frequency	0 to 3.47	0.184	0.339	−0.011	0.135	0.103	0.091	0.079	—				
9. Global-absolute frequency	0 to 4.11	0.190	0.368	0.107	0.036	−0.019	0.053	0.091	0.371	—			
10. Local-absolute frequency	0 to 4.33	0.319	0.502	0.256	0.207	−0.006	0.075	0.216	0.361	0.519	—		
11. Previous-absolute frequency	0 to 3.78	0.556	0.690	0.502	0.511	0.408	−0.127	0.457	0.331	0.359	0.476	—	
12. Distant-absolute frequency	0 to 4.33	0.136	0.300	0.049	0.039	0.024	0.075	−0.133	0.425	0.455	0.378	0.294	—

Note: Correlations were averaged across the eight health behaviors using Fisher z-transformation and back-transformed for reporting. Self-absolute and other-absolute frequency estimates were log transformed before going into the analyses.

6.2.2 | Main Analyses

6.2.2.1 | Examining Direct Comparisons—Do the Results of Study 1 Replicate?. Before turning to our main question about the predictive validity of other-absolute frequency estimates versus direct-comparison judgments, we first sought to replicate the key finding from Study 1: That direct comparisons reliably predict self-LBF judgments beyond absolute frequency estimates. As in Study 1, we found that a model including the five direct comparison measures as predictors of self-LBF judgments provided a better fit than a model including only self-absolute frequency estimates, $\chi^2(5) = 360.67$, $p < 0.001$. The marginal R^2 increased from 55.08% for the base model to 63.40% for the model including direct comparisons, indicating that the direct comparisons explained an additional 8.32% of variance in LBF judgments. This generally matches the pattern observed in Study 1, again suggesting that self-LBF responses are shaped by both perceived absolute behavior frequency and comparative assessments of behavior.

Next, we examined the relative contribution of each direct comparison type. Figure 2 plots the model coefficients of the model that included self-absolute frequency estimates and all direct comparisons as predictors; see specifically the gray-triangle icons for the *direct comparison model*. Self-absolute frequency again remained a strong predictor of self-LBF judgments ($\beta = 0.44$, SE = 0.02, $t(1566.98) = 20.79$, $p < 0.001$, *95% CI* [0.40; 0.49]). Three of the four relevant direct comparisons also again emerged as a significant predictor, and their relative contributions were generally comparable to those of Study 1. Global comparisons ($\beta = 0.19$, SE = 0.03, $t(1539.09) = 7.35$, $p < 0.001$, *95% CI* [0.14; 0.24]) and comparisons to expert benchmarks ($\beta = 0.13$, SE = 0.02, $t(1538.98) = 6.06$, $p < 0.001$, *95% CI* [0.09; 0.17]) were the strongest direct comparison predictors, followed by comparisons to local peers ($\beta = 0.09$, SE = 0.02, $t(1549.72) = 4.44$, $p < 0.001$, *95% CI* [0.05; 0.13]). Comparisons to the previous self ($\beta = 0.03$, SE = 0.01, $t(1565.02) = 1.83$, $p = 0.067$, *95% CI* [−0.00; 0.05]) were only marginally significant, and distant social group comparisons ($\beta = 0.03$, SE = 0.02, $t(1566.80) = 1.37$, $p = 0.172$, *95% CI* [−0.01; 0.07]) were not a significant predictor of self-LBF judgments in Study 2.

To assess the unique contribution of each comparison type in terms of additional explained variance, we again fit a series of models in which each direct comparison predictor was entered alongside self-absolute frequency. Each of these models was then compared to the base model with only self-absolute frequency (see Section A of the Supporting Information for a full reporting of these models). As in Study 1, all five comparisons significantly improved model fit (all $p\text{s} < 0.001$), and each was a significant predictor within its respective model. Changes in marginal R^2 showed that global and local comparisons explained the greatest additional variance (7.73% and 5.14%, respectively), followed by distant social group comparisons (5.10%) and expert benchmark comparisons (3.80%). Comparisons to the previous self again accounted for the smallest increase in explained variance (0.84%). These results reaffirm the finding that, although all comparison types carry some predictive value, certain comparisons—particularly global and local standards—may play a more central role in shaping how people respond to self-LBF questions.

Next, we once again assessed the possibility that methodological similarities could be driving the relationship between direct comparisons and self-LBF judgments by examining the extent to which the four relevant direct comparison types explain variance in self-LBF judgments beyond self-absolute frequency estimates and distant social group (i.e., nonrelevant) comparisons. The conclusions of this analysis are largely the same as for Study 1 (see Section B of the Supporting Information or full reporting). Across all models, adding each conceptually relevant direct comparison led to a significant improvement in model fit (all $p\text{s} < 0.001$), and each comparison emerged as a significant

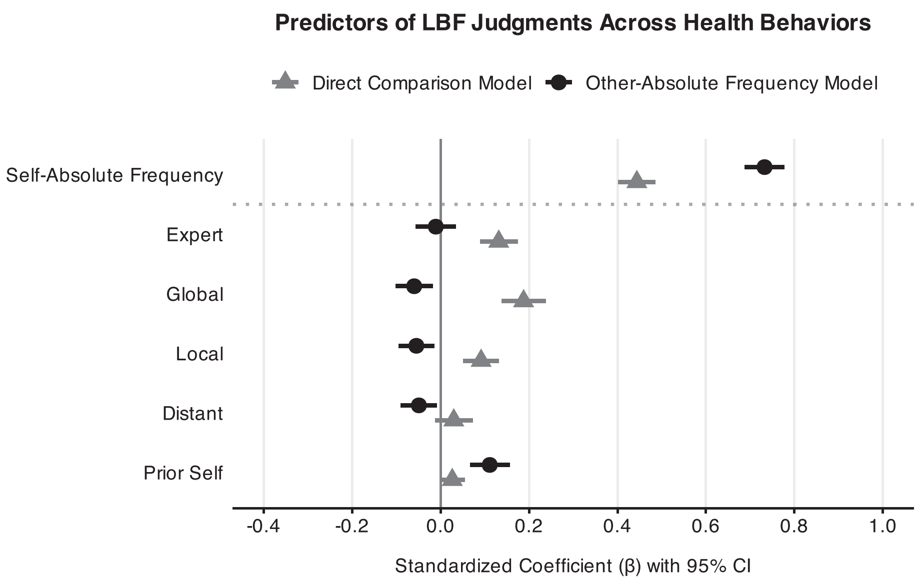

FIGURE 2 | Study 2. Model coefficients for the cross-classified multilevel model including self-absolute frequency estimates (above dotted line) and the five other-absolute frequency measures (other-absolute frequency model) and the cross-classified multilevel model including self-absolute frequency estimates and the five direct comparison measures (direct comparison model). Error bars represent 95% confidence intervals based on profile-likelihood estimation. Dotted horizontal line separates self-absolute frequency estimates from measures for the five reference groups.

predictor when included alongside absolute frequency estimates and the distant social comparison. The additional variance explained by each predictor varied, with global comparisons yielding the largest gain in marginal R^2 (3.47%), followed by local (2.07%) and expert (1.00%) comparisons. Comparisons to participants' previous selves again accounted for a smaller yet still significant portion of additional variance (0.44%). These findings replicate those from Study 1 and reinforce the conclusion that the predictive influence of relevant comparisons cannot be attributed solely to shared-method variance with the outcome measure.

