
RESEARCH ARTICLE

The Real World is the Ingroup World: A Normative Explanation of Response-Scale Effects in the Subjective Report of Behaviors

Fabio Lorenzi-Cioldi, Vincenzo Iacoviello and Clara Kulich

The ranges of response alternatives presented in frequency scales influence respondents' behavioral estimates. This research aimed at complementing the existing cognitive interpretations of this scale effect (e.g., Schwarz, 1994). We propose a normative interpretation, according to which targets associated with generic behavioral norms, and with ingroup norms, lead participants to rely more strongly on the scale's frequency properties. Studies 1–2 showed stronger scale effects when participants reported behaviors over an extended (vs. short) time period, specifically when they reported behaviors of themselves (vs. people in general). Studies 3–6 showed that the use of a scale's properties in self-reports increased when participants were led to believe that the scale's frequency alternatives mirrored typical ingroup (vs. outgroup) behaviors. Finally, Study 7 demonstrated the role of ingroup identification in the production of these scale effects. Collectively, these findings suggest that behavioral estimates based on shared norms override memory scanning when behavior is difficult to retrieve, and when relevant ingroup information is available.

Keywords: response scales; frequency alternatives; subjective states; norms; behavioral reports

Survey questions about people's behaviors are often questions on the frequency of occurrence of these behaviors. Past research has demonstrated substantial effects of the ranges of frequency alternatives presented in a response scale on how respondents report their behavior (e.g., Bradburn, Sudman, & Wansink, 2004; Krosnick & Presser, 2010; Schaeffer & Presser, 2003; Schuman & Presser, 1981; Schwarz, 1990, 1994, 1999; Schwarz et al., 2008). Instead of devoting effort to retrieve from memory a stored information about the behavior, respondents often estimate the frequency of this behavior based on the characteristics of the response scale, in particular by assuming that the midpoint of the scale represents the typical, average behavior. Seminal evidence for this phenomenon comes from research by Schwarz, Hippler, Deutsch, and Strack (1985). The authors asked participants to estimate the average number of hours spent per day watching television. Participants received either a low frequency scale, with increments of thirty minutes from 'up to a half hour' to 'more than two and a half hours', or a high frequency scale with similar increments but from 'up to two and a half hours' to 'more than four and a half hours'. The findings demonstrated a robust scale effect. Participants

responding with the high frequency scale provided larger estimates of their behavior than participants responding with the low frequency scale. This finding has been replicated in a wide range of scenarios and settings (Schwarz et al., 1991; Tourangeau & Smith, 1996; see Schwarz, 1994, for a review).

Two main and non-mutually exclusive interpretations of scale effects have been proposed. One interpretation focuses on the respondents' desire to convey favorable impressions of themselves by not appearing unusual, eccentric, or even deviant (e.g., Mingay et al., 1994). Accordingly, respondents avoid selecting options from either extreme of the scale, preferring middle options. This impression-management strategy would automatically produce a greater estimation of the targeted behavior in the high frequency than in the low frequency scale. The other interpretation rests on a principle of cognitive economy (e.g., Kahneman, 2011; Wyer & Srull, 1986). Respondents do not necessarily reconstruct all the occurrences of a behavior from memory, but rather estimate, quickly and with little effort, the frequency of the behavior. In doing so, they anchor their judgment on the information provided by the frequency alternatives of the response scale, assuming that the median choice represents the average or typical amount of the targeted behavior in the population at large. This, however, is likely to be contingent on the respondents' actual knowledge of the empirical distribution of the behavior (Schwarz &

Oyserman, 2001). Such knowledge depends on various factors, and foremost on shared behavioral norms, in society at large or in specific groups.

Generic Behavioral Norms

A strategy of recall-and-count of behavioral episodes is less feasible when respondents provide estimates of another person's behavior than of their own, and when they provide estimates of behaviors that occurred over a long period of time than a short one. In the former cases, reports are likely to rely less strongly on memory and more strongly on the shared behavioral standards inferred from the frequency alternatives of the response scale, producing greater scale effects. Schwarz and Bienias (1990) replicated the television estimation task by asking students to report television consumption of specific targets (oneself and a close friend), and of a more generic target (a typical student on campus). Scale type did not affect the judgment about the self and only marginally affected the judgment about the close friend. This can reasonably be attributed to awareness of one's own behavior and to familiarity with the habits and lifestyle of a person the respondent is close to. Conversely, the scale effect proved to be pronounced when the estimation was carried out on a more distal target, the typical student. The first aim of the present research is to further test the generic norm perspective, by examining temporal frame as a moderator of the scale effect. More specifically, we argue that the difficulty to retrieve information over an extended period of time should produce stronger scale effects than if retrieval concerns information from recent experiences, for which memory is still accessible. From this follows that estimates of one's own behavior should be more strongly affected by the type of scale in distant time frames (compared to proximal), while judgements of other people should not be affected by differential time frames as participants should always rely on the normative features of the scale when making judgment for such an abstract target.

Ingroup Behavioral Norms

According to a normative perspective, the tendency to rely on the scale's properties should increase as the inferred norm is perceived as relevant to the self. In this vein, Sudman, Bradburn, and Schwarz (1996) argued that scale effects may be reduced or eliminated when the informational value of the response alternatives is called into question. To illustrate their claim, they mentioned the hypothetical case in which university students would be asked to use scales that have supposedly been taken from a survey on an outgroup's (e.g., the elderly) typical behaviors (p. 221). To our knowledge, this specific idea has never been tested empirically. Therefore, the second aim of the present research was to examine scale effects in settings in which ingroup standards versus outgroup standards are salient. If participants (e.g., women) were informed that the researcher based the design of a response scale on the findings obtained from studying the habits of the gender ingroup (other women) rather than the gender outgroup (men), we would expect an amplification of the difference between the self-reports made on the two scale formats.

Furthermore, a germane assumption of this normative perspective is that ingroup identification should play a role in the emergence of scale effects. The social identity perspective maintains that respondents are more greatly influenced by the desire to conform to the behavioral norms of their ingroup than of an outgroup (e.g., Hogg & Abrams, 1998; Miller & Prentice, 1996; Reese, Steffens, & Jonas, 2013; Turner, 1991). This literature shows that different norms – that is, different sets of expectations concerning individuals' beliefs, modes of conduct, and behaviors – are attached to different groups, and that nonconformity to ingroup norms increases a person's psychological uncertainty (see Hogg, 2000) and decreases self-esteem (e.g., Leary, 2005; Iacoviello et al., 2017). Such unpleasant state may invigorate the respondents' motivation to rely on the response alternatives provided by a response scale in order to endorse the typical occurrence of a behavior in their own group, and therefore to achieve conformity with this group. As high identifiers are particularly prone to conform to the group norm (e.g., Chao, Zhang, & Chiu, 2009), we expect the scale effect to increase as a function of ingroup identification. Such a tendency would support the normative interpretation of scale effects by moving beyond explanations based on individual concerns (e.g., individualistic premises as represented by cognitive economy, impression-management motives, and self-presentation concerns). Indeed, if people map their responses onto the scale, not only because the scale is informative of the ingroup's behaviors when memory is poor, but because they want to *conform* to the ingroup norm, then the extent to which the group 'matters', as assessed by ingroup identification, should be a likely moderator of the scale effect.