6.2.2.2 | Do Pairwise Comparisons of Absolute Frequencies Play a Role?.

In Study 2, we added the other-absolute frequency measures in order to learn more about the nature of the comparison processes that are influencing LBF judgments. One possibility is that being asked for an LBF judgment triggers a pairwise comparison for formulating the LBF response. The pairwise comparison would be between a person's estimate of their own frequency of the behavior and a referent estimate—such as an expert's standard or the estimated frequency with which other people do the behavior. If such pairwise comparisons with other-absolute frequency estimates trigger and drive LBF responses, then the other-absolute frequency estimates should carry the same predictive value as direct-comparison judgments. In contrast, if people are making pairwise comparisons and are relying more on gist-like impressions of relative standing, these other-absolute estimates may be less strongly associated with self-LBF judgments. To examine whether other-absolute frequency estimates predict self-LBF judgments beyond self-absolute frequency estimates in a similar manner as do direct comparisons, we used the same analysis approach as for direct comparisons. Namely, we compared a model including self-absolute frequencies and the five types of other-absolute frequency estimates to the base model with self-absolute frequency estimates as the sole predictor of self-LBF judgments (again with all absolute frequencies log transformed).

A likelihood ratio test showed that the model that included other-absolute frequency estimates provided a significantly better fit, $\chi^2(5) = 56.37$, $p < 0.001$. However, the marginal R^2 increased only from 54.08% to 57.39% with the inclusion of the five other-absolute frequency estimates, indicating that absolute estimates about others' behavioral frequencies accounted only for an additional 3.31% of the variance in self-LBF judgments, above and beyond the self-absolute frequency estimates. This relatively modest gain in predictive power contrasts with the stronger influence of direct comparisons (13% increase in marginal R^2 for Study 1 and 8.32% increase for Study 2) and provides initial evidence against the idea that comparative judgments are predictive because people's self-LBF judgments are shaped by pairwise comparisons of absolute frequencies. If participants were relying on absolute-frequency based computations for their comparisons, one would expect other-absolute frequency estimates to contribute substantially more explanatory power.

When we examine the marginal predictive power of each type of other-absolute frequency estimates individually, the overall conclusion remains the same: They were not strong predictors. Figure 2 plots coefficients from the model that includes self-absolute frequency estimates and the five other-absolute frequency estimates as predictors; see specifically the black-circle icons for the *other-absolute frequency model*. For interpreting those coefficients for the other-absolute frequency measures, we first note that the pairwise-comparison account would expect them to be negative rather than positive.[6] The coefficients for the global, local, and distant social estimates were significantly negative, but their magnitudes were modest: $\beta = -0.06$, SE = 0.02, $t(1575.94) = -2.78$, $p = 0.006$, 95% *CI* [−0.10, −0.02]; $\beta = -0.05$, SE = 0.02, $t(1570.36) = -2.66$, $p = 0.008$, 95% *CI* [−0.10, −0.01]; and $\beta = -0.05$, SE = 0.02, $t(1573.00) = -2.36$, $p = 0.018$, 95% *CI* [−0.09, −0.01]; respectively. The coefficient for the expert estimate was not significant, $\beta = -0.01$, SE = 0.02, $t(1505.83) = -0.47$, $p = 0.636$, 95% *CI* [−0.06, 0.03], and the coefficient for previous-self estimate was significant but in the opposite (positive) direction, $\beta = 0.11$, SE = 0.02, $t(1569.14) = 4.86$, $p < 0.001$, 95% *CI* [0.07, 0.16].

In another way of examining the marginal predictive power of each type of other-absolute frequency estimate, we ran a series of individual model comparison analyses, just as we had done for the direct comparisons. Specifically, we fit five models—each with the self-absolute frequency measure and one other-absolute frequency measure as predictors—and compared them to the base model that had only the self-absolute frequency measure as a predictor. Table 6 summarizes these models.

Although four of the five models showed a significant improvement over the base model, the magnitudes of the improvements were all quite small—ranging from 0.19% to 1.15% for the additional amount of variance explained. As reported earlier, the additional amounts of variance explained by adding direct comparison measures over the base model were generally larger; the highest was for global direct comparisons, which explained 7.73% additional variance. Even in the analyses that controlled for shared-method variance, the global direct comparisons explained 3.47% additional variance. In summary, the additional predictive utilities of other-absolute frequencies were small (with the highest at 1.15%) relative to the additional predictive utilities of some of the direct comparisons, even after accounting for shared-method variance.

6.2.2.3 | How Are Direct-Comparison Judgments Themselves Constructed?.

One question that may arise is how the direct-comparison judgments themselves are constructed and whether they truly reflect intuitive gist judgments that go beyond numeric comparisons, or whether they partially arise from computations based on absolute frequency estimates. To probe this, we conducted an exploratory analysis in which we computed indirect comparison indices as the difference between participants' self- and other-absolute frequency estimates and assessed their correspondence with the direct comparison measures. Across judgment targets, these indirect indices were moderately correlated with the corresponding direct comparisons (*rs* ranging from 0.470 to 0.574) and explained roughly 30% of the variance in those ratings (R^2s ranging from 0.24 to 0.38). This pattern suggests that direct comparisons may incorporate some information that is also captured by difference-based computations, but that a substantial portion of the variance in direct-comparison judgments is not well accounted for by these numeric contrasts alone—consistent with the idea that direct comparisons reflect more gist-like assessments.

TABLE 6 | Analyses examining the added predictive power of each other-absolute frequency measure above absolute frequency estimates in Study 2.

							95% CI for β		Model comparison results		
Model	**Predictors**	β	**SE** β	***df***	***t***	***p***	**Lower-bound**	**Upper-bound**	**Marginal** R^2	**ΔMarginal** R^2	***Sig.*** X^2
Base	Self-absolute frequency	0.75	0.02	1570.43	45.59	< 0.001	0.72	0.78	0.541	—	—
Base + expert frequency	Self-absolute frequency	0.76	0.02	1564.30	44.42	< 0.001	0.73	0.79	0.537	−0.003	0.057
	Expert frequency	−0.04	0.02	1540.05	−1.90	0.058	−0.09	0.00			
Base + global frequency	Self-absolute frequency	0.78	0.02	1565.03	45.02	< 0.001	0.74	0.81	0.550	0.009	< 0.001
	Global frequency	−0.09	0.02	1549.61	−4.73	< 0.001	−0.13	−0.05			
Base + local frequency	Self-absolute frequency	0.79	0.02	1563.64	41.96	< 0.001	0.75	0.82	0.543	0.002	< 0.001
	Local frequency	−0.07	0.02	1568.82	−3.91	< 0.001	−0.11	−0.04			
Base + distant frequency	Self-absolute frequency	0.77	0.02	1568.04	45.39	< 0.001	0.73	0.80	0.552	0.012	< 0.001
	Distant frequency	−0.08	0.02	1518.47	−4.23	< 0.001	−0.12	−0.04			
Base + previous-self frequency	Self-absolute frequency	0.70	0.02	1564.30	31.22	< 0.001	0.65	0.74	0.548	0.007	< 0.001
	Previous-self frequency	0.08	0.02	1566.13	3.59	< 0.001	0.04	0.13			

Note: Standardized predictors. 95% confidence intervals based on profile-likelihood estimation.

Next, we explored the predictive value of the indirect comparison indices for self-LBF judgments. Relative to the base model including only self-absolute frequency estimates (marginal $R^2 = 0.54$), adding all indirect comparison indices explained only an additional 2.51% of variance, similar in magnitude to the increment observed when adding the other-absolute estimates themselves (3.31%). By comparison, adding the full set of direct comparison measures explained 8.32% of additional variance. Thus, although difference-based indices derived from absolute frequencies capture part of the relational information that influences LBF responses, they were less predictive than participants' direct-comparative judgments. Taken together, these exploratory findings suggest that direct comparisons are unlikely to arise solely from arithmetic transformations of verbatim frequency estimates. Instead, they may reflect a partially independent layer of comparative reasoning—one that incorporates but is not reducible to numeric differences—consistent with a more gist-like process of evaluating one's standing relative to various referents.

6.3 | Interim Discussion of Study 2

Study 2's results generally replicate the findings of Study 1. As in Study 1, we found that direct comparison measures of various types predicted LBF judgments above and beyond self-absolute frequency estimates of behavior; reaffirming that LBF judgments reflect not only a person's beliefs about behavior frequency but also their beliefs about relative standing. As in Study 1, comparisons to the average person and expert benchmarks had the strongest influence, followed by local comparisons and—to a lesser extent—comparisons to the previous self.

Importantly, Study 2 also provides additional insights into the cognitive basis of these comparative effects. Specifically, we compared the predictive power of direct-comparison judgments with that of absolute frequency estimates for the relevant referent groups (e.g., others and experts). Although both types of measures explained variance in LBF judgments beyond self-absolute estimates, direct-comparison judgments consistently accounted for more variance (8.32%) than the corresponding other-absolute frequency estimates (3.31%). Standardized coefficients likewise tended to indicate stronger effects for the direct comparisons. Moreover, even after controlling for potential shared-method variance—given that both direct comparison and LBF judgments rely on the same response format—direct comparisons retained a slight predictive advantage. These findings are more consistent with a gist-based interpretation: Rather than retrieving and comparing explicit absolute frequency standards, participants may rely on broad, categorical impressions of relative standing (e.g., "more than most" or "about the same as others"). Such gist-based representations may integrate various sources of comparison information in a holistic manner, bypassing the need for detailed numerical comparisons.

7 | General Discussion

Prior work has found that responses to Likert-type behavioral frequency questions can involve comparison processes (Robert Pace and Friedlander 1982; Schaeffer 1991; Sudman et al. 1996; Wänke 2002; Wright et al. 1994). However, little is known about which types of comparisons are most influential. In the present work, we addressed this gap by examining the extent to which different reference standards—including population norms, expert benchmarks, close peers, distant others, and the self over time—systematically influence LBF judgments about health-related behaviors.

In both Studies 1 and 2, we found that participants' self-LBF judgments were significantly predicted by a range of comparative standards, even after controlling for absolute estimates of their own behavioral frequency. This finding supports prior research suggesting that LBF judgments are not just pure reflections of recalled behavioral frequencies but also a reflection of how one situates those frequencies in a broader comparative context (Cole and Korkmaz 2013; Robert Pace and Friedlander 1982; Schaeffer 1991; Wänke 2002). Notably, although all comparison types were statistically significant predictors of self-LBF judgments, comparisons to global population norms and expert recommendations emerged as the strongest contributors. Comparisons to local peers were also a notable contributor, with comparisons to one's previous self only playing a small role in shaping self-LBF judgments.

That expert benchmarks emerged as especially strong predictors of self-LBF judgments is particularly notable. Prior research has shown that expert guidelines can shape behavior itself (Cummings et al. 1979; Larson and Story 2009; Michie et al. 2012; Prochaska et al. 2006). The present findings extend this work by indicating that such benchmarks—at least in the context of health behaviors—may also guide how individuals interpret and use LBF questions. At the same time, the fact that comparisons to both global and local peers also predicted LBF responses is consistent with the notion that people tend to evaluate themselves relative to others in their social environment (Festinger 1954). Although prior work has emphasized the importance of local comparisons (Alicke et al. 1995; Zell and Alicke 2010), the current results suggest that individuals may also—perhaps even to a larger extent—draw on more distal or global comparison standards to the average person when forming LBF judgments about health behaviors.

Although the findings above are important for interpreting LBF judgments in health-related domains, it is reasonable to speculate that the relative contributions of different comparison groups may shift across contexts and individuals. As evidenced by other research, interpretations of vague verbal quantifiers can differ depending on the target behavior (Pepper and Prytulak 1974), as well as across demographic groups (Borgers et al. 2003; Krumpal et al. 2008; Schaeffer 1991; Schneider and Stone 2016; Wright et al. 1994). Because LBF questions rely on such vague quantifiers to anchor scale endpoints, it is likely that the patterns we observed are—to some extent—specific to the present health-behavior context. For example, we found that—in the present context—comparisons to one's prior self were a relatively weak predictor of LBF judgments. However, one possibility is that participants perceived little change in their own health-related behaviors over time, leading to a clustering of responses around the scale midpoint and comparatively low response variability. It is possible that comparisons to one's prior self may be more predictive of LBF judgments in contexts where behaviors are less stable. Regardless, the fact that such a wide range of possible comparison targets has a unique influence on self-LBF judgments is indicative of the general importance of comparisons in shaping LBF judgments. Going forward, a key question is not whether comparison processes are at play, but which comparison sources are most salient or influential in a given context. Understanding this could help clarify when LBF responses will and will not map onto objective behavioral frequencies.

The present findings also have implications for theories on why LBF questions are subject to comparison processes. As outlined in the introduction of this paper, range-frequency theory (Parducci 1965) posits that judgments are shaped not only by the absolute value of the target (in this case, one's own behavior frequency) but also by its position within a perceived range and distribution of relevant values. However, the theory is agnostic regarding where these comparison values come from. Our results can now speak to this. If LBF judgments are indeed shaped by a target value's position within a perceived distribution of relevant values, our findings indicate that the values making up this distribution possibly stem from a broad range of comparison sources, including expert recommendations and knowledge about the behavior frequencies of the average person and close peers. This finding dovetails with more recent process-level accounts, such as DbS theory (Stewart et al. 2006), which explicitly propose that reference distributions are constructed in the moment through cognitive sampling from diverse sources in memory and the environment.