Hypotheses

The baseline hypothesis (H1) of the present studies is that a high frequency scale should produce higher estimations of the target's behavior, compared to a low frequency scale. The remaining hypotheses are the central focus of the present research, as they describe moderations of this scale effect. We argue that distal targets (the last 30 days, and people in general), in comparison with more proximal targets (yesterday, and the self), are likely to be more poorly represented in memory, so that respondents are led to base their estimates on the assumed empirical distribution of the behavior as it is represented in the scale's frequency alternatives. The second hypothesis thus concerns the tendency to estimate behavior with the use of the scale's properties, rather than retrieving and counting single episodes of the behavior from memory, according to the temporal frame and type of target. Specifically, H2a predicts that the scale effect should be of greater magnitude when people report on the behavior of a target on an extended time length (the last 30 days) compared to when they report on a short time length (yesterday). Moreover, H2b predicts that time frame should only affect judgements of the self and not judgments of people in general, as participants should rely more exclusively on the normative properties of the scale, regardless of time frame, when making judgements of this latter target.

Following Sudman et al. (1996), we extend this interpretation of the scale effect to self-reports in group contexts in our third hypothesis. H3 relates to the alleged origin of the alternatives of the frequency scale, reflecting the typical behavior of either a membership group or an outgroup. The impact of the scale type should be accentuated when respondents are informed that the alternatives presented in the response scale echo their ingroup members' behavioral standards rather than the behavioral standards of an outgroup.

However, this hypothesis does not explicitly distinguish between two normative explanations of the scale effect: one based on informational process (i.e., an ingroup scale telling more about oneself than an outgroup scale) and the other on the motive towards conformity (i.e., self-ingroup identification boosting the informational effect). As a matter of fact, people may make greater use of a scale attributed to an ingroup than to an outgroup because they consider the former as more informative of their own behavior, and/or because they are motivated to conform to the ingroup. H4 disentangles these interpretations by predicting that highly identified group members would show a stronger scale effect than the lowly identified, in particular on prototypical ingroup traits.

Current Research

We conducted seven studies aimed at demonstrating that scale effects can be accounted for by people's tendency to conform to generic norms or ingroup norms. The baseline hypothesis (H1) is tested in all of the studies. Studies 1–3 are concerned with H2a and H2b, that is, the expectation that judgments made on a long period of time should be more prone to scale effects than a short period of time, and especially when judgments are made on the self (vs. people in general). In Studies 3–6, we use split sample procedures to manipulate the origin of the scale, which will be experimentally attributed to either the participant's ingroup or outgroup. A variety of social groups (based on gender, region of residence, and social stigma) will be considered. As predicted in H3, the scale effect should emerge more strongly when the frequency alternatives of the scale are attributed to the ingroup than to the outgroup. Finally, Study 7 is concerned with H4. In line with the conformity explanation, the expectation is that participants' use of the scale's frequency alternatives for typical ingroup characteristics should be boosted

among the highly identified with the group, compared to the lowly identified.

Type of frequency scale was the primary independent variable in all studies. In past research, scale effects have predominantly been investigated with response alternatives naming occurrences of concrete behaviors (for instance, number of hours spent watching television, number of drunk glasses of alcohol, etc.). What happens when the behaviors and events being investigated deal with subjective phenomena? This is ordinary practice in much social and psychological research, such as the reporting of emotions or behaviors according to personality traits.

Low and high frequency scales were devised by adapting the seminal Schwarz et al.'s (1985) scale formats in order to tap the subjective, rather than objective, occurrence of behaviors. Thus, unlike the seminal scales, which were based on a progression of temporal intervals, the presently devised scales comprised a progression of frequency options on an ordinal continuum (see **Table 1**). In the scale consisting of response alternatives discriminating at low frequencies, the first three options rose by small magnitudes and the fourth option more abruptly. The scale consisting of response alternatives discriminating at high frequencies worked in precisely the opposite way.

We coded participants' estimates following Schwarz et al.'s (1985) procedure. Responses located between 'never' and 'sometimes' were coded 0 (choices 1 to 3 in the low frequency scale, and choice 1 in the high frequency scale). Responses located between 'sometimes' and 'all the time' were coded 1 (choice 4 in the low frequency scale, and choices 2 to 4 in the high frequency scale). Thus, for each characteristic and type of scale, responses indicated the frequency 'from sometimes to all the time'. With the use of these frequency scales, participants were asked to supply estimates of their own behavior (Studies 1–7) and the behavior of 'people in general' (Study 2) on four general characteristics (taken from Lorenzi-Cioldi, 1988; Studies 1–6), or on two prototypical and two non-prototypical ingroup characteristics (Study 7). These judgments were elicited in specific time spans (short and/or extended past, depending on the study). We finally computed the dependent variable by summing the responses on the four characteristics, producing a score that ranged between 0 and 4 (Studies 1–6), and between 0 and 2 for each of the prototypical and the non-prototypical subsets (Study 7).¹

Table 1: Frequency scales used in this research.

High frequency scale	Low frequency scale
1. <i>From never to sometimes</i>	1. <i>From never to very rarely</i>
2. From sometimes to fairly often	2. <i>From very rarely to occasionally</i>
3. From fairly often to often	3. <i>From occasionally to sometimes</i>
4. From often to all the time	4. From sometimes to all the time
Equivalence of responses offered in high and low frequency scales	
1 (italics)	1–3 (italics)
2–4 (bold)	4 (bold)

We provide supplementary materials with this article. They contain exact wording of all manipulations, as well as all measures that were presented in the questionnaires but are not part of the present article. We performed data analysis with the sample sizes provided herein. No additional data were sought for any of the studies after initial data analysis. Participants in all studies were randomly assigned to experimental conditions. For Studies 1 to 4, and 6, data were collected as part of an exercise in post-graduate seminars. Students distributed paper-pencil questionnaires in public places (e.g., coffee shops, libraries, train stations). Datasets of all studies are available on the Open Science Foundation platform (https://osf.io/j5cuk/?view_only=69e1bf5069094b9ab2c1ff63e8b9e821).

Study 1

In this study, all the judgments pertained to the self, and were made twice, in the short and the extended time period.

Method

Participants

We asked a convenient sample of Swiss French-speaking individuals to participate in a survey about how people in general perceive themselves. A total of 171 persons (86 women and 85 men, $M_{age} = 26.61$, $SD_{age} = 6.94$) agreed to fill in the questionnaire. Seventy-two were employed and 99 were students.

Procedure

Participants were invited to estimate the frequency with which they personally engaged in behaviors related to each of four characteristics ('assertive', 'warm', 'selfish', and 'annoyed'). These judgments referred to 'yesterday' and to 'the last 30 days'.² Participants answered on either the low or the high frequency scale.