It remains an open question whether respondents typically integrate multiple comparative sources into a single mental distribution of relevant values, or whether different individuals rely on different types of comparison sources when forming their distributions. Both possibilities are consistent with the current data, which show that a variety of comparison sources are predictive of LBF responses across participants. From a DbS perspective, it is also plausible that respondents draw a limited, contextually cued set of comparison information—sometimes drawn from multiple sources, sometimes dominated by a single source—depending on which information is most accessible at the moment of judgment. What the present results make clear is that the distributions underlying LBF judgments—however they are constructed—can reflect a wide array of comparison values, offering one explanation for why such judgments often diverge from absolute frequency estimates.

7.1 | Gist vs. Verbatim Representations of Relative Behavior Frequency

A key feature of Study 2 was the inclusion of other-absolute frequency questions for the different reference groups, which allowed us to probe whether comparative influences on LBF judgments are best understood as arising from verbatim pairwise comparisons of absolute frequency estimates or from more gist-like representations of relative behavioral standing. We did this by comparing the added predictive value of direct-comparative judgments to that of absolute frequency estimates for the same reference groups. Direct-comparison judgments explained more additional variance in LBF responses than did their corresponding absolute-frequency counterparts, even when controlling for the shared-method benefits of direct comparison measures. This suggests that, when it comes to the influence of comparisons on LBF judgments, participants were not performing pairwise comparisons between two frequency estimates but were instead drawing on more abstract, gist-like impressions of relative frequency.

The finding that the influence of comparisons on LBF judgments is gist-like in nature is consistent with a DbS framework (Stewart et al. 2006), which proposes that judgments are constructed by sampling comparison values and determining the target's rank among them. DbS assumes ordinal comparisons drive evaluation, making judgments sensitive to relative position rather than precise magnitude. We do not need to assume that these ordinal comparisons occur only at the moment of judgment. People have repeated opportunities over time to notice how their own behavior compares to that of others as well as to norms suggested by experts. From these many pairwise comparisons, a gist representation of relative standing can gradually form and be updated. When asked for an LBF judgment, a person's response may reflect this accumulated gist rather than an on the spot comparison of absolute frequency estimates. This rank-based process provides a natural explanation for why LBF judgments appear to rely on abstract impressions of behavioral standing rather than precise numeric differences.

Broadly speaking, this conclusion also aligns with the conceptual framework proposed by fuzzy-trace theory (e.g., Brainerd and Reyna 2002, 2004; Reyna and Brainerd 1995), which proposes that individuals encode both verbatim (precise and quantitative) and gist (categorical and qualitative) representations of information but tend to rely more heavily on gist when making judgments—particularly in contexts that are ambiguous, low stakes, or effortful to quantify. In the present case, the tendency for direct comparisons to outperform other-absolute-frequency estimates suggests that people may not engage in verbatim numerical comparisons when responding to LBF questions. Instead, they may rely on intuitive gist-based assessments of their relative standing, informed by accessible and meaningful categories (e.g., "more than most," "about average," "less than recommended"). This process may reflect a broader cognitive strategy, whereby individuals integrate diverse inputs—memories, social context, normative beliefs/expert recommendations—into a coherent, gist-based summary judgment about comparative standing, bypassing the need for detailed numerical comparisons.

7.2 | Ruling Out a Shared-Method Explanation for the Predictiveness of Direct Comparisons

Another goal of ours was to disentangle the conceptual impact of direct comparisons from any methodological advantages they may have in predicting LBF judgments, due to their shared response format with self-LBF judgments. To explore this, we examined whether comparisons to a conceptually irrelevant standard—the average European—also predicted self-reported frequency judgments. If so, that would indicate that methodological similarities between LBF questions and direct comparison measures are contributing to the association between LBF judgments and direct comparisons. Somewhat surprisingly, in Study 1 (but not Study 2), comparisons to the average European did account for unique variance beyond what was explained by self-absolute estimates, suggesting that surface-level methodological features (e.g., response format or question framing) may contribute to the predictive strength of comparison measures. However, additional analyses controlling for the shared-method advantage showed that conceptually meaningful comparisons—to the average person, local peers, expert benchmarks, and the prior self—continued to account for unique variance in self-LBF judgments beyond both absolute frequency estimates and shared-method benefits. This indicates that the predictiveness of direct comparisons cannot be fully explained by shared methodological features. Moreover, although the comparison to the average European also explained unique variance, the relative effect size differed substantially across comparison targets. In particular, global and expert comparisons produced larger and more consistent effects than did comparisons to the prior self or the distant other. If the incremental predictive power of the comparison measures were driven primarily by increased reliability from asking similar questions repeatedly, a more uniform pattern of contributions across comparison targets would be expected. Instead, the observed heterogeneity in effect sizes suggests that these measures capture meaningful, content-specific information over and above shared methodological features. As such, these results confirm that comparisons play a meaningful psychological role in shaping how people respond to LBF questions.

7.3 | Implications for Survey Research and Psychological Measurement

From an applied perspective, our findings carry important implications for the use of LBF items in psychological and behavioral research. LBF questions are widely used in academic and applied contexts both as *practical indicators* of how often people engage in behaviors and as measures of self-perceived behavioral frequency (e.g., Likert 1932; Schwarz and Oyserman 2001; Tourangeau et al. 2000; see Jebb et al. 2021, for review). Although their ease of administration and cognitive simplicity are clear advantages (Al Baghal 2014; Bradburn and Miles 1979; Burton and Blair 1991; Tourangeau et al. 2000), our findings confirm that these measures are not simple translations of internal frequency estimates. In fact, we found that self-absolute frequency estimates explained only around half of the variance in LBF judgments, which aligns with other work showing that LBF judgments do not consistently or directly map onto objective numerical estimates (e.g., Budescu and Wallsten 1985; Conrad et al. 1998; Schwarz 1999; Wright et al. 1994; Woltz et al. 2012).

One reason for these inconsistencies is that people do not interpret vague frequency scales in isolation—they interpret them through the lens of social and normative comparisons. Prior work has shown that making specific reference groups salient—through framing, wording changes, or contextual cues—can shift how vague quantifiers are interpreted (Wänke 2002), and that such interpretations can also vary systematically across demographic groups (Robert Pace and Friedlander 1982; Schaeffer 1991). However, a novelty of the present work is that it demonstrates that people draw on a wide array of comparison standards—including population norms, expert recommendations, peers, and past selves—even when responding to standard LBF questions that provide no explicit reference group or contextual framing. This suggests that comparison-based reasoning is a default feature of how people interpret vague frequency questions, not just a product of experimental manipulations. Moreover, it highlights the need for caution when interpreting LBF data as direct reports of internal behavioral frequency estimates.

With that said, the interpretive flexibility of LBF questions is not inherently problematic, and the tendency for LBF questions to prompt comparison-based reasoning may have important implications for their predictive utility. In the domain of health psychology, studies comparing the effectiveness of *absolute* versus *comparative* risk perception measures have found that comparative judgments—those in which individuals evaluate their own risk relative to others—can be at least as predictive of behavioral intentions as absolute risk estimates, and in some cases more strongly associated with behavioral intentions (e.g., Klein 2002; Miller et al. 2020; Portnoy et al. 2014; Renner and Reuter 2012; Rose 2010; but see Hay et al. 2016; Janssen et al. 2011). Given that LBF questions inherently promote such comparisons, their relative framing may not merely introduce response variability but may actually enhance their predictive power above and beyond absolute numeric frequency questions in certain contexts.

Finally, all measurement formats carry characteristic biases: Likert-type scales can show central-tendency and category-boundary effects (Douven 2018), numeric frequency reports can exhibit rounding and magnitude-related distortions (Honda et al. 2022), and visual analog scales can be shaped by their perceptual and interface design (Matejka et al. 2016). Thus, rather than any single format providing bias-free measurement, each introduces its own constraints, making it essential to select the response format that best matches one's analytic aims. As such, researchers should consider the goals of measurement when choosing between LBF and other formats. When the goal is to obtain precise or objective measures of behavior frequency—such as for modeling behavioral patterns, making behavioral assessments, or establishing population baselines—asking participants to recall recent instances of a behavior from memory rather than eliciting behavioral frequency along vague quantifier scales can yield greater accuracy and comparability (e.g., Petersen et al. 2025). Conversely, when the goal is to predict future behavior or capture perceptions of behavior frequency—especially in contexts where social or other comparisons salient—LBF judgments may be more informative.

7.4 | Limitations and Future Directions

While the present studies offer a comprehensive examination of the cognitive processes underlying LBF judgments, several limitations warrant consideration. First, the behavioral domains we examined were all health-related, which may constrain the generalizability of our findings to other types of behaviors (e.g., moral, occupational, or leisure behaviors). Second, although our use of the "distant other" (i.e., typical Europeans) as an irrelevant comparison standard provided a useful control for shared-method variance between predictors and the LBF outcome, it may not have been entirely construct-irrelevant in the way we intended. Specifically, participants who lacked a clear sense of how often Europeans engage in the target behaviors may have substituted more familiar standards—such as local or global norms—when answering the distant other comparison questions (Kahneman and Frederick 2002). If so, this would have inflated the predictive power of the distant-other judgments and made them a more conservative control than intended. Importantly, however, even with a potentially overconservative control method, our key relevant comparison predictors still explained substantial additional variance in LBF judgments, reinforcing the conclusion that direct comparisons play a psychologically meaningful role in how people interpret and respond to LBF questions.

Although our analyses focused on group-level patterns, it is also likely that individual respondents differ in the reference groups they rely on when interpreting vague quantifier labels in LBF questions. That is, rather than applying a uniform standard, respondents may draw upon different combinations of comparison targets, depending on personal relevance, accessibility, or interpretation of the LBF item. Future research could explore these individual-level differences more directly, for example, by modeling person-specific reliance on reference groups or examining how trait-level variables (e.g., social comparison orientation) moderate the influence of different standards. Such work could help clarify the extent to which the observed effects are driven by consistent use of particular standards versus heterogeneous strategies across individuals.

While the present work is novel in that it highlights how LBF judgments are shaped by a variety of comparisons that appear gist-based in nature, more work is still needed to examine the exact structure, stability, and causal influence of gist representations when it comes to impressions of behavioral frequency. For example, an important open question is whether gist representations perhaps also exert an influence on how people generate responses to self-absolute frequency measures. Moreover, prior work suggests that gist representations tend to be more stable and enduring than verbatim representations (Brainerd and Reyna 1990; Reyna et al. 2016), which raises the possibility that LBF judgments and absolute frequency estimates about the same behavior may follow different trajectories over time. Disentangling these dynamics could clarify when people are likely to revise their beliefs about their own behavior versus maintain stable, qualitative impressions (perhaps despite contradictory evidence).

Finally, an important question that cannot be fully addressed by the present studies concerns the psychological mechanisms through which comparison information shapes LBF judgments. The present results establish that multiple comparison standards reliably influence LBF responses in a gist-consistent manner, but it remains open whether comparisons shape perceived behavioral frequency itself or, alternatively, shift how respondents interpret and apply the verbal response scale (e.g., what qualifies as "often" or "a lot"). Both processes predict similar empirical patterns but carry different implications for the cognitive meaning of LBF judgments.

7.5 | Conclusion

By demonstrating that Likert-type behavioral frequency judgments are shaped by a range of comparison standards—and that these influences likely operate through gist-based rather than verbatim comparison processes—the present studies highlight the cognitive complexity underlying what may appear to be simple survey items. They also underscore the need to consider their comparative and interpretive nature—especially when behavioral frequency carries social, normative, or evaluative meaning. As researchers continue to use Likert-type scales, recognizing their cognitive underpinnings will be essential for valid and interpretable measurement.

Author Contributions

J.E.M and P.D.W: conceptualization and methodology. **J.E.M and P.D.W:** data collection and curation. **J.D.S, J.E.M, I.T.P, and P.D.W:** formal analysis. **J.D.S:** writing – original draft. **J.D.S, J.E.M, I.T.P, and P.D.W:** writing – reviewing and editing. **P.D.W:** supervision.

Funding

This work was supported by the National Science Foundation (SES-1851738) and Eunice Kennedy Shriver National Institute of Child Health and Human Development (HD098235).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data, materials, and code are available at: https://osf.io/c8abu.

Endnotes

[1] We note that this counterbalancing condition contained an error in the survey design, which caused our three demographics questions (age, gender, and race)—which were meant to appear at the end of the survey—to appear as a block mixed in with the five blocks of direct comparison items. For example, some participants would have answered the expert direct comparison items for each behavior, then the three demographic questions, and then the four other blocks of direct comparison items. We did not observe significant differences between counterbalancing conditions, so it is unlikely that this error had a notable impact on participants' responses.

[2] Although we log transformed all count variables in Studies 1 and 2 to account for positive skew, the conclusions of our main analyses remain largely the same when count variables are not transformed.

[3] General model specification: Self-LBF Judgments ~ Fixed Effects + (1 | participant) + (1 | behavior).

[4] We note that responses to the previous-self comparison measure ($SD = 1.32$) were less variable relative to other comparison items ($SDs \approx 1.58$–1.67), with many participants selecting the midpoint of the scale. This restricted range may have somewhat attenuated its predictive associations with LBF judgments.

[5] We note that the predictive power of the direct comparison with the average-European question seems to go beyond shared method variance, which potentially makes this a more conservative test of the extent to which each conceptually relevant comparison has predictive value beyond methodological similarity. Nonetheless, we report on the planned analyses.

[6] Consider, for example, a participant who estimates that the average person performs a behavior at a relatively *high* rate (higher than their own rate). If that person makes a pairwise comparison to formulate their self-LBF judgment, they should give a relatively *low* self-LBF judgment.

References

Al Baghal, T. 2014. "Is Vague Valid? The Comparative Predictive Validity of Vague Quantifiers and Numeric Response Options." *Survey Research Methods* 8, no. 3: 169–179. https://doi.org/10.18148/srm/2014.v8i3.5813.

Alicke, M. D., M. L. Klotz, D. L. Breitenbecher, T. J. Yurak, and D. S. Vredenburg. 1995. "Personal Contact, Individuation, and the Better-Than-Average Effect." *Journal of Personality and Social Psychology* 68, no. 5: 804–825. https://doi.org/10.1037/0022-3514.68.5.804.