Results

Participants' self-reports were submitted to a full-factorial Scale Type \times Temporal Frame ANOVA, with repeated measures on the second factor. The analysis first revealed a main effect of temporal frame, $F(1,169) = 57.15$, $p < 0.001$, $\eta_p^2 = 0.25$. It came as no surprise that the occurrence of responses 'from sometimes to all the time' was greater for the extended period of time than for the short period ($M_s = 2.12$ and 1.56 , $SD_s = 1.23$ and 0.99 , respectively). The analysis also showed the in H1 expected main effect of scale type, $F(1,169) = 889.87$, $p < 0.001$, $\eta_p^2 = 0.84$. Behavioral estimates were greater in the high frequency scale than in the low frequency scale ($M_s = 2.42$ and 1.27 , $SD_s = 0.86$ and 0.75 , respectively). Consistent with H2a, this main effect was qualified by a Scale Type \times Temporal Frame interaction, $F(1,169) = 87.30$, $p < 0.001$, $\eta_p^2 = 0.34$. Though the scale discrepancy was significant for both the short temporal frame, $F(1,169) = 41.13$, $p < 0.001$, $\eta_p^2 = 0.20$, and the extended temporal frame, $F(1,169) = 89.31$, $p < 0.001$, $\eta_p^2 = 0.35$, it was larger in the extended temporal frame ($M_s = 2.85$ and 1.41 , $SD_s = 1.03$ and 0.96 , for the high and the low frequency scales, respectively) than in the short one ($M_s = 2.00$ and 1.13 , $SD_s = 0.95$ and 0.82).³

Discussion

The findings provided support for H1, showing that the high frequency scale leads to an overestimation of the behavior compared to the low frequency scale. They also supported H2a, as judgments carried out for the extended reference period were more sensitive to scale type than judgments for the short reference period. In all likelihood, the extended period of time prevented participants from scanning their memory to retrieve and count specific episodes of their behavior. Self-reports were therefore based more strongly on the frequency alternatives of the proposed scale. In sum, respondents gave more weight to the assumed empirical distribution of the targeted behavior suggested by the scale in the context of a distal (the last 30 days) than a proximal (yesterday) temporal context.

Study 2

Study 2 sought to replicate and extend the previous findings by comparing the scale effect on self-perception (proximal target) and perception of 'people in general' (distal target), in the long and the short temporal frames. According to H2b, the expectation was that judgments made on a long temporal frame (i.e., the last 30 days) would be more prone to scale effects than on a short temporal frame (i.e., yesterday), and especially when the judgements are made on the self (vs. people in general).

Method

Participants

A total of 119 Swiss French-speaking individuals (58 women and 61 men, $M_{age} = 30.18$, $SD_{age} = 10.05$) agreed to complete a questionnaire about people's self-perception. A sensitivity power analysis using G*Power (Faul et al., 2009) suggests that this sample size provided 80% power to detect effect sizes of $\eta_p^2 = 0.01$ or greater ($\alpha = 0.05$). Considering that the effect size of the Scale Type \times Temporal Frame interaction in Study 1 was of greater magnitude (i.e., $\eta_p^2 = 0.34$), the present sample size appears well-powered.

Procedure

Participants estimated the frequency with which they personally engaged in behaviors related to each of the four characteristics and how frequently people in general engaged in these behaviors. As in Study 1, participants answered on either the high or the low frequency scale, and these judgments were made for both targets in the extended and the short time length.⁴

Results

The full-factorial Scale Type \times Temporal Frame \times Target ANOVA on participants' judgments, with repeated measures on the last two factors, confirmed the main effect of type of scale, $F(1,117) = 162.22$, $p < 0.001$, $\eta_p^2 = 0.58$. The occurrence of responses 'from sometimes to all the time' was greater in the high frequency scale than in the low frequency scale ($M_s = 2.89$ and 1.36 , $SD_s = 0.65$ and 0.66 , respectively). The analysis then showed a Scale Type \times Temporal Frame interaction, $F(1,117) = 9.81$, $p = 0.002$, $\eta_p^2 = 0.08$, and a Scale Type \times Target interaction, $F(1,117) = 26.94$, $p < 0.001$, $\eta_p^2 = 0.19$. Of importance, these effects

were qualified by a Scale Type \times Temporal Frame \times Target interaction, $F(1,117) = 4.47$, $p = 0.04$, $\eta_p^2 = 0.04$ (see **Figure 1**). Indeed, the Scale Type \times Temporal Frame interaction was only significant for self-reports, $F(1,117) = 11.15$, $p = 0.001$, $\eta_p^2 = 0.09$. In support of H2b, inspection of the means showed that the scale effect was of greater magnitude on the extended temporal frame ($M_s = 3.02$ and 1.56 , $SD_s = 0.93$ and 0.83 , for the high and the low frequency scales, respectively) than on the short temporal frame ($M_s = 2.09$ and 1.30 , $SD_s = 0.98$ and 0.74). When the target being judged was 'people in general', the corresponding Scale Type \times Temporal Frame interaction did not reach significance, $F(1,117) = 2.03$, $p = 0.16$, $\eta_p^2 = 0.02$, indicating that the scale discrepancy was not different according to the temporal frame. As a consequence, when judgments were made on the short temporal frame, the scale discrepancy was larger for the target 'people in general' than for the target 'self', $F(1,117) = 28.80$, $p < 0.001$, $\eta_p^2 = 0.20$. The same pattern was observed when judgments were made on the extended temporal frame, though it was of weaker magnitude, $F(1,117) = 10.08$, $p = 0.002$, $\eta_p^2 = 0.08$.

Discussion

Testifying to the robustness of the scale effect, the findings revealed that participants' claim to behave in assertive, warm, selfish, and annoyed manners 'from sometimes to all the time' was stronger when these judgments were to be mapped onto a high compared to a low frequency scale. Importantly, in support of H2b, the findings revealed that the reference period affected self-reports, with the extended time length producing a larger-scale effect than the short time length. Conversely, estimates of the behavior of people in general were consistently based on the scale's properties, regardless of the reference period. In line with previous literature (e.g., Schwarz, 1990), these findings support the interpretation that respondents are less familiar with the behavior of a generic target ('people

in general') than with their own behavior, but that as the temporal frame gets longer, and therefore retrieval processes are increasingly difficult, even self-descriptions become strongly affected by the response scale's properties.

Study 3

In Studies 3–6, we tested H3 by experimentally manipulating the presumed ingroup or outgroup origin of the response scale. In Study 3, half of the participants were told that the researcher had devised the frequency scale using the findings from a current opinion survey in which responses had been collected from other members of their gender group. The other half were given the opposite information and were therefore led to believe that the response scale was based on the responses of the gender outgroup. As in Studies 1 and 2, self-descriptions were then elicited with regard to the extended and the short temporal frames. The main expectation was that the impact of scale type would be more pronounced when the scale was attributed to the gender ingroup than when attributed to the gender outgroup. Furthermore, consistent with the findings from the preceding studies, we also predicted that reference period would moderate this effect, such that the extended time frame would accentuate the norm moderation of the scale effect (as participants should mainly rely on the normative properties of the scale when answering on an extended temporality scale).