Alicke, M. D., E. Zell, and D. L. Bloom. 2010. "Mere Categorization and the Frog-Pond Effect." *Psychological Science* 21, no. 2: 174–177. https://doi.org/10.1177/0956797609357718.

Bates, D., M. Maechler, B. Bolker, et al. 2015. "Package 'lme4'." *Convergence* 12, no. 1: 2. http://dk.archive.ubuntu.com/pub/pub/cran/web/packages/lme4/lme4.pdf.

Borgers, N., J. Hox, and D. Sikkel. 2003. "Response Quality in Survey Research With Children and Adolescents: The Effect of Labeled Response Options and Vague Quantifiers." *International Journal of Public Opinion Research* 15, no. 1: 83–94. https://doi.org/10.1093/ijpor/15.1.83.

Bradburn, N. M., and C. Miles. 1979. "Vague Quantifiers." *Public Opinion Quarterly* 43, no. 1: 92–101. https://doi.org/10.1086/268494.

Brainerd, C. J., and V. F. Reyna. 1990. "Gist Is the Grist: Fuzzy-Trace Theory and the New Intuitionism." *Developmental Review* 10, no. 1: 3–47. https://doi.org/10.1016/0273-2297(90)90003-M.

Brainerd, C. J., and V. F. Reyna. 2002. "Fuzzy-Trace Theory and False Memory." *Current Directions in Psychological Science* 11, no. 5: 164–169. https://doi.org/10.1111/1467-8721.00192.

Brainerd, C. J., and V. F. Reyna. 2004. "Fuzzy-Trace Theory and Memory Development." *Developmental Review* 24, no. 4: 396–439. https://doi.org/10.1016/j.dr.2004.08.005.

Bruchmann, K., and A. T. Evans. 2013. "Abstract Mind-Sets and Social Comparison: When Global Comparisons Matter." *Social Psychological and Personality Science* 4, no. 4: 427–433. https://doi.org/10.1177/1948550612464661.

Brun, W., and K. H. Teigen. 1988. "Verbal Probabilities: Ambiguous, Context-Dependent, or Both?" *Organizational Behavior and Human Decision Processes* 41, no. 3: 390–404. https://doi.org/10.1016/0749-5978(88)90036-2.

Buckingham, J. T., and M. D. Alicke. 2002. "The Influence of Individual Versus Aggregate Social Comparison and the Presence of Others on Self-Evaluations." *Journal of Personality and Social Psychology* 83, no. 5: 1117–1130. https://doi.org/10.1037//0022-3514.83.5.1117.

Budescu, D. V., and T. S. Wallsten. 1985. "Consistency in Interpretation of Probabilistic Phrases." *Organizational Behavior and Human Decision Processes* 36, no. 3: 391–405. https://doi.org/10.1016/0749-5978(85)90007-X.

Burton, S., and E. Blair. 1991. "Task Conditions, Response Formulation Processes, and Response Accuracy for Behavioral Frequency Questions in Surveys." *Public Opinion Quarterly* 55, no. 1: 50–79. https://doi.org/10.1086/269241.

Cole, J. S., and A. Korkmaz. 2013. "Estimating College Student Behavior Frequencies: Do Vague and Enumerated Estimation Strategies Yield Similar Results?" *Journal of Applied Research in Higher Education* 5, no. 1: 58–71. https://doi.org/10.1108/17581181311310270.

Conrad, F. G., N. R. Brown, and E. R. Cashman. 1998. "Strategies for Estimating Behavioural Frequency in Survey Interviews." *Memory* 6, no. 4: 339–366. https://doi.org/10.1080/741942603.

Cummings, K. M., A. M. Jette, B. M. Brock, and D. P. Haefner. 1979. "Psychosocial Determinants of Immunization Behavior in a Swine Influenza Campaign." *Medical Care* 17, no. 6: 639–649. https://doi.org/10.1097/00005650-197906000-00008.

Douven, I. 2018. "A Bayesian Perspective on Likert Scales and Central Tendency." *Psychonomic Bulletin & Review* 25, no. 3: 1203–1211. https://doi.org/10.3758/s13423-017-1344-2.

Festinger, L. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7, no. 2: 117–140. https://doi.org/10.1177/001872675400700202.

Garson, G. D. 2012. *Testing Statistical Assumptions*. Statistical Associates Publishing.

Hay, J. L., M. Ramos, Y. Li, S. Holland, D. Brennessel, and M. M. Kemeny. 2016. "Deliberative and Intuitive Risk Perceptions as Predictors of Colorectal Cancer Screening Over Time." *Journal of Behavioral Medicine* 39, no. 1: 65–74. https://doi.org/10.1007/s10865-015-9667-9.

Honda, H., R. Kagawa, and M. Shirasuna. 2022. "On the Round Number Bias and Wisdom of Crowds in Different Response Formats for Numerical Estimation." *Scientific Reports* 12, no. 1: 8167. https://doi.org/10.1038/s41598-022-11900-7.

Inchley, J., D. Currie, S. Budisavljevic, et al. 2020. Spotlight on Adolescent Health and Well-Being. Findings From the 2017/2018 Health Behaviour in School-Aged Children (HBSC) Survey in Europe and Canada. International Report. Volume 1. Key Findings.

Janssen, E., L. van Osch, H. de Vries, and L. Lechner. 2011. "Measuring Risk Perceptions of Skin Cancer: Reliability and Validity of Different Operationalizations." *British Journal of Health Psychology* 16, no. 1: 92–112. https://doi.org/10.1348/135910710X514120.

Jebb, A. T., V. Ng, and L. Tay. 2021. "A Review of Key Likert Scale Development Advances: 1995–2019." *Frontiers in Psychology* 12: 637547.

Kahneman, D., and S. Frederick. 2002. "Representativeness Revisited: Attribute Substitution in Intuitive Judgment." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, 49–81. Cambridge University Press. https://doi.org/10.1017/CBO9780511808098.004.

Klein, W. M. 1997. "Objective Standards Are Not Enough: Affective, Self-Evaluative, and Behavioral Responses to Social Comparison Information." *Journal of Personality and Social Psychology* 72, no. 4: 763–774. https://doi.org/10.1037//0022-3514.72.4.763.

Klein, W. M. 2002. "Comparative Risk Estimates Relative to the Average Peer Predict Behavioral Intentions and Concern About Absolute Risk." *Risk, Decision, and Policy* 7: 193–202. https://doi.org/10.1017/S1357530902000613.

Klein, W. M. 2003. "Self-Prescriptive, Perceived, and Actual Attention to Comparative Risk Information." *Psychology and Health* 18, no. 5: 625–643. https://doi.org/10.1080/0887044031000118465.

Krumpal, I., H. Rauhut, D. Böhr, and E. Naumann. 2008. "How Likely Is 'Likely'? Subjective Perception and Communication of Victimization Probabilities." *Methods, Data, Analyses* 2, no. 1: 25. https://doi.org/10.12758/mda.2008.001.