Method

Participants

Participants were a convenient sample of 394 Swiss French-speaking people (199 women and 195 men, $M_{age} = 27.04$, $SD_{age} = 7.26$). A sensitivity power analysis suggested the sample size is well-powered as we are able to detect effect sizes of $\eta_p^2 = 0.01$ or greater and considering that

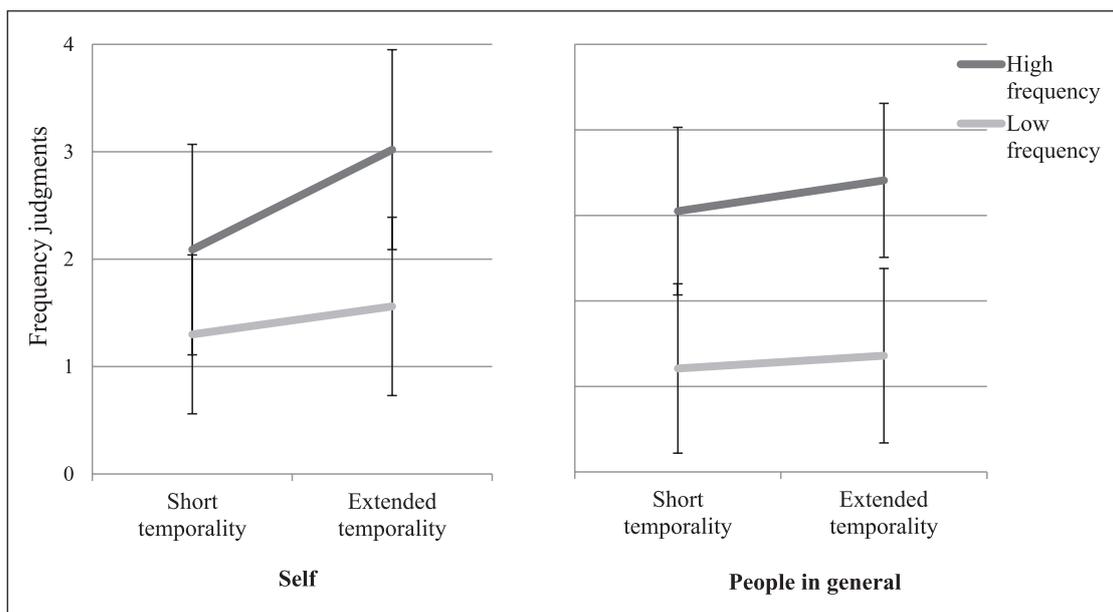


Figure 1: Participants' estimates of the frequency of their behavior on the four characteristics according to temporal frame, target, and scale type (Study 2).

the effect sizes of Studies 1 and 2 were greater than this threshold ($\eta_p^2 = 0.34$ and $\eta_p^2 = 0.04$, respectively).

Procedure

Participants were informed that a vast survey was currently being carried out investigating various aspects of 'men's and women's daily lives'. Half of the participants were then told that thus far only men's responses had been collected and fully analyzed, while the other half were given the same information about women's responses. Participants were then informed that the researchers had used this analysis to design the frequency alternatives of a response scale. The presented scale was actually of high or low frequency. Thus, the ingroup norm condition combined men responding on the frequency scales attributed to other men, and women responding on the frequency scales attributed to other women. The outgroup condition combined men responding on the women's scales and women responding on the men's scales. The remainder of the study was identical to Study 1, with self-reports assessed for 'the last 30 days' and 'yesterday'.

Results

The full-factorial Scale Type \times Temporal Frame \times Norm \times Participant Gender ANOVA on participants' self-reports, with repeated measures on the second factor, showed a main effect of type of scale, $F(1, 386) = 281.85$, $p < 0.001$, $\eta_p^2 = 0.42$, a main effect of temporal frame, $F(1, 386) = 74.19$, $p < 0.001$, $\eta_p^2 = 0.16$, and a Scale Type \times Temporal Frame interaction, $F(1, 386) = 44.59$, $p < 0.001$, $\eta_p^2 = 0.10$. As in Studies 1 and 2, the scale discrepancy was larger in the extended temporal frame ($M_s = 2.90$ and 1.34 , $SD_s = 0.89$ and 0.91 , for the high and low frequency scales, respectively) than in the short temporal frame ($M_s = 2.10$ and 1.23 , $SD_s = 0.89$ and 0.86). Importantly, this pattern of means was further qualified by a Scale Type \times Temporal Frame \times Norm interaction, $F(1, 386) = 4.31$, $p = 0.04$, $\eta_p^2 = 0.01$ (see **Table 2**). Contrast analyses showed the expected Scale Type \times Norm interaction in the extended temporal frame, $F(1, 386) = 4.55$, $p = 0.03$, $\eta_p^2 = 0.01$. The scale discrepancy was larger in the ingroup norm condition ($M_s = 2.98$ and 1.22 , $SD_s = 0.87$ and 0.77 , for the high and low frequency scales, respectively) than in the outgroup norm condition ($M_s = 2.82$ and 1.45 , $SD_s = 0.90$ and 1.02). Conversely, no Scale Type \times Norm interaction was observed in the short temporal frame, $F(1, 386) = 0.08$, $p = 0.77$, $\eta_p^2 < 0.01$. The Scale Type \times Temporal Frame \times Norm \times Participant Gender interaction was not significant, $F(1, 386) = 0.28$, $p = 0.61$, $\eta_p^2 < 0.01$, suggesting that this overall pattern of means did not differ according to participant gender.

Discussion

The findings from Study 3 demonstrated once again the importance of the type of frequency scale. But they also provided initial evidence for the crucial hypothesis (H3) that anchoring the scale to an ingroup or an outgroup impacts on how intensely individuals allow themselves to be influenced by the frequency alternatives of the

scale when making self-descriptions. Our findings further revealed that it is possible to eliminate the norm moderation of the scale effect. When self-reports were made on the short temporal frame, people seemed to devote cognitive effort to recall-and-count behavioral episodes. As a consequence, the scale effect was not moderated by the normative properties of the scale. Conversely, the norm moderation was present when self-reports pertained to the extended period of time, boosting participants' tendency to model their estimates on the ingroup's scale properties. In more general terms, these findings speak in favor of our normative interpretation of the scale effects. The next three studies aimed at replicating the norm moderation of the scale effect by varying the type of group, which provided the alleged norm of the response scale.

Study 4

Method

Participants

French-speaking participants living in the region of Geneva (Switzerland) were asked to participate in a survey about 'the habits of Geneva's citizens'. A total of 41 people (21 women and 20 men, $M_{age} = 24.51$, $SD_{age} = 3.84$) agreed to complete the questionnaire. It is noteworthy that a sensitivity power analysis suggested the sample size may be underpowered as we are able to detect effect sizes of $\eta_p^2 = 0.17$ or greater and considering that the effect size of the Scale Type \times Norm interaction in the extended temporal frame in Study 3 (which is similar to the design of the present study) was weaker than this threshold ($\eta_p^2 = 0.01$).

Procedure

The procedure was similar to Study 3, except that self-descriptions were elicited exclusively for the extended time period. Participants were informed about a vast opinion survey that had just been carried out in the Canton of Geneva and the Canton of Vaud (a neighboring Swiss region), and that only the responses of people living in the Canton of Geneva (ingroup norm condition) or the responses of people living in the Canton of Vaud (outgroup norm condition) had been analyzed and used to devise the response scale. They then estimated the frequency of the behaviors related to the four characteristics on either the high or the low frequency scale.

Results and Discussion

A Scale Type \times Norm ANOVA on participants' self-reports first substantiated the main effect of scale type, $F(1,37) = 37.32$, $p < 0.001$, $\eta_p^2 = 0.50$. Consistent with H3, this effect was qualified by a Scale Type \times Norm interaction, $F(1,37) = 4.84$, $p = 0.03$, $\eta_p^2 = 0.12$ (see **Table 2**). The scale discrepancy was larger in the ingroup norm condition ($M_s = 3.67$ and 1.36 , $SD_s = 0.49$ and 1.03 , for high and low frequency scales, respectively) than in the outgroup norm condition ($M_s = 2.75$ and 1.67 , $SD_s = 0.97$ and 0.82). Despite varying the context for self-description, Study 4 replicated the findings from Study 3, showing the expected larger scale effect in the condition where participants were primed with an ingroup norm.

Table 2: Mean Frequencies (Standard Deviations in Brackets) of the Behaviors Related to the Summed Four Characteristics, According to Temporal Frame, Scale Type and Group Norm (Studies 3–6).

Temporal frame	Scale type	Study 3 (Gender group)		Study 4 (People from Geneva)		Study 5 (Hearing loss people)		Study 6 (Weight group)	
		Ingroup	Outgroup	Ingroup	Outgroup	Ingroup	Outgroup	Ingroup	Outgroup
Extended	High frequency	2.98 (0.87)	2.82 (0.90)	3.67 (0.49)	2.75 (0.97)	2.76 (1.03)	2.05 (0.97)	2.90 (0.79)	2.65 (1.04)
	Low frequency	1.22 (0.77)	1.45 (1.02)	1.36 (1.03)	1.67 (0.82)	1.18 (0.85)	1.32 (0.58)	1.73 (0.88)	1.92 (0.91)
Short	High frequency	2.08 (0.82)	2.11 (0.96)	–	–	–	–	–	–
	Low frequency	1.24 (0.84)	1.22 (0.89)	–	–	–	–	–	–

Study 5

Method

Participants and procedure

Data were collected in the context of a Master thesis. The student distributed paper-pencil questionnaires in an institution for people with hearing loss, where participants were invited to participate in a survey about 'hearing loss and well-being'. Seventy-seven of them (34 women and 43 men, $M_{age} = 32.47$, $SD_{age} = 11.09$) agreed to complete the questionnaire. All of them had total loss of hearing from at least 15 years of age, and the majority (62.3%) since birth. The procedure was identical to Study 4, except that the norm conditions were based on a supposedly recent survey conducted on people with (ingroup) and without (outgroup) hearing loss. A sensitivity power analysis using G*Power (Faul et al., 2009) suggested that this sample size allowed to detect effect sizes of $\eta_p^2 = 0.09$ or greater ($\alpha = 0.05$). Considering that the effect size in Study 3 ($\eta_p^2 = 0.01$) was weaker than this threshold, but that the effect size in Study 4 ($\eta_p^2 = 0.12$) was greater, preliminary knowledge gives inconclusive evidence on the power of the present study.

Results and Discussion

The full-factorial Scale Type \times Norm ANOVA on participant self-reports showed the main effect of scale type, $F(1,73) = 33.85$, $p < 0.001$, $\eta_p^2 = 0.32$, and this effect was once again qualified by a Scale Type \times Norm interaction, $F(1,73) = 4.50$, $p = 0.04$, $\eta_p^2 = 0.06$ (see **Table 2**). In line with H3, the tendency to overestimate the frequency of one's behaviors in the high frequency scale, as compared with the low frequency scale, was larger in the ingroup norm condition ($M_s = 2.76$ and 1.18 , $SD_s = 1.03$ and 0.85 , for the high and the low frequency scales, respectively) than in the outgroup norm condition ($M_s = 2.05$ and 1.32 , $SD_s = 0.97$ and 0.58).

Study 6

In this study, the group norm pertained to people's weight, distinguishing between people of normal weight and overweight people.

Method

Participants

A convenient sample of 80 participants agreed to complete a questionnaire 'about people's physical appearance and well-being' (40 women and 40 men, $M_{age} = 32.97$, $SD_{age} = 12.62$). A sensitivity power analysis suggested that we are able to detect effect sizes of $\eta_p^2 = 0.09$ or greater. Since only the effect size of Study 4 ($\eta_p^2 = 0.12$), but not Studies 3 and 5 ($\eta_p^2 = 0.01$ and $\eta_p^2 = 0.06$, respectively), is greater than this threshold, there is still doubt about the present study being well-powered.

Procedure

The procedure was similar to Studies 4 and 5. Participants were informed that a vast survey about people's physical appearance and well-being had just been carried out. Half of them were told that thus far only the responses of people of normal weight had been analyzed, while the other

half were told that only the responses of overweight people had been analyzed. The researcher further informed participants that the findings from this initial analysis had been used to design a response scale. Participants then examined the scale, which was either of high or low frequency, and reported the frequency of their own behavior on the four characteristics. At the end of the survey, participants provided their demographics, which included their height and weight.

Norm variable

Participant height and weight were used to compute participants' body mass index (BMI). Forty-eight participants had a calculated BMI of 18.5 to 25, and 25 had a BMI higher than 25. According to BMI standards,⁵ the former were considered to be people of normal weight and the latter as overweight people. Seven participants had a calculated BMI lower than 18.5. They were excluded from the analyses because they were classified as underweight, and therefore did not fit in with the ingroup-outgroup norm distinction between normal weight and overweight. The ingroup norm condition combined participants with normal weight responding on the scale allegedly based on responses from normal weight participants in the earlier study, and overweight people responding on the scale allegedly based on responses from overweight participants. The outgroup norm condition was based on the reverse combination of participant BMI and scale's origin.

Results and Discussion

The full-factorial Scale Type \times Norm \times BMI Membership (normal weight vs. overweight) ANOVA on participant self-reports first substantiated the main effect of type of scale, $F(1, 65) = 11.70$, $p < 0.01$, $\eta_p^2 = 0.15$. This effect was qualified by a Scale Type \times Norm interaction, $F(1, 65) = 3.60$, $p = 0.03$, $\eta_p^2 = 0.07$ (see **Table 2**). As expected, the scale discrepancy was of larger magnitude in the ingroup condition ($M_s = 2.90$ and 1.73 , $SD_s = 0.79$ and 0.88 , respectively) than in the outgroup condition ($M_s = 2.65$ and 1.92 , $SD_s = 1.04$ and 0.91). This pattern of means did not vary according to BMI groups, $F(1, 65) < 1$, $p = 0.43$, $\eta_p^2 = 0.01$.

These findings provide support for H3 and complement those of Studies 3 and 4, which used gender and region of residence, respectively, to create ingroup and outgroup norms with no particular value asymmetries. In contrast, Studies 5–6 introduced group norms of more obvious value discrepancies. Indeed, people with hearing loss and overweight people are potentially disparaged social categories.

It is worth mentioning that in the present study the normal weight versus overweight participant group assignment was more implicit than the group memberships primed in the previous studies. Indeed, we created the participant ingroup and outgroup based on the post-hoc calculation of participant body mass index. Contrary to Studies 3–5, where participants were ostensibly assigned to a group membership (on the basis of gender in Study 3, place of residence in Study 4, and deafness in Study 5), in the present study participants were not explicitly told to

which group they belonged. Accordingly, in our research, scale origin (normal weight vs. overweight) was of little informational value to the participants. Despite this unobtrusive character of group assignment, the findings revealed a similar pattern of means as in the previous studies. The next study will examine the role of ingroup identification. To the extent that the conformity motive prevails over a mere informational tendency, we should observe that highly identified group members show stronger scale effects than lowly identified group members on an identical ingroup scale.

Study 7

Method

Participants

In a post-graduate research seminar, students had the task to recruit about 20 participants each. They did so by asking people in their surroundings and by posting the link to the study on social media. Participants were invited to participate in an online survey on 'people's aesthetic preferences'. A total of 273 Swiss French-speaking people (185 women, 74 men, and 14 gender-unspecified, $M_{\text{age}} = 24.19$, $SD_{\text{age}} = 6.43$) accepted to participate. A sensitivity power analysis suggested that we are able to detect effect sizes of $\eta_p^2 = 0.03$ or greater. Since the effect sizes of all studies were greater (all $\eta_p^2 > 0.06$), except Study 1 ($\eta_p^2 = 0.01$), we are fairly confident that the present study is well-powered.