Larson, N., and M. Story. 2009. "A Review of Environmental Influences on Food Choices." *Annals of Behavioral Medicine* 38, no. suppl_1: s56–s73. https://doi.org/10.1007/s12160-009-9120-9.

Likert, R. 1932. "A Technique for the Measurement of Attitudes." *Archives of Psychology* 22, no. 140: 55.

Matejka, J., M. Glueck, T. Grossman, and G. Fitzmaurice. 2016. "The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales." In *Proceedings of the 2016 CHI Conference* *on Human Factors in Computing Systems*, 5421–5432. Association for Computing Machinery. https://doi.org/10.1145/2858036.2858063.

Michie, S., C. Whittington, Z. Hamoudi, F. Zarnani, G. Tober, and R. West. 2012. "Identification of Behaviour Change Techniques to Reduce Excessive Alcohol Consumption." *Addiction* 107, no. 8: 1431–1440. https://doi.org/10.1111/j.1360-0443.2012.03845.x.

Miller, J. E., P. D. Windschitl, T. A. Treat, and A. M. Scherer. 2020. "Comparisons as Predictors of People's Beliefs About the Importance of Changing Their Health Behaviors." *European Journal of Health Psychology* 27: 14–29. https://doi.org/10.1027/2512-8442/a000043.

Mussweiler, T., and K. Rüter. 2003. "What Friends Are for! The Use of Routine Standards in Social Comparison." *Journal of Personality and Social Psychology* 85, no. 3: 467–481. https://doi.org/10.1037/0022-3514.85.3.467.

Mussweiler, T., and F. Strack. 2000. "The "Relative Self": Informational and Judgmental Consequences of Comparative Self-Evaluation." *Journal of Personality and Social Psychology* 79, no. 1: 23–38. https://doi.org/10.1037/0022-3514.79.1.23.

Nakagawa, S., P. C. Johnson, and H. Schielzeth. 2017. "The Coefficient of Determination R 2 and Intra-Class Correlation Coefficient From Generalized Linear Mixed-Effects Models Revisited and Expanded." *Journal of the Royal Society Interface* 14, no. 134: 20170213. https://doi.org/10.1098/rsif.2017.0213.

Nakagawa, S., and H. Schielzeth. 2013. "A General and Simple Method for Obtaining R2 From Generalized Linear Mixed-Effects Models." *Methods in Ecology and Evolution* 4, no. 2: 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x.

O'brien, R. M. 2007. "A Caution Regarding Rules of Thumb for Variance Inflation Factors." *Quality & Quantity* 41, no. 5: 673–690. https://doi.org/10.1007/s11135-006-9018-6.

Parducci, A. 1965. "Category Judgment: A Range-Frequency Model." *Psychological Review* 72, no. 6: 407–418. https://doi.org/10.1037/h0022602.

Pepper, S., and L. S. Prytulak. 1974. "Sometimes Frequently Means Seldom: Context Effects in the Interpretation of Quantitative Expressions." *Journal of Research in Personality* 8, no. 1: 95–101. https://doi.org/10.1016/0092-6566(74)90049-X.

Petersen, I. T., Z. Demko, P. Doebler, L. Sabel, J. J. Oleson, and R. F. Krueger. 2025. "How Often Is "Often"? Improving Assessment of the Externalizing Spectrum Using Absolute Frequency." *Psychological Assessment* 38: 295–306. https://doi.org/10.1037/pas0001441.

Podsakoff, P. M., S. B. MacKenzie, and N. P. Podsakoff. 2012. "Sources of Method bias in Social Science Research and Recommendations on How to Control It." *Annual Review of Psychology* 63, no. 1: 539–569. https://doi.org/10.1146/annurev-psych-120710-100452.

Portnoy, D. B., A. R. Kaufman, W. M. Klein, T. A. Doyle, and M. De Groot. 2014. "Cognitive and Affective Perceptions of Vulnerability as Predictors of Exercise Intentions Among People With Type 2 Diabetes." *Journal of Risk Research* 17, no. 2: 177–193. https://doi.org/10.1080/13669877.2013.794153.

Prochaska, J. J., W. F. Velicer, J. O. Prochaska, K. Delucchi, and S. M. Hall. 2006. "Comparing Intervention Outcomes in Smokers Treated for Single Versus Multiple Behavioral Risks." *Health Psychology* 25, no. 3: 380–388. https://doi.org/10.1037/0278-6133.25.3.380.

Renner, B., and T. Reuter. 2012. "Predicting Vaccination Using Numerical and Affective Risk Perceptions: The Case of A/H1N1 Influenza." *Vaccine* 30, no. 49: 7019–7026. https://doi.org/10.1016/j.vaccine.2012.09.064.

Reyna, V. F., and C. J. Brainerd. 1995. "Fuzzy-Trace Theory: Some Foundational Issues." *Learning and Individual Differences* 7, no. 2: 145–162. https://doi.org/10.1016/1041-6080(95)90028-4.

Reyna, V. F., J. C. Corbin, R. B. Weldon, and C. J. Brainerd. 2016. "How Fuzzy-Trace Theory Predicts True and False Memories for Words, Sentences, and Narratives." *Journal of Applied Research in Memory and Cognition* 5, no. 1: 1–9. https://doi.org/10.1016/j.jarmac.2015.12.003.

Robert Pace, C., and J. Friedlander. 1982. "The Meaning of Response Categories: How Often Is "Occasionally," "Often," and "Very Often?"" *Research in Higher Education* 17, no. 3: 267–281. https://doi.org/10.1007/bf00976703.

Rose, J. P. 2010. "Are Direct or Indirect Measures of Comparative Risk Better Predictors of Concern and Behavioural Intentions?" *Psychology and Health* 25, no. 2: 149–165. https://doi.org/10.1080/08870440802340164.

Schaeffer, N. C. 1991. "Hardly Ever or Constantly? Group Comparisons Using Vague Quantifiers." *Public Opinion Quarterly* 55, no. 3: 395–423. https://doi.org/10.1086/269270.

Schneider, S., and A. A. Stone. 2016. "The Meaning of Vaguely Quantified Frequency Response Options on a Quality of Life Scale Depends on Respondents' Medical Status and Age." *Quality of Life Research* 25, no. 10: 2511–2521. https://doi.org/10.1007/s11136-016-1293-7.

Schönbrodt, F. D., and M. Perugini. 2013. "At What Sample Size Do Correlations Stabilize?" *Journal of Research in Personality* 47, no. 5: 609–612. https://doi.org/10.1016/j.jrp.2013.05.009.

Schwarz, N. 1999. "Self-Reports: How the Questions Shape the Answers." *American Psychologist* 54, no. 2: 93–105. https://doi.org/10.1037/0003-066X.54.2.93.

Schwarz, N., and D. Oyserman. 2001. "Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction." *American Journal of Evaluation* 22, no. 2: 127–160. https://doi.org/10.1177/109821400102200202.

Stewart, N., N. Chater, and G. D. Brown. 2006. "Decision by Sampling." *Cognitive Psychology* 53, no. 1: 1–26. https://doi.org/10.1016/j.cogpsych.2005.10.003.