Procedure

A minimal group procedure was used to create an intergroup context (see Lorenzi-Cioldi, 1998). Participants engaged in an aesthetic judgement task in which they expressed preferences in the field of arts. The experimenter told the participants that they were going to see pieces of paintings by two contemporary painters named 'Dusek' and 'Tausig'. The participants' task would simply be to choose the preferred paintings. Participants were eventually assigned to the group Dusek or the group Tausig according to their preferences (in actuality, they were randomly assigned to one of the two groups). Right after the group assignment, we assessed ingroup identification using a pictorial measure of the inclusion of the self within the ingroup (see Aron, Aron, & Smollan, 1992). Participants were shown seven pairs of circles that increasingly overlapped. In each pair, a small circle represented them as an individual and a larger circle represented their membership group. Participants were asked to choose the pair that best represented 'their level of identification with the ingroup'. The closer the circles, the greater the ingroup identification ($M = 3.57$, $SD = 1.84$).

Participants then estimated the frequency of the behaviors related to four characteristics on either the high or the low frequency scales. Two characteristics were prototypical of the ingroup ('creativity' and 'imagination'), and two characteristics were non-prototypical ('pragmatism' and 'logical mind'). These characteristics were chosen on the basis of a pilot study that showed that the former characteristics are related to the field of art, whereas the latter are not (see Iacoviello, Lorenzi-Cioldi, & Chipeaux, 2018).

To provide a further check of the characteristics' prototypicality, we asked participants from the present study to rate these characteristics on a 7-point scale (1 = *not at all typical*, 7 = *entirely typical*) and submitted these judgements to a principal components analysis with varimax rotation. The two prototypical characteristics loaded on component 1 ($\lambda = 1.93$, 48.46% of the total variance, with loadings greater than 0.91), and the two non-prototypical characteristics loaded on component 2 ($\lambda = 1.26$, 31.56%, with loadings greater than 0.98). The self-descriptions always took place in the extended time frame.

Results

To test H4, we performed a repeated-measures ANCOVA on participants' self-descriptions, with characteristics' prototypicality of the ingroup (within-subjects variable), scale type (between-subjects variable), ingroup identification (centered continuous variable), and their interactions, as predictors. The analysis first showed a main effect of characteristics' prototypicality, $F(1, 269) = 54.84$, $p < 0.001$, $\eta_p^2 = 0.17$. The score of the prototypical characteristics was globally higher than the score of the non-prototypical characteristics ($M_s = 1.07$ and 0.54 , and $SE_s = 0.05$ and 0.13 , respectively). The analysis corroborated the main effect of scale type, $F(1, 269) = 143.09$, $p < 0.001$, $\eta_p^2 = 0.35$, and showed a marginally significant Scale Type \times Identification interaction, $F(1, 269) = 3.70$, $p = 0.06$, $\eta_p^2 = 0.01$. The scale discrepancy was larger among high identifiers (+1 SD; $M_s = 3.45$ and 1.71 , and $SE_s = 0.12$ and 0.13 , for the high and low frequency scales, respectively) than among low identifiers (-1 SD; $M_s = 3.23$ and 1.98 , and $SE_s = 0.13$ and 0.13). Importantly, this effect was further qualified by a Scale Type \times Identification \times Prototypicality interaction, $F(1, 269) = 9.90$, $p = 0.002$, $\eta_p^2 = 0.04$. Inspection of simple effects showed that, on the prototypical characteristics, the scale discrepancy was larger among high identifiers ($M_s = 1.62$ and 0.46 , and $SE_s = 0.09$ and 0.10 , for the high and low frequency scales, respectively) than among low identifiers ($M_s = 1.36$ and 0.83 , and $SE_s = 0.10$ and 0.10), $F(1, 269) = 10.97$, $p = 0.001$, $\eta_p^2 = 0.04$. This tendency was not reliable on the non-prototypical characteristics (among high identifiers: $M_s = 1.83$ and 1.25 , and $SE_s = 0.08$ and 0.08 ; among low identifiers: $M_s = 1.88$ and 1.15 , and $SE_s = 0.08$ and 0.08), $F(1, 269) = 0.86$, $p = 0.35$, $\eta_p^2 < 0.01$.⁶

Discussion

Study 7 added to the preceding studies the critical assessment of ingroup identification. By distinguishing between participants who were highly and lowly identified with their group, the findings showed that the scale effect was of greater magnitude among highly identified participants, especially when judgements were made on the characteristics that were prototypical of the ingroup. This finding allows to distinguish between two likely normative interpretations of the findings about scale effects. In the preceding studies, the larger scale effect that occurred when participants were presented with a scale attributed to the ingroup rather than to the outgroup could either stem from participants' motivation to conform to an

ingroup norm, *or/and* from the ingroup's scale providing more information about the group members' typical behavior. In the present study, despite the fact that all of the participants were presented with a frequency scale with identical informational value about their ingroup's behaviors, only highly identified participants showed a conspicuous scale effect on the prototypical ingroup's characteristics. Therefore, these findings help gaining confidence that ingroup conformity is an important factor accounting for scale effects.

General Discussion

The format of a response scale is an integral part of a question, not simply a means to catch pre-formed opinions. Response scales are not neutral vehicles for displaying opinions and facts. They carry meaning in themselves, revealing much to the respondents who, by examining the scale's response alternatives, draw inferences about the very nature of the targeted phenomenon in the real world (see Tourangeau, 2018, for an account of cognitive processes involved in survey responses). In the present research, we drew a distinction between *generic* norms, which point to behavioral standards in society at large, and *group* norms, which point to behavioral standards in specific groups, and we devised new frequency scales to assess reports of behavior related to personality traits. Scale effects were then examined in controlled experimental designs, using split sample procedures.

People often aim to present themselves as 'normal', and therefore tend to select responses located around the midpoint of the response scale (Schwarz et al., 2008). But such impression-management strategy is not the whole story. In Studies 1–3, this was demonstrated by two main findings: First, self-reports revealed a smaller scale effect than reports about others, presumably because self-descriptions benefit from more factual information than descriptions of other people. Second, the length of the recall period affected reliance on the scale's properties even for self-description, with an extended reference period increasing the scale effect. Apparently, distal targets depleted efforts to scan one's memory. When respondents are faced with such targets, they clearly abandon a retrieve-and-count strategy in favor of the effortless strategy of picking the median response alternative in the frequency scale, because such a response is likely to represent the norm of the behavior.

Studies 3–7 moved one step forward to test this normative interpretation of the scale effect. As Sudman et al. (1996) suggested, the informational value of the response alternatives of a scale may be contingent on a salient group framework. To provide an empirical test of this conjecture, we drew on common evidence in the field of social psychology: In general, individuals are more motivated to conform to ingroup norms than to outgroup norms (e.g., Adida et al., 2016; Hogg & Abrams, 1998; Turner, 1991; Wright, Gaskell, & O'Muircheartaigh, 1994). In support of this reasoning, the findings of Studies 3–6 revealed scale effects of greater magnitude when respondents were led to believe that the response scale was anchored in the typical

behavior of other ingroup members than when they were led to believe that it was anchored in an outgroup.

The findings from Study 7 provided further insights about the normative dynamics at play in scale effects by disentangling between a purely informational process (i.e., participants rely on the ingroup scale because it is more relevant to provide information about themselves) and a process based on the socially shared motive for conformity (i.e., participants rely on the ingroup scale because they are motivated to stick to the ingroup norm). By showing that the scale effect was particularly pronounced among highly identified group members who self-reported on the ingroup's prototypical characteristics, the findings speak in favor of the conformity explanation over the informational one. Indeed, as the scale was always based on ingroup behaviors, the informational value of the scale was of similar relevance for high- and low-identified members. Despite this similarity, high identifiers, compared to low identifiers, demonstrated a greater motivation to conform to the ingroup norm.

Altogether, the findings from this research deviate in significant ways from a simple 'recall-and-count' psychological model of reporting retrospective behavior. The currently favored interpretation of scale effects rests on the assumption that respondents infer from the scale the researcher's knowledge and expectations about the distribution of the behavior in the real world. Our findings add a piece of evidence in favor of the interpretation that people view the real world through their ingroup's lens.

Limitations and Future Research

The findings from this research strongly supported our hypotheses using a variety of targets, reference periods, and group memberships, testifying to the robustness of a conformity explanation of the scale effect. However, this research made an exclusive use of scales composed of so-called vague quantifiers (as opposed to plain numerical values) to assess the occurrence of subjective behaviors. Vague quantifiers may elicit different interpretations among people and across content domains (e.g., Schwarz et al., 1988; Tourangeau, Rips, & Rasinski, 2000). For instance, behaving *selfish* 'from sometimes to all the time' does hardly underpin identical quantities for everybody, and hardly the same quantities as for a trait like *warm*. Moreover, these judgments may suffer from variations due to 'shifting standards': for instance, men may alter their estimations of being selfish when comparing their behavior to the behavior of other men or to the behavior of women (e.g., Biernat & Manis, 1994). The use of vague and general quantifiers was nonetheless motivated by two reasons. First, subjective states (such as those represented in trait adjectives), compared to concrete or mundane behaviors, are not ideally suited for numerical quantification. Commenting on a survey on the frequency with which participants experienced a series of feeling states, Bradburn and Miles (1979) concluded that 'While it is relatively easy for respondents to report how many times in the past week they have been to a movie or how many hours they watched television yesterday, they seem to

have a great deal of difficulty in putting precise numbers on subjective states' (p. 100; see also Schimmack, 2002; Schuman & Kalton, 1985; Turner & Martin, 1985). Second, though participants may have interpreted vague quantifiers differently one from another and across trait adjectives, there is evidence that these variations do not affect, and thus do not preclude, intergroup comparisons in controlled experimental designs (see Nelson Laird, Korkmaz, & Chen, 2008; Wright, Gaskell, & O'Muircheartaigh, 1994). Such comparisons were the only focus in our research using the split sample approach (see also Visser, Krosnick, & Lavrakas, 2000). Nonetheless, future research should ascertain whether vague quantifiers provide effective anchor points to the participants, and their relationship with actual numerical quantities (see Pohl, 1981).

A second limitation should be acknowledged. In addition to the factors involved in the hypotheses of the present research, it is likely that the saliency and the subjective importance of a behavior may be important factors in producing scale effects. Though saliency and importance are plausibly substantially correlated, our research suggests that they may operate in two distinct directions. On the one hand, as Studies 1–3 suggest, saliency of a behavior may facilitate a retrieval and counting strategy of the behaviors, thus reducing the scale effect. On the other hand, subjective importance may invigorate an ingroup rather than an outgroup context, so that, as our Studies 3–7 imply, high importance may produce a more powerful scale effect. Further research is warranted to disentangle these interpretations.

Conclusions

There is a long history of empirical research on the impact of scale formats (e.g., Schwarz, 1994). The present research adds to our knowledge by showing that social norms influence the respondents' use of the properties of a response scale in self-reports and others' reports. These findings have implications for questionnaire construction. Despite the fact that the proposed high and low frequency scales in the present research have poor ecological validity (in that they purposely distorted the range of alternatives from equidistance, by being highly positively or negatively skewed), and that the use of convenience samples further limited the external validity of the results (see Rockwood, Sangster, & Dillman, 1997; Sudman & Schwarz, 1989), our findings call attention to the role of implicit norms of behavior in the questions asked. Respondents' judgments always and necessarily take place in a normative context. When this context is not made clear in the questionnaire, respondents (sometimes involuntarily) turn their attention to the implicit context proposed by the researcher. Most of the time, this implicit context is likely to uncover the content of a (familiar) ingroup norm. As a consequence, the scale's frequency alternatives may be interpreted differently by respondents as a function of transient but salient group memberships during the survey. This would result in responses that are hardly comparable across different subgroups of respondents. The findings of the present research clearly underscore the need

to go beyond implicit norms in order to construct more adequate questionnaires.

Notes

- ¹ A series of MANOVAs showed no noticeable differences between the four characteristics in Studies 1–6.
- ² We controlled for order of presentation of temporal frame (yesterday vs. the last 30 days).
- ³ The effects did not vary as a function of order of temporal frame.
- ⁴ We controlled for order of presentation of target (self vs. people in general) and temporal frame (yesterday vs. the last 30 days).
- ⁵ See https://en.wikipedia.org/wiki/Body_mass_index.
- ⁶ This pattern of results did not differ according to the kind of group (Dusek or Tausig).

Acknowledgements

The authors would like to thank Nathalie Rodriguez for assistance with data collection in Study 5.

Competing Interests

The authors have no competing interests to declare.

References

- Adida, C. L., Ferree, K. E., Posner, D. N., & Robinson, A. L.** (2016). Who's asking? Interviewer coethnicity effects in African survey data. *Comparative Political Studies*, *49*(12), 1630–1660. DOI: <https://doi.org/10.1177/0010414016633487>
- Aron, A., Aron, E. N., & Smollan, D.** (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*, 596–612. DOI: <https://doi.org/10.1037/0022-3514.63.4.596>
- Biernat, M., & Manis, M.** (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, *66*(1), 5–20. DOI: <https://doi.org/10.1037/0022-3514.66.1.5>
- Bradburn, N. M., & Miles, C.** (1979). Vague quantifiers. *Public Opinion Quarterly*, *43*, 92–101. DOI: <https://doi.org/10.1086/268494>
- Bradburn, N. M., Sudman, S., & Wansink, B.** (2004). *Asking questions. The definitive guide to questionnaire design*. San Francisco: Jossey-Bass.
- Chao, M. M., Zhang, Z. X., & Chiu, C. Y.** (2009). Adherence to perceived norms across cultural boundaries: The role of need for cognitive closure and ingroup identification. *Group Processes and Intergroup Relations*, *13*, 69–89. DOI: <https://doi.org/10.1177/1368430209343115>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G.** (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160. DOI: <https://doi.org/10.3758/BRM.41.4.1149>
- Hogg, M. A.** (2000). Subjective uncertainty reduction through self-categorization: A motivational theory of social identity processes. *European Review of*