Sudman, S., N. M. Bradburn, and N. Schwarz. 1996. *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass.

Suls, J., and L. Wheeler. 2000. "A Selective History of Classic and Neo-Social Comparison Theory." In *Handbook of Social Comparison: Theory and Research*, 3–19. Springer US. https://doi.org/10.1007/978-1-4615-4237-7_1.

Tabachnick, B. G., and L. S. Fidell. 2007. *Experimental Designs Using ANOVA*, Vol. 724. Thomson/Brooks/Cole.

Tourangeau, R., L. J. Rips, and K. Rasinski. 2000. The Psychology of Survey Response. https://doi.org/10.1017/CBO9780511819322.

Wänke, M. 2002. "Conversational Norms and the Interpretation of Vague Quantifiers." *Applied Cognitive Psychology* 16, no. 3: 301–307. https://doi.org/10.1002/acp.787.

Woltz, D. J., M. K. Gardner, J. C. Kircher, and J. J. Burrow-Sanchez. 2012. "Relationship Between Perceived and Actual Frequency Represented by Common Rating Scale Labels." *Psychological Assessment* 24, no. 4: 995–1007. https://doi.org/10.1037/a0028693.

Wright, P. M., G. C. McMahan, and A. McWilliams. 1994. "Human Resources and Sustained Competitive Advantage: A Resource-Based Perspective." *International Journal of Human Resource Management* 5, no. 2: 301–326. https://doi.org/10.1002/acp.2350080506.

Zell, E., and M. D. Alicke. 2009. "Self-Evaluative Effects of Temporal and Social Comparison." *Journal of Experimental Social Psychology* 45, no. 1: 223–227. https://doi.org/10.1016/j.jesp.2008.09.007.

Zell, E., and M. D. Alicke. 2010. "The Local Dominance Effect in Self-Evaluation: Evidence and Explanations." *Personality and Social Psychology Review* 14, no. 4: 368–384. https://doi.org/10.1177/1088868310366144.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Analyses examining the added predictive power of each direct comparison measure above absolute frequency estimates in Study 2. **Table S2:** Analyses examining the added predictive power of each conceptually relevant comparison measure above absolute frequency estimates and distant social group comparisons in Study 2.

Never…Often? Comparisons that Shape People's Likert-Type Ratings of Behavior Frequencies

Jeremy D. Strueder, Jane E. Miller, Isaac T. Petersen, Paul D. Windschitl

Supplemental Materials

Section A: Details on Models Testing Added Predictive Power of Direct Comparisons in Study 2 p. 2-3

Section B: Details on Models Assessing the Role of Methodological Similarities in Study 2 p. 4-5

Section A: Details on Models Testing Added Predictive Power of Direct Comparisons in Study 2

Table S1 contains the full results of the analyses examining the added predictive power of each direct comparison measure above absolute frequency estimates in Study 2.

Table S1. Analyses examining the added predictive power of each direct comparison measure above absolute frequency estimates in Study 2.

							95% CI for β		Model Comparison Results		
Model	**Predictors**	**β**	**SE β**	***df***	***t***	***p***	**Lower-bound**	**Upper-bound**	**Marginal R^2**	**ΔMarginal R^2**	***Sig.* X^2**
Base	Self-Absolute Frequency	0.75	0.02	1570.43	45.59	<.001	0.717	0.782	0.541	-	-
Base + Expert Comparison	Self-Absolute Frequency	0.56	0.02	1571.52	28.07	<.001	0.52	0.60	0.579	0.038	<.001
	Expert Comparison	0.28	0.02	1560.62	14.36	<.001	0.24	0.31			
Base + Global Comparison	Self-Absolute Frequency	0.51	0.02	1572.24	24.62	<.001	0.46	0.55	0.618	0.077	<.001
	Global Comparison	0.33	0.02	1569.94	17.26	<.001	0.30	0.37			
Base + Local Comparison	Self-Absolute Frequency	0.60	0.02	1572.24	32.13	<.001	0.56	0.64	0.592	0.051	<.001
	Local Comparison	0.25	0.02	1569.31	14.50	<.001	0.22	0.28			
Base + Distant Comparison	Self-Absolute Frequency	0.61	0.02	1572.82	31.26	<.001	0.57	0.65	0.592	0.051	<.001
	Distant Comparison	0.23	0.02	1572.37	12.01	<.001	0.19	0.26			
Base + Previous Self Comparison	Self-Absolute Frequency	0.72	0.02	1571.64	42.27	<.001	0.69	0.75	0.549	0.008	<.001
	Previous Self Comparison	0.09	0.02	1569.87	5.65	<.001	0.06	0.12			

Note. Standardized predictors. 95% confidence intervals based on profile-likelihood estimation.

Section B: Details on Models Assessing the Role of Methodological Similarities in Study 2

Table S2 contains the full results of the analyses examining the extent to which the four relevant direct comparison types explain variance in self-LBF judgments beyond self-absolute frequency estimates and distant social group (i.e., non-relevant) comparisons in Study 2.

Table S2. Analyses examining the added predictive power of each conceptually relevant comparison measure above absolute frequency estimates and distant social group comparisons in Study 2.

							95% CI for β		Comparison to Base Model		
Model	**Predictors**	**β**	**SE β**	***df***	***t***	***p***	**Lower-bound**	**Upper-bound**	**Marginal R^2**	**ΔMarginal R^2**	***Sig.*** X^2
Base	Self-Absolute Frequency	0.61	0.02	1572.82	31.26	<.001	0.57	0.65	0.592	-	-
	Distant Comparison	0.23	0.02	1572.37	12.01	<.001	0.19	0.26			
Base + Global Comparison	Self-Absolute Frequency	0.49	0.02	1570.49	23.39	<.001	0.45	0.53	0.626	0.035	<.001
	Distant Comparison	0.08	0.02	1565.93	4.02	<.001	0.04	0.13			
	Global Comparison	0.27	0.02	1566.58	12.59	<.001	0.24	0.33			
Base + Local Comparison	Self-Absolute Frequency	0.55	0.02	1571.73	27.59	<.001	0.51	0.59	0.612	0.021	<.001
	Distant Comparison	0.14	0.02	1570.45	6.75	<.001	0.10	0.18			
	Local Comparison	0.20	0.02	1567.42	10.39	<.001	0.16	0.23			
Base + Expert Comparison	Self-Absolute Frequency	0.52	0.02	1571.12	25.26	<.001	0.48	0.57	0.602	0.010	<.001
	Distant Comparison	0.13	0.02	1571.90	6.49	<.001	0.09	0.17			
	Expert Comparison	0.21	0.02	1560.17	10.03	<.001	0.17	0.26			
Base + Previous Self Comparison	Self-Absolute Frequency	0.60	0.02	1571.02	30.26	<.001	0.56	0.63	0.596	0.004	<.001
	Distant Comparison	0.21	0.02	1571.67	11.38	<.001	0.18	0.25			
	Previous Self Comparison	0.06	0.01	1569.09	4.26	<.001	0.03	0.09			

Note. Standardized predictors. 95% confidence intervals based on profile-likelihood estimation.