- Social Psychology*, 11, 223–255. DOI: <https://doi.org/10.1080/14792772043000040>
- Hogg, M. A., & Abrams, D.** (1998). *Social identifications*. New York: Routledge.
- Iacoviello, V., Berent, J., Frederic, N. S., & Pereira, A.** (2017). The impact of ingroup favoritism on self-esteem: A normative perspective. *Journal of Experimental Social Psychology*, 71, 31–41. DOI: <https://doi.org/10.1016/j.jesp.2016.12.013>
- Iacoviello, V., Lorenzi-Cioldi, F., & Chipeaux, M.** (2018). When ingroup status moderates the identification-similarity relationship: A social-identity perspective. *Self and Identity*, 18, 685–708. DOI: <https://doi.org/10.1080/15298868.2018.1513375>
- Kahneman, D.** (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Krosnick, J. A., & Presser, S.** (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research*, 2nd Edition (pp. 263–313). London: Emerald Group Publishing.
- Leary, M. R.** (2005). Sociometer theory and the pursuit of relational value: Getting to the root of self-esteem. *European Review of Social Psychology*, 16, 75–111. DOI: <https://doi.org/10.1080/10463280540000007>
- Lorenzi-Cioldi, F.** (1988). *Individus dominants et groupes dominés* [Dominant individuals and dominated groups]. Grenoble: Presses Universitaires.
- Lorenzi-Cioldi, F.** (1998). Group status and perceptions of homogeneity. *European Review of Social Psychology*, 9, 31–75. DOI: <https://doi.org/10.1080/14792779843000045>
- Miller, D. T., & Prentice, D. A.** (1996). The construction of social norms and standards. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 799–829). New York: Guilford Press.
- Mingay, D. J., Shevell, S. K., Bradburn, N. M., & Ramirez, C.** (1994). Self and proxy reports of everyday events. In N. Schwarz & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 235–250). New York: Springer-Verlag. DOI: https://doi.org/10.1007/978-1-4612-2624-6_16
- Nelson Laird, T. F., Korkmaz, A., & Chen, P.-S. D.** (2008). How often is “often” revisited: The meaning and linearity of vague quantifiers used on the National Survey of student engagement. *Paper presented at the Annual Meeting of the American Educational Research Association*, April 15, San Diego.
- Pohl, N. F.** (1981). Scale considerations in using vague quantifiers. *The Journal of Experimental Education*, 49, 235–240. DOI: <https://doi.org/10.1080/00220973.1981.11011790>
- Reese, G., Steffens, M. C., & Jonas, K. J.** (2013). When black sheep make us think: Information processing and devaluation of in- and outgroup norm deviants. *Social Cognition*, 31, 482–503. DOI: https://doi.org/10.1521/soco_2012_1005
- Rockwood, T. H., Sangster, R. L., & Dillman, D. A.** (1997). The effect of response categories on questionnaire answers. *Sociological Methods & Research*, 26, 118–140. DOI: <https://doi.org/10.1177/0049124197026001004>
- Schaeffer, N. C., & Presser, S.** (2003). The science of asking questions. *Annual Review of Sociology*, 29, 65–88. DOI: <https://doi.org/10.1146/annurev.soc.29.110702.110112>
- Schimmack, U.** (2002). Frequency judgements of emotions: The cognitive basis of personality assessment. In P. Sedlmeier & T. Betsch (Eds.), *Frequency processing and cognition* (pp. 189–204). New York: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780198508632.003.0012>
- Schuman, H., & Kalton, G.** (1985). Survey methods. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (pp. 635–697). New York: Random House.
- Schuman, H., & Presser, S.** (1981). *Questions and answers in attitude surveys. Experiments on question form, wording, and context*. New York: Academic Press.
- Schwarz, N.** (1990). Assessing frequency reports of mundane behaviors: Contributions of cognitive psychology to questionnaire construction. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 98–119). London: Sage.
- Schwarz, N.** (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. In: M. P. Zanna (Ed.), *Advances in Experimental Social Psychology*, 26, 123–162. New York: Academic Press. DOI: [https://doi.org/10.1016/S0065-2601\(08\)60153-7](https://doi.org/10.1016/S0065-2601(08)60153-7)
- Schwarz, N.** (1999). Self-reports. How the questions shape the answers. *American Psychologist*, 54, 93–105. DOI: <https://doi.org/10.1037/0003-066X.54.2.93>
- Schwarz, N., & Bienias, J.** (1990). What mediates the impact of response alternatives on frequency reports of mundane behaviors? *Applied Cognitive Psychology*, 4, 61–72. DOI: <https://doi.org/10.1002/acp.2350040106>
- Schwarz, N., Bless, H., Bohner, G., Harlacher, U., & Kellenbenz, M.** (1991). Response scales as frames of reference: The impact of frequency range on diagnostic judgement. *Applied Cognitive Psychology*, 5, 37–49. DOI: <https://doi.org/10.1002/acp.2350050104>
- Schwarz, N., Hippler, H. J., Deutsch, B., & Strack, F.** (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly*, 49, 388–395. DOI: <https://doi.org/10.1086/268936>
- Schwarz, N., Knäuper, B., Oyserman, D., & Stich, C.** (2008). The psychology of asking questions. In: D. de Leeuw, J. J. Hox & D. A. Dillman (Eds.), *International handbook of survey methodology* (pp. 18–34). New York: Lawrence Erlbaum.
- Schwarz, N., & Oyserman, D.** (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, 22, 127–160. DOI: <https://doi.org/10.1177/109821400102200202>
- Schwarz, N., Strack, F., Müller, G., & Chassein, B.** (1988). The range of response alternatives may

- determine the meaning of the question: Further evidence on informative functions of response alternatives. *Social Cognition*, 6, 107–117. DOI: <https://doi.org/10.1521/soco.1988.6.2.107>
- Sudman, S., Bradburn, N. M., & Schwarz, N.** (1996). *Thinking about answers. The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Sudman, S., & Schwarz, N.** (1989). Contributions of cognitive psychology to advertising research. *Journal of Advertising Research*, 29, 43–53.
- Tourangeau, R.** (2018). The survey response process from a cognitive viewpoint. *Quality Assurance in Education*, 26(2), 169–181. DOI: <https://doi.org/10.1108/QAE-06-2017-0034>
- Tourangeau, R., Rips, L. J., & Rasinski, K.** (Eds.) (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9780511819322>
- Tourangeau, R., & Smith, T. W.** (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60(2), 275–304. DOI: <https://doi.org/10.1086/297751>
- Turner, J. C.** (1991). *Social influence*. Buckingham: Open University Press.
- Turner, C. F., & Martin, E.** (1985). *Surveying subjective phenomena* (vol. 2). New York: Russell Sage.
- Visser, P. S., Krosnick, J. A., & Lavrakas, P. J.** (2000). Survey research. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 223–252). Cambridge University Press.
- Wright, D. B., Gaskell, G. D., & O'Muircheartaigh, C. A.** (1994). How much is “Quite a bit”? Mapping between numerical values and vague quantifiers. *Applied Cognitive Psychology*, 8, 479–496. DOI: <https://doi.org/10.1002/acp.2350080506>
- Wyer, R. S., & Srull, T. K.** (1986). Human cognition in its social context. *Psychological Review*, 93, 322–359. DOI: <https://doi.org/10.1037/0033-295X.93.3.322>

How to cite this article: Lorenzi-Cioldi, F., Iacoviello, V., & Kulich, C. (2021). The Real World is the Ingroup World: A Normative Explanation of Response-Scale Effects in the Subjective Report of Behaviors. *International Review of Social Psychology*, 34(1): 13, 1–13. DOI: <https://doi.org/10.5334/irsp.426>

Submitted: 30 March 2020

Accepted: 30 April 2021

Published: 02 July 2021

Copyright: © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



International Review of Social Psychology is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS