Article

Different Styles, Different Times: How Response Times Can Inform Our Knowledge About the Response Process in Rating Scale Measurement

Mirka Henninger[1] (iD) and Hansjörg Plieninger[1] (iD)

Abstract

When respondents use different ways to answer rating scale items, they employ so-called response styles that can bias inferences drawn from measurement. To describe the influence of such response styles on the response process, we investigated relations between extreme, acquiescent, and mid response style and response times in three studies using multilevel modeling. On the response level, agreement and midpoint, but not extreme responses were slower. On the person level, response times increased for extreme, but not for acquiescence or mid response style traits. For all three response styles, we found negative cross-level interaction effects, indicating that a response matching the response style trait is faster. The results demonstrate that response styles facilitate the choice of specific category combinations in terms of response speed across a wide range of response style trait levels.

Keywords

response process, response styles, response times, speed–distance hypothesis, data quality, cognitive effort

Rating scales are often used to measure latent variables such as beliefs, attitudes, or personality traits as they are convenient to apply and evaluate. However, the response to a rating scale item does not only reflect the trait to be measured but also the way a respondent perceives and uses the rating scale. The so-called response styles (Paulhus, 1991) can be regarded as latent traits that describe the respondents' tendencies to prefer certain types of categories over others irrespective of item content. For example, a bias towards choosing the highest and lowest categories is called extreme response style (ERS), a tendency to generally agree with the item is called acquiescence response style (ARS), and a preference toward the middle category is called mid response style (MRS; see van Vaerenbergh & Thomas, 2013, for a review and definitions of additional response styles).

Response styles seem to be ubiquitous in rating data (e.g., Böckenholt & Meiser, 2017; Eid & Rauber, 2000; Meiser & Machunsky, 2008; Wetzel et al., 2013). Moreover, response styles have been shown to be consistent across different content traits (Weijters et al., 2010b; Wetzel et al., 2013), and to be stable personality characteristics that persist over time (Weijters et al., 2010c; Wetzel et al., 2016). Thus, rating scales do not only capture information on the latent content trait but also on response styles. Such response styles can distort measurement precision (Bolt et al., 2014; Wetzel &

Carstensen, 2017), inflate relations between measured variables (Abad et al., 2018; Böckenholt & Meiser, 2017), or bias cross-group comparisons, for example, in cross-cultural research (Bolt et al., 2014; Rollock & Lui, 2016).

Attempts to explain response styles through demographic, personality, and situational variables yielded mixed results. The effects of gender and age on ERS are inconsistent across studies (e.g., Hamilton, 1968; Moors, 2008; van Vaerenbergh & Thomas, 2013; Weijters et al., 2010c), but intelligence, occupational status, and education seem to reduce ERS (e.g., Bolt & Johnson, 2009; Meisenberg & Williams, 2008). On the one hand, ERS increases with certain personality traits, such as intolerance of ambiguity, simplistic thinking, and decisiveness (Naemi et al., 2009), on the other hand, the relation of response styles and the Big Five have been found to be positive, negative, or nonexistent (e.g., Austin et al., 2006; Couch & Keniston, 1960; Grimm & Church, 1999; He & Van De Vijver, 2013; Hibbing et al., 2017; van Dijk et al., 2009; Wetzel & Carstensen, 2017). Situational variables, such as reducing the number of

---

[1]University of Mannheim, Mannheim, Germany

Corresponding Author:
Mirka Henninger, Department of Psychology, University of Mannheim, L13, 15, Mannheim 68161, Germany.
Email: m.henninger@uni-mannheim.de

response categories and inducing cognitive load increases the magnitude of ERS and ARS, respectively (Cabooter, 2010; Knowles & Condon, 1999; Weijters et al., 2010a), while at the same time alternative response formats have been shown not only to reduce but also increase response styles (Böckenholt, 2017, Plieninger et al., 2019).

The inconsistent results with respect to personality and situational variables demonstrate how little is still known about response styles as a psychological process. An analysis of response times may be a means to this end: Fekken and Holden (1994) argued that the time respondents take to provide a self-report response is a behavioral representation of the underlying cognitive process. They showed that response times are meaningful indicators for the trait to be measured on a personality test. Since responses are not only indicators of the trait to be measured but also of response styles, the time accompanying the responses should also be an indicator of processes related to content as well as response styles. Knowledge about the cognitive processes that influence response category selection through response styles will help us evaluate the often made claim that response styles are a result of reduced cognitive effort (e.g., Aichholzer, 2013; Krosnick, 1999), and to evaluate the magnitude of impact that response styles have on data quality.

Response Times in Rating Scale Measures

Response times have been used to assess cognitive processes in experimental psychology (e.g., Heck & Erdfelder, 2016) and served as collateral information in item response theory models for ability testing (e.g., van der Linden et al., 2010). However, there is little research investigating response times in personality measurement and even fewer assessing the relationship of response times and response styles.

Response Times in Personality Measurement

Response times have served as an indicator of respondents' motivation and deliberation in surveys. Fast responses have been associated with low motivation of the respondent (Callegaro et al., 2009), lower validity (Neubauer & Malle, 1997), and poor data quality (Zhang & Conrad, 2013). Furthermore, items that appear later in the survey are responded faster than earlier items and with a lower variability in the responses, which might be an indicator of decreasing motivation of respondents toward the end of the survey (Callegaro et al., 2009; Galesic & Bosnjak, 2009; Wise & DeMars, 2005, 2006; Yan & Tourangeau, 2008). Similarly, while shorter response times are associated with reports of desirable attitudes and behavior, longer response times have been linked to responses that are given more

carefully, such as faked responses or the reporting of undesirable attitudes (Andersen & Mayerl, 2017; Dunn et al., 1972; McIntyre, 2011; Neubauer & Malle, 1997; van Hooft & Born, 2012).

Another view on response times links fast responses to high confidence in the rating. Fast responses have been associated with the accessibility of the trait being measured, as respondents whose attitudes were important to them responded faster (Tourangeau et al., 1991). Similarly, fast response times are associated with a high consistency in item responses since respondents take less time to decide for a response option when they are certain about it (Arndt et al., 2017; Germeroth et al., 2015; McIntyre, 2011). In line with that, slow responses are considered to indicate cognitive effort in the response process. When respondents try to find the best answer to the item, response times increase, especially for complicated or ambiguous questions (Bassili & Scott, 1996; Dunn et al., 1972; Hanley, 1965; Rogers, 1973). Similarly, item complexity such as the number of clauses, characters, or cognitive operations required for a response increases response times (Kulas & Stachowski, 2009; Lenzner et al., 2010; Sauer et al., 2011; Yan & Tourangeau, 2008).

In sum, fast responses can have two interpretations: they may indicate a spontaneous response mode, in which respondents demonstrate low motivation and deliberation, but may also indicate confidence in the rating as the optimal response is highly accessible. Slower responses are the result of a careful, effortful or deliberate cognitive process, either due to thought-out decisions or item complexity.

Response Times in Response Style Research

In this research project, we examine the relation between extreme, acquiescent, and mid responding and response times to describe cognitive processes in rating scale usage. Herein, we differentiate between specific *responses* (e.g., extreme, agree, or mid responses) that are given faster or slower than other responses, and *respondents* (with different ERS, ARS, or MRS levels) that may respond faster or slower than other respondents across items.

It is important to separate the level of item responses and the level of respondents, because the effects may be different across levels. For example, it might be the case that persons high on ERS are generally faster, but that extreme responses across all respondents are given slower compared to non-extreme responses. Such phenomena are often referred to as *Simpson's paradox* or *ecological fallacy* (e.g., Kievit et al., 2013). For example, "the relation between coffee consumption and neuroticism is positive in each individual, but those individuals who drink more coffee are generally less neurotic" (Borsboom et al., 2009, p. 72). Thus, separating intra-individual and interindividual effects may provide new insights into the nature of response tendencies that have been so far been unobserved.

Effects of Current Responses on Response Times at the Response Level. Regarding the relation of item responses and response times, Casey and Tryon (2001) showed that a majority of participants gave faster responses in the extreme categories than in the neighboring non-extreme category. This result may suggest a negative main effect of extreme responses on response times, but the stability and magnitude of the effect remains unclear.

Hypothesis 1a: Extreme responses may result in shorter response times (although evidence for this effect is based on only one investigation by Casey and Tryon (2001).

Agree responses might be related to task complexity and a result of cognitive burden when items are hard to interpret. Agreement to both reversed and non-reversed items occurs with complex rather than easy items and results in higher cognitive demand and longer response times (Hanley, 1965; Rogers, 1973; Swain et al., 2008). In addition, Knowles and Condon (1999) showed in an experimental investigation that under high cognitive load, respondents tended to agree with the items more often. As cognitive load has been associated with longer response times, this effect further supports the hypothesis that agree responses lead to longer response latencies.

Hypothesis 1b: We expect that agree responses are given slower than non-agree responses since agree responses have been shown to result in longer response times (Hanley, 1965; Knowles & Condon, 1999; Rogers, 1973; Swain et al., 2008). Slower agree responses may indicate task complexity and increased cognitive demand.

Kulas and Stachowski (2009) found that respondents took longest to give a response in the middle category. The authors argued that it is cognitively less demanding to agree or disagree than to choose the midpoint. Especially when respondents cannot decide for a directed response, the choice of the indecisive midpoint may indicate a well evaluated, and therefore cognitively demanding judgment process that becomes visible through response times.

Hypothesis 1c: Mid responses may take longer than directed responses based on the evidence and considerations presented by Kulas and Stachowski (2009). Similar to the process underlying ARS, slower responses may indicate cognitive burden in evaluating the item, leading to a thought-out item response.

Effects of Response Style Traits on Response Times at the Respondent Level. In contrast to effects at the item response levels, there is little evidence pointing toward directed effects for response style trait levels on response times.

Therefore, analyses on the level of the respondent are exploratory. For ERS, there are no studies suggesting any positive or negative effect of ERS trait levels on response times. On the one hand, fast respondents showed higher variability in their responses than slow respondents (Neubauer & Malle, 1997). High variability in the responses is associated with high ERS levels, as the variance in the responses increases when extreme categories are chosen more often which may be indirect evidence that high ERS trait levels reduce response times. On the other hand, Naemi et al. (2009) found no main effect of ERS levels on response times.

Exploratory Analysis 2a: We will explore the effects of ERS trait levels on response times. As the effects reported by Neubauer and Malle (1997) are indirect, and no effects were found by Naemi et al. (2009), no prediction can be made on whether high ERS levels should lead to shorter, faster, or unchanged response times.

For ARS, Mayerl (2013) argued that measured attitudes are stronger influenced by acquiescence when respondents answered in a fast, automatic-spontaneous response mode. The descriptive response times by Knowles and Condon (1999) also indicate lower response times for respondents with high ARS levels than for respondents with low ARS levels.

Exploratory Analysis 2b: We will explore the effects of ARS trait levels on response times. First results (Knowles & Condon, 1999; Mayler, 2013) point toward a decrease in response times for higher ARS trait levels, but evidence is sparse.

To our knowledge, there is no literature to build on in order to predict effects of MRS levels on response times.

Exploratory Analysis 2c: We will explore the effects of response style trait levels for MRS on response times.

Interaction Effects Between the Current Response and Response Style Traits on Response Times. Besides main effects of item responses and respondents' response style traits, interaction effects may occur such that respondents with higher response style traits are faster when they give responses matching their response style trait. For example, a respondent with high ERS trait levels may be faster when giving an extreme response, and slower when giving a non-extreme response.

In terms of ERS, Naemi et al. (2009) showed that the combination of ERS and specific personality traits jointly decrease response times. This pattern speaks in favor of a more complex relation between ERS and response times. In terms of ARS, Knowles and Condon (1999) found an

interaction effect in such a way that respondents with high levels of ARS were faster when they agreed than when they disagreed with an item, and faster when they agreed than non-ARS respondents.

For MRS, there is no literature directly pointing toward an interaction effect for MRS and response times. However, response time for choices of the mid response option may be longer for respondents that have weighed the pros and cons of either side of the item, but that do not have a general tendency to prefer the middle category over the other response options. In contrast, respondents with a high MRS trait, using the mid response option abundantly may have faster response times when giving a mid response than respondents with low MRS trait levels (see Kulas & Stachowski, 2009).

Speed–Distance Hypothesis. An important theory that further supports the idea of an interaction effect between response style traits and item responses on response times is the speed–distance hypothesis. It predicts that response times decrease with increasing distance between the trait level of the respondent and item difficulty (Akrami et al., 2007; McIntyre, 2011). Larger distances result in a higher confidence to give a clear-cut response, while smaller distances imply high uncertainty about the item response (see also Ferrando & Lorenzo-Seva, 2007; Ranger & Ortner, 2011, for two item response theory models based on the speed–distance relationship).

Evidence for the speed–distance hypothesis is abundant. For example, Fekken and Holden (1992) show that response times for respondents with high trait levels that agree with the item respond fast, while respondents with high trait levels that disagree with the item respond slowly. Similarly, Casey and Tryon (2001) and McIntyre (2011) argued that pronounced self-schemata guide responses and decrease response times. In contrast, respondents with low trait knowledge or respondents that answer contrary to their self-schemata give slow responses (see also Dunn et al., 1972; Germeroth et al., 2015; Kuiper, 1981). The complex relationship between the trait level, the given response and response times even holds for peer ratings. Fuhrman and Funder (1995) found that high self-ratings were predictive of higher as well as quicker peer ratings; peer ratings were slower when the trait was rated high, but the current item was disagreed with. In short, the speed–distance hypothesis assumes that the more likely a response, the faster it will be given.

Based on the speed–distance hypothesis, we predict that the closer the observed response matches the response style trait, the faster the response will be. For example, a person with high ERS levels will take little time to give an extreme response. In contrast, when deviating from their ERS trait by giving a non-extreme response, the respondent will take more time. This reasoning is also in line with the evidence that, under high confidence, responses are given faster (Germeroth et al., 2015; McIntyre, 2011; Tourangeau et al., 1991) while responses involving high cognitive effort are given slower (e.g., Kulas & Stachowski, 2009; Lenzner et al., 2010; Sauer et al., 2011; Yan & Tourangeau, 2008).

> **Hypothesis 3a-c:** We predict for ERS, ARS, and MRS that responses that are in line with the response style traits will be given faster, whereas responses that are opposite to the response style trait will be accompanied by longer response times.

Method

Collecting Response Time Data

In collaboration with three research groups (Fladerer et al., 2019; Pfister, 2018; Plieninger et al., 2019), we recorded response times for each response in three studies. The first study was conducted in the context of a Bachelor thesis (Pfister, 2018) in which we initially investigated the relation between implicit personality measures and response styles in rating scale items with five categories. Large parts of the sample consist of undergraduate university students from social science programs. The second study originated from a collaboration with Plieninger et al. (2019), wherein the authors compared different response formats using six response categories; we collected response times in the Likert condition. The sample was recruited through the nonrepresentative and noncommercial online panel SoSci Survey. The third study consists of responses to 5- and 7-point rating scales on Leadership and Team Collaboration and was conducted in collaboration with Fladerer et al. (2019). In this study, data were collected from employees of dealerships and sales branches in Germany. All three studies were conducted online in one wave and rating scale items were presented on several pages with no more than 14 items presented on a single webpage. Table 1 provides information on the sample characteristics, while Table 2 summarizes the number of items, response options, and scales used in the three studies. The online supplementary material furthermore provides information on item characteristics, such as number of words, syllables, whether the item was negated or reversed, and the items' content type.

When planning the three studies, we aimed at validly measuring response styles by using heterogeneous items without a common trait (see De Beuckelaer et al., 2010; Greenleaf, 1992), while at the same time making the study conditions as close to real measurement situations as possible. Thus, we designed three studies accordingly: Study 1 focused on measurement of response styles, therefore only heterogeneous items (i.e., items without a common trait) were selected from various scales (De Beuckelaer et al.,

Table I. Sample Characteristics.

	N	Gender	Age groups in years				
			15-25	26-35	36-45	45-65	66+
Study 1	161	80% female	77%	17%	0%	3%	2%
Study 2	154	53% female	20%	34%	15%	16%	15%
Study 3	786	30% female	22%	29%	21%	28%	0%

Note. N = number of participants after exclusions.

Table 2. Overview of the Data Used for Analyses.

	I	K	Scales	Cronbach's α
Study 1	39	5	*Heterogeneous* (no common trait; 39 items)	—
Study 2	54	6	Honesty–Humility (10 items)	.79
		6	Personal Need for Structure (12 items)	.65
		6	*Heterogeneous* (no common trait; 32 items)	—
Study 3	45	5	Identity Leadership Inventory (14 items)	.96
		5	Social Identification (6 items)	.82
		7	Perceived Organizational Support (8 items)	.87
		7	Collective Self-Esteem (7 items)	.84
		7	Resilience (10 items)	.83

Note. I = number of items; K = number of response categories. Validated scales originated from: Honesty–Humility scale (Lee & Ashton, 2006); Personal Need for Structure scale (Machunsky & Meiser, 2006); Identity Leadership Inventory (Steffens et al., 2014); Social Identification scale (Mael & Asiforth, 1992); Perceived Organizational Support scale (Eisenberger et al., 1986); Collective Self-Esteem scale (Riggs et al., 1994); Resilience CD-RISC scale (Sarubin et al., 2015). Cronbach's alpha is computed in the final sample used in these analyses.

2010; Greenleaf, 1992). Study 2 served as an intermediate step employing heterogeneous items as well as two content trait scales. Study 3 used items of five different, validated scales from organizational psychology to ensure that the results obtained from Study 1 and Study 2 can be generalized to applied measurement situations.

We used JavaScript to track the response as well as the time in milliseconds associated with each mouse click. The response times for a given item was then operationalized as the time difference between the current and the preceding mouse click. For future research or applications, we made the JavaScript code to collect response times available on OSF.[1]

Data Preprocessing

Since data were collected online in the three studies, careful attention was paid to retain only valid data. The first two studies contained several validity checks to ensure data quality (e.g., items wherein participants could indicate that, for example, they have been distracted during the study and one Bogus item, see also Meade & Craig, 2012). Based on the validity checks, we excluded 26 and 44 participants in Studies 1 and 2, respectively. In the third study, respondents who answered to at least 30 out of 45 items were included in the analyses.

When participants were directed back to the preceding page because they omitted one or more items, we decided to exclude responses to the initially omitted items from analyses since they may be imprecise with respect to response times when respondents have to reorientate themselves on the survey page (see also Höhne & Schlosser, 2018).[2] As we collected response times for each mouse click, we evaluated whether respondents answered to items more than once. Across all items and respondents, 9% of responses were changed in Study 1, 8% in Study 2, and 6% in Study 3. Assuming that a spontaneous response was the best indicator of the underlying response process, we used the initial response to an item in cases where participants later modified their response.

Response Times. Based on a Box–Cox transformation test, we log transformed response times to obtain a normal distribution of the response time variable. Assuming that very slow and very fast responses may not be the result of a valid response process, we excluded responses that deviated ±2 standard deviation from the individual respondent's mean response time (a common approach in response time analyses, see Bassili & Fletcher, 1991; Mayerl & Urban, 2008; Mulligan et al., 2003). Through this procedure, 302 out of 6,268 responses (4.8%) in Study; 1,358 out of 8,279 responses (4.3%) in Study 2; and 1,739 out of 34,854 responses (5.0%)

Table 3. Descriptive Sample Statistics of Log Response Times in the Three Studies.

	Min.	1st Quart.	Median	Mean	3rd Quart.	Max.
Study 1	−1.02	1.25	1.55	1.58	1.89	4.48
Study 2	−0.16	1.45	1.78	1.81	2.14	4.21
Study 3	−1.33	1.32	1.72	1.75	2.15	5.60

Table 4. Recoding of Item Responses Into Dichotomous Response Style Indicators for Different Number of Response Categories.

Number of categories	Response type	Initial response						
		−3	−2	−1	0	1	2	3
5	$X_{in}^{Extreme}$	—	1	0	0	0	1	—
	X_{in}^{Agree}	—	0	0	0	1	1	—
	X_{in}^{Mid}	—	0	0	1	0	0	—
6	$X_{in}^{Extreme}$	1	0	0	—	0	0	1
	X_{in}^{Agree}	0	0	0	—	1	1	1
	X_{in}^{Mid}	—	—	—	—	—	—	—
7	$X_{in}^{Extreme}$	1	0	0	0	0	0	1
	X_{in}^{Agree}	0	0	0	0	1	1	1
	X_{in}^{Mid}	0	0	0	1	0	0	0

Note. X_{in} = response to item i by person n for extreme, agree, and mid responses; no midpoint response was modeled for scales with an even number of response categories.

in Study 3 were excluded which led to an approximately normal distribution of log response times on the sample level. Table 3 shows the descriptive sample statistics of log response times in the three data sets.

Response Style Indicators. We recoded item responses to obtain dichotomous response style indicators (see De Beuckelaer et al., 2010; Greenleaf, 1992; Wetzel & Carstensen, 2017). For extreme responses, a response was coded 1 if it was in either one of the two extreme categories and 0 otherwise. For agreement responses, responses in the agreement categories (i.e., categories above the midpoint) were coded 1 and 0 otherwise. A response was coded a midpoint response with value 1, if the midpoint was chosen and 0 otherwise; midpoint responses were not defined in case of a scale with an even number of categories. Table 4 gives an overview of the scoring rules for scales with different numbers of response categories.

Multilevel Modeling Approach

We used a multilevel modeling approach to predict individual log response times based on responses of respondent n to item i using item responses (Level 1), respondents' response styles (Level 2) and their cross-level interaction as predictor variables.

On Level 1 (response level), we used three dichotomous variables ($X_{in}^{Extreme}$, X_{in}^{Agree}, X_{in}^{Mid}) that indicated whether a

given response was an extreme, agreement, or midpoint response, respectively (see Table 4). Hence, Level 1 variables described whether the current item response was indicative of a specific response style. In addition, we entered effect-coded item fixed effects $\sum_{i=2}^{I} \beta_i X_i^{item}$ using X_1^{item} as a reference to control for differences in response times due to item features, such as item length or complexity.[3] Thus, the Level 1 model equation is given by:

$$log\, ResponseTimes_{in} = \sum_{i=2}^{I} \beta_i X_i^{item} + \beta_{0n} + \beta_{1n} X_{in}^{Extreme} + \beta_{2n} X_{in}^{Agree} + \beta_{3n} X_{in}^{Mid} + e_{in}$$

Level 2 (respondent level) variables were trait scores of response styles ERS, ARS, and MRS $\left(\theta_n^{ERS}, \theta_n^{ARS}, \theta_n^{MRS}\right)$ for each respondent. The trait scores reflected interindividual differences in response styles. Rather than using, for example, manifest sum scores which may lead to biased regression estimates (see Lüdtke et al., 2008), we used a latent aggregation procedure. It takes sampling error into account when Level 1 variables $\left(X_{in}^{Extreme}, X_{in}^{Agree}, X_{in}^{Mid}\right)$ are combined to form Level 2 variables $\left(\theta_n^{ERS}, \theta_n^{ARS}, \theta_n^{MRS}\right)$. Therewith, we account for unreliability in Level 2 predictors and can correct for biases in between-group regression coefficients (see also Lüdtke et al., 2008; Lüdtke et al., 2011; Marsh et al., 2009).

On Level 2, we specified a random intercept for respondents to account for differences in response times between

respondents. We defined the parameters of the random intercept as a function of the latent response style traits $\left(\theta_n^{ERS}, \theta_n^{ARS}, \theta_n^{MRS}\right)$. The model equation for the intercept is given by:

$$\beta_{0n} = \gamma_{00} + \gamma_{01}\theta_n^{ERS} + \gamma_{02}\theta_n^{ARS} + \gamma_{03}\theta_n^{MRS} + u_{0n}$$

Besides, we defined varying slope parameters $\beta_{1n}, \beta_{2n}, \beta_{3n}$ for each response type $\left(X_{in}^{Extreme}, X_{in}^{Agree}, X_{in}^{Mid}\right)$ as a function of the respective latent response style trait $\left(\theta_n^{ERS}, \theta_n^{ARS}, \theta_n^{MRS}\right)$ to study the effects of response styles on response times through cross-level interactions. The model equation for the slope parameters is given by:

$$\beta_{1n} = \gamma_{10} + \gamma_{11}\theta_n^{ERS} + u_{1n}$$

$$\beta_{2n} = \gamma_{20} + \gamma_{21}\theta_n^{ARS} + u_{2n}$$

$$\beta_{3n} = \gamma_{30} + \gamma_{31}\theta_n^{MRS} + u_{3n}$$

The resulting joint model equation is thus given by:

$$log\ Response\ Times_{in} = \sum_{i=2}^{I} \beta_i X_i^{item} +$$
$$\gamma_{00} + \gamma_{01}\theta_n^{ERS} + \gamma_{02}\theta_n^{ARS} + \gamma_{03}\theta_n^{MRS} +$$
$$\gamma_{10}X_{in}^{Extreme} + \gamma_{11}\theta_n^{ERS}X_{in}^{Extreme} +$$
$$\gamma_{20}X_{in}^{Agree} + \gamma_{21}\theta_n^{ARS}X_{in}^{Agree} +$$
$$\gamma_{30}X_{in}^{Mid} + \gamma_{31}\theta_n^{MRS}X_{in}^{Mid} +$$
$$u_{0n} + u_{1n}X_{in}^{Extreme} + u_{2n}X_{in}^{Agree} +$$
$$u_{3n}X_{in}^{Mid} + e_{in}$$

In summary, the model captures differences in response times due to simple interindividual differences (via β_{0n}) and differences due to item characteristics (via β_i). Thus, further effects can be interpreted as the changes of the respondent's response time to an average item from his or her average response time. Main effects on Level 1 $\left(\gamma_{10}, \gamma_{20}, \gamma_{30}\right)$ indicate whether specific responses (e.g., $X_{in}^{Extreme}$) take longer, main effects on Level 2 $\left(\gamma_{01}, \gamma_{02}, \gamma_{03}\right)$ indicate whether specific respondents (e.g., with high θ_n^{ERS}) take longer and cross-level interaction effects $\left(\gamma_{11}, \gamma_{21}, \gamma_{31}\right)$ indicate whether specific responses (e.g., $X_{in}^{Extreme}$) take longer for certain levels of latent response style traits (e.g., for high θ_n^{ERS}).

All analyses were conducted using R (R Core Team, 2019) with M*plus* Automation (Hallquist & Wiley, 2018) using M*plus* version 7.4 (Muthén & Muthén, 2012) for model fit.[4] M*plus* code for model fit is provided on OSF. We set the level of significance to $\alpha = .05$.

Results

Figure 1 and Table 5 provide the estimates of the multilevel analysis for the three data sets. Since response times were log-transformed, the exponential of the estimate (x) is interpreted as a proportional change ($x*100\%$) in the dependent variable (see, e.g., Lo & Andrews, 2015, p. 12).

On Level 1, agree and mid responses significantly increased response times (Hypotheses 1b and 1c), while there is a null effect for extreme responses (in contrast to Hypothesis 1a). Substantively, giving an agree response X_{in}^{Agree} increased response times compared with the respondent's average response time by 28% in Study 1, by 54% in Study 2, and by 21% in Study 3. Similarly, giving a mid response X_{in}^{Mid} increased the average response time of the respondent by 20% in Study 1, and by 19% in Study 3 compared with a directed response.

On Level 2, there was a significant positive main effect of respondents' ERS levels θ_n^{ERS} on response times in all three data sets (Exploratory Analysis 2a). When ERS levels increased by, for example, 0.3 response times increased by 22% in Study 1, by 15% in Study 2, and by 8% in Study 3. There were no significant Level 2 main effects for ARS and MRS, so higher levels of acquiescence or MRS did neither increase nor decrease response times (Exploratory Analysis 2b-c).

In all three studies, there was a negative cross-level interaction effect between the type of item responses $\left(X_{in}^{Extreme}, X_{in}^{Agree}, X_{in}^{Mid}\right)$ and respondents' response styles ($\theta_n^{ERS}, \theta_n^{ARS}, \theta_n^{MRS}$; Hypotheses 3a-c). High levels of response styles in combination with a response that matches the response styles significantly accelerated the response time of the respondent. So, when ERS levels increased by 0.3, and an extreme response was given, respondents were 7%, 11%, or 12% faster in Studies 1, 2, and 3, respectively. In case of ARS, an increase of 0.3 in ARS levels jointly with an agree response decreased respondents' response time by 16%, 23%, or 12% in the three data sets. For MRS, a mid response in combination with an increase in MRS levels by 0.3 decreased response times by 15% in Study 1 and 25% in Study 3. The interpretation of these cross-level interactions will be further illuminated in the following paragraph (see also Figure A1 in Appendix A).

Interpreting Interaction Effects With the Johnson–Neyman Technique

The upper rows of Figures 2, 3, and 4 show raw data scatterplots of response times in seconds (minimum inner 80% quantile) and model-based prediction lines as a function of the latent response style aggregate for extreme, agree, and midpoint responding, respectively. Please note that prediction lines are slightly bent due to reconversion of log response times (that are the basis of the linear model) into response times in seconds. In the lower row, Johnson–Neyman plots
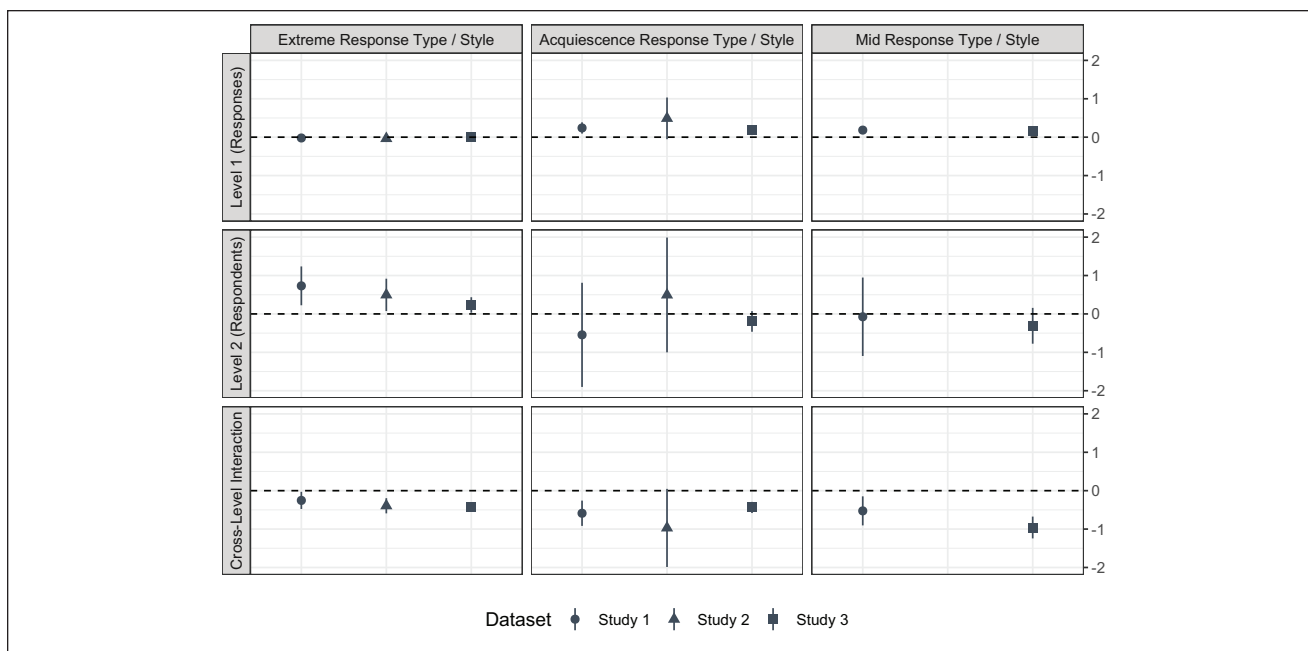
Figure 1. Fixed effects estimates of the multilevel analysis (error bars reflect 95% confidence intervals).

Table 5. Summary of Multilevel Model Estimates Predicting Log Response Times.

Predictors	Study 1			Study 2			Study 3		
	B	SE	p	B	SE	p	B	SE	p
Level 1 (Responses)									
Intercept (γ_{00})	1.61	0.34	<.001	1.38	0.37	<.001	1.81	0.11	<.001
ERS (γ_{10})	−0.03	0.03	.432	−0.03	0.03	.181	<−0.01	0.02	.803
ARS (γ_{20})	0.24	0.07	<.001	0.43	0.16	.008	0.19	0.04	<.001
MRS (γ_{30})	0.18	0.05	<.001	—	—	—	0.17	0.03	<.001
Level 2 (Respondents)									
ERS (γ_{01})	0.67	0.24	.005	0.47	0.21	.024	0.26	0.10	.012
ARS (γ_{02})	−0.58	0.64	.363	0.30	0.59	.607	−0.21	0.13	.100
MRS (γ_{03})	−0.16	0.47	.742	—	—	—	−0.25	0.24	.290
Cross-level interaction									
ERS (γ_{11})	−0.24	0.11	.030	−0.40	0.10	<.001	−0.44	0.05	<.001
ARS (γ_{12})	−0.60	0.15	<.001	−0.87	0.31	.005	−0.43	0.07	<.001
MRS (γ_{13})	−0.52	0.19	.007	—	—	—	−0.97	0.14	<.001
Variance components									
Intercept (u_{on})	0.09	0.01	<.001	0.08	0.01	<.001	0.10	0.01	<.001
ERS slope (u_{In})	<0.01	<0.01	.771	<0.01	<0.01	.914	0.01	<0.01	.001
ARS slope (u_{2n})	<0.01	<0.01	.753	<0.01	<0.01	.251	0.01	0.01	.015
MRS slope (u_{3n})	<0.01	<0.01	.659	—	—	—	0.02	0.01	.004
Residual (e_{in})	0.11	0.01	<.001	0.12	0.01	<.001	0.22	0.01	<.001

Note. SE = standard error; ERS = extreme response style; ARS = acquiescence response style; MRS = mid response style. All significance tests are two-sided.

illustrate the change in the effect of an item response on response times as a function of the latent response style aggregate (see Bauer & Curran, 2005; Preacher et al., 2006, for details on this technique in multilevel models). For example, the Johnson–Neyman technique displays how the effect of giving an extreme response $\left(X_{in}^{Extreme} \right)$ on response times
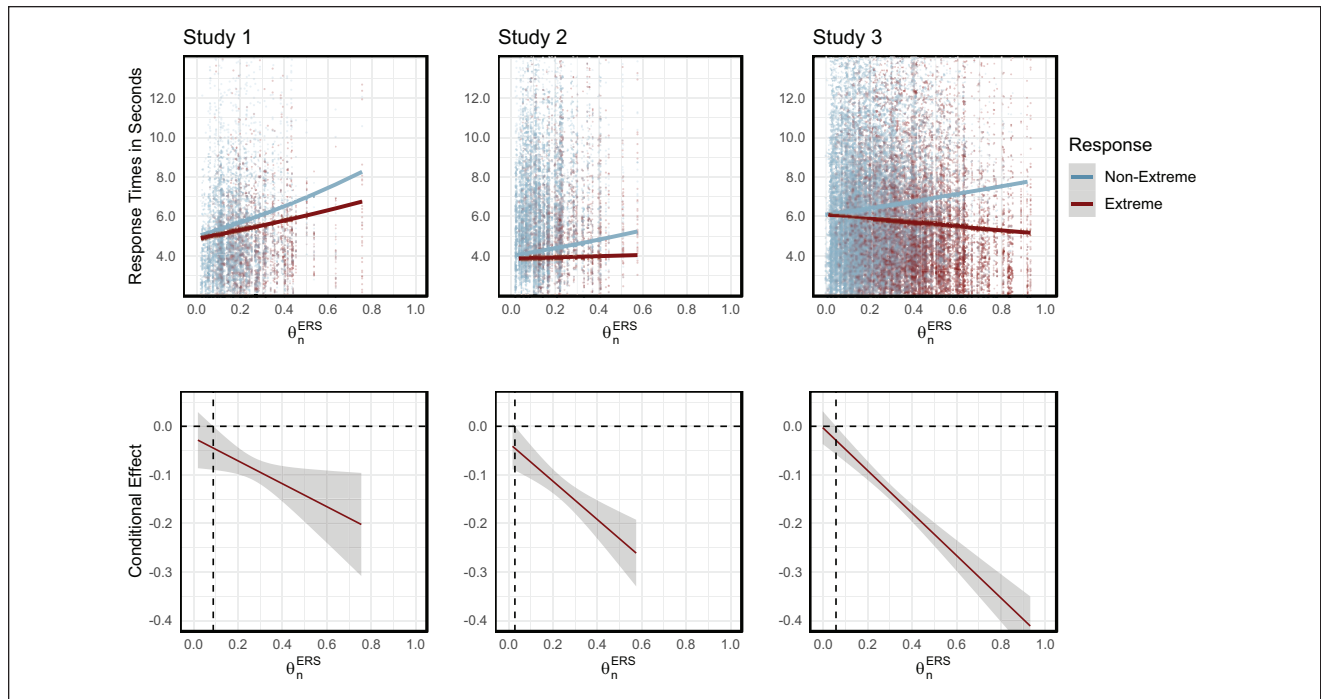
Figure 2. Scatterplots with model-based prediction lines (upper panel) and Johnson–Neyman plots (lower panel) to illustrate the effect of extreme response style (ERS) levels θ_n^{ERS} and an extreme response on response times.
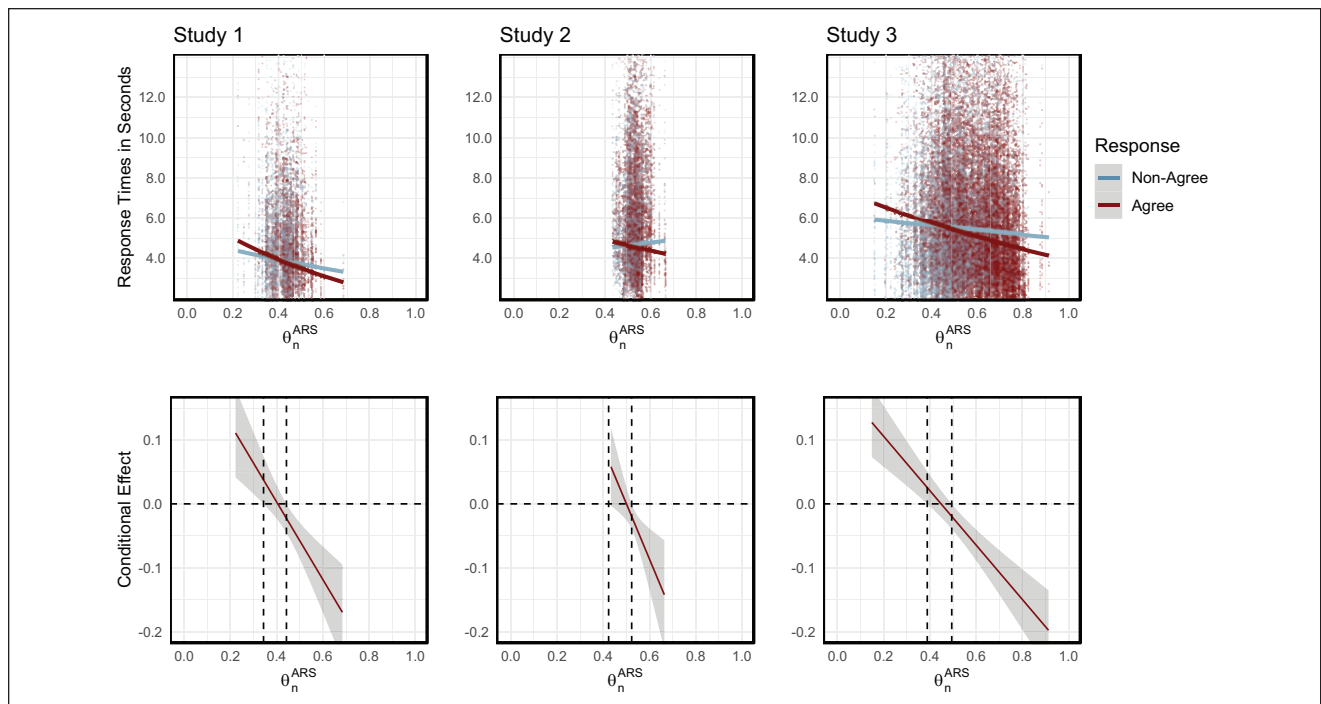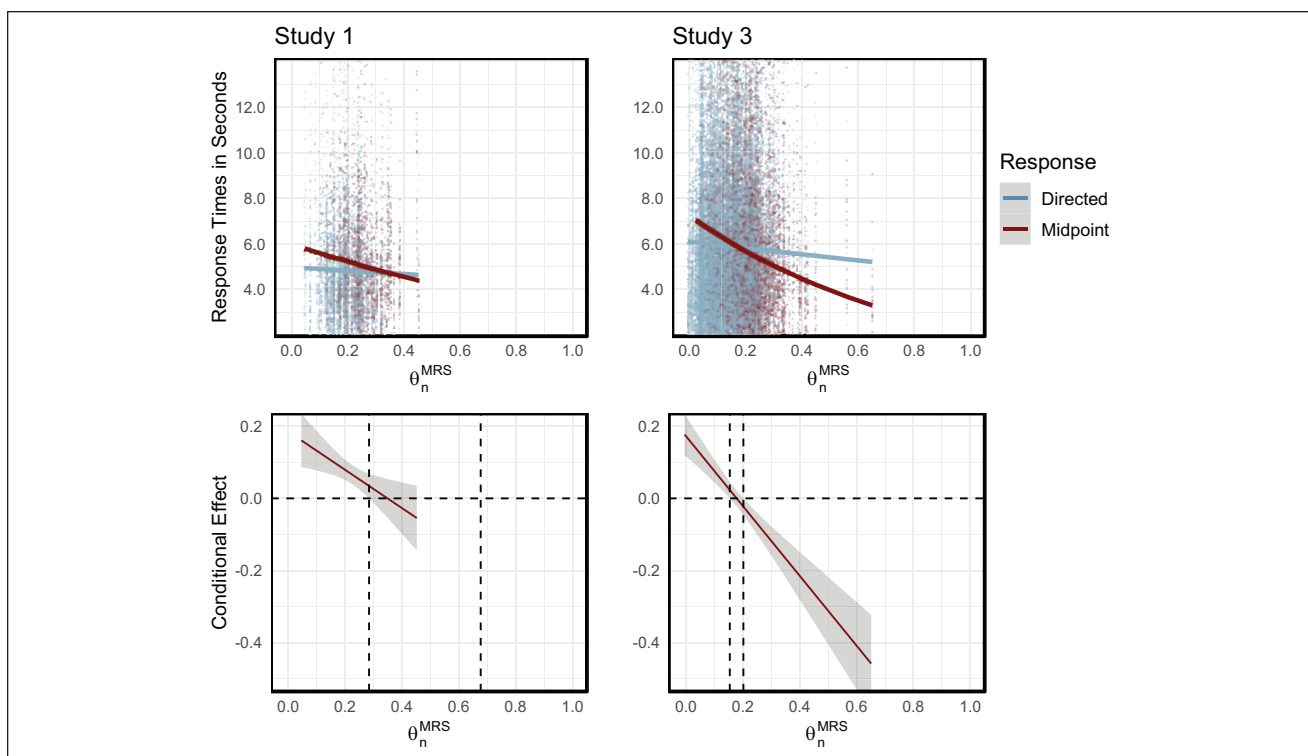

Figure 3. Scatterplots with model-based prediction lines (upper panel) and Johnson–Neyman plots (lower panel) to illustrate the effect of acquiescence response style (ARS) levels θ_n^{ARS} and an agree response on response times.

(y-axis) changes for different levels of the latent ERS aggregate (θ_n^{ERS}; x-axis) and identifies regions of significance,

hence regions where the effect is significantly positive, significantly negative, or not significantly different from zero.

Figure 4. Scatterplots with model-based prediction lines (upper panel) and Johnson–Neyman plots (lower panel) to illustrate the effect of mid response style (MRS) levels θ_n^{MRS} and a mid response on response times (no MRS effect was modeled in Study 2 due to the use of a rating scale with an even number of categories).

Confidence bands represent the uncertainty in the conditional effect and dashed, vertical lines represent the boundaries of the regions of significance.

Extreme Response Style. In the upper row of Figure 2, the positive Level 2 main effect of θ_n^{ERS} is apparent when averaging over extreme and non-extreme responses. The cross-level interaction leads to the fact that the lines for extreme and non-extreme responses are not parallel. This cross-level interaction is further illustrated in the lower row using the Johnson–Neyman technique. These plots show the effect of giving an extreme response on response time as a function of the latent ERS estimate on the *x*-axis. These plots indicate that the higher the ERS trait level $\left(\theta_n^{ERS}\right)$, the stronger was the negative effect of extreme compared with non-extreme responses on response times. This conditional effect was significantly negative for $\theta_n^{ERS} > .09$ across data sets as illustrated by the dashed line marking the boundary of the region of significance. Very low levels of ERS do not impact the effect of an extreme response on response time. Stated differently, responses were slowest when respondents with high ERS levels selected a non-extreme response category, which seems to be a more carefully considered category choice the higher the ERS level.

Acquiescence Response Style. Figure 3 illustrates the interaction effect for ARS which followed a disordinal pattern. Hence, for low ARS levels giving an agree response increased response times, while for high ARS levels giving an agree response decreased response times. Across studies, the conditional effect was significantly positive for $\theta_n^{ARS} < .34$ and significantly negative for $\theta_n^{ARS} > .52$. Hence, responses were faster when respondents with low ARS selected a non-agree response category and respondents with high ARS levels selected an agree response category.

Mid Response Style. Figure 4 shows the interaction effect for MRS in the studies with an odd number of response categories. In Study 1, the effect of MRS responses on the effect of MRS latent aggregate on response time was significantly positive for low levels of MRS $\left(\theta_n^{MRS} < .29\right)$, where response times increased when a mid response was given. The upper boundary was $\theta_n^{MRS} > .69$, implying that for MRS levels above this boundary response times decreased when a mid response was given even though there was no data available for this range of MRS in Study 1. We see a pronounced disordinal interaction in Study 3 indicating that giving a midpoint response significantly increased response times for low MRS levels $\left(\theta_n^{MRS} < .15\right)$, while it decreased response times for MRS levels higher than $\theta_n^{MRS} > .20$. We

can conclude that lower MRS levels lead to higher response times when a midpoint response was given, while a mid response for higher MRS trait levels results in shorter response times.

Discussion

In this research project, we investigated the effects of extreme, agree, and mid responding on response times. Although response times are frequently used to describe cognitive processes, they have rarely been linked to personality traits or response tendencies. However, response times can provide useful insights into the cognitive processes underlying rating scale usage and the use of response styles. In three different studies, we employed a multilevel modeling approach that allowed us to separate the effect of item responses from respondents' response styles as intraindividual and interindividual effects may differ from each other. Furthermore, the multilevel model approach allowed us to study the cross-level interaction effect of response types and response styles on response times.

On the level of individual item responses, we investigated the effect of dichotomous indicators of extreme, agree, and mid responses on response times and found consistent main effects across the three studies. In contrast to Hypothesis 1a, there was no difference in response times between extreme and non-extreme responses in any of the data sets which contradicts the results by Casey and Tryon (2001). In accordance with Hypothesis 1b, response times increased when agree responses were given. This is in line with evidence presented by Swain et al. (2008), Hanley (1965), and Rogers (1973) indicating that agree responses might be related to cognitive burden. Similarly, response times increased when a midpoint response was given which is in line with Hypothesis 1c. Hence, choosing the midpoint seems to be a deliberate process where respondent weigh the different alternatives, and choose the midpoint as a final response. The results corroborate findings by Kulas and Stachowski (2009) indicating that the midpoint was the response option with the longest response latency.

On the level of the respondent, we explored the influence of response style traits ERS, ARS, and MRS on response times (Exploratory Analyses 2a-c). In all three data sets, we found a positive main effect of the ERS trait on response times. Thus, the higher the ERS trait, the more time does the respondent take to respond. Particularly when responses are non-extreme, respondents with high ERS levels seem to take more time to respond. No main effects were found for ARS in any of the three data sets contradicting the results by Mayerl (2013) and the descriptive results by Knowles and Condon (1999). Neither did we find a main effect for MRS on the respondent level.

The multilevel analysis used here yielded original evidence for cross-level interactions, hence matching effects of response styles and item responses. As predicted in Hypotheses 3a-c, there were significant negative cross-level interaction effects of item responses and response style traits on response times across all data sets and across all response styles. Thus, giving a response that is in line with the response style trait decreases response times or, stated differently, the response style trait facilitates the choice of certain categories in terms of response speed. The illustration with the Johnson–Neyman technique (Figures 2, 3, and 4) also brought novel insights into the range of response style trait levels $\left(\theta_n^{ERS}, \theta_n^{ARS}, \theta_n^{MRS}\right)$ for which category choice was affected or unaffected. Please note that when respondents' latent response style trait lies in the area over which the cross-level interaction effect is not significantly different from zero (area within the boundaries of the region of significance), response times are equal for both response options (e.g., an extreme or a non-extreme response). This area might therefore demarcate the range over which response styles have the smallest impact on the response. In our analyses, these "neutral" response style levels were identified to be very low for ERS across data sets (ERS: $\theta_n^{ERS} < .09$), indicated by an ordinal interaction effect. In contrast, for ARS and MRS moderate response style trait levels were identified as "neutral" ($ARS : .34 < \theta_n^{ARS} < .52$; MRS: $.15 < \theta_n^{MRS} < .20$; MRS in Study 3), as indicated by a disordinal interaction effect. Across all three response styles, the range of these "neutral" levels was very small. Therefore, a preference (or avoidance) for certain response category types in terms of response speed is consequential for a majority of respondents and there exists almost no level of response style for which the category choice is not facilitated by response tendencies (see also Figure A1 in Appendix A, for an illustration of the frequency of different effects of response types on response times in the data sets). These differences in response speed reflect the subjective conviction of respondents with a certain response style about selecting specific response options (Gross et al., 1995).

Theoretical Implications

Cognitive Processes Underlying Response Style Usage. The analyses and results of the current investigation show that extreme responding is qualitatively distinct from acquiescent and mid responding and follows a different cognitive process. Based on the visualization of the Johnson–Neyman technique, respondents with moderate and high ERS trait levels take longer to give non-extreme responses (see Figure 2). Furthermore, only at very low ERS trait levels, extreme and non-extreme responses have similar response times. Since overall high ERS trait levels are accompanied by longer response times and, the negative cross-level interaction indicates that respondents with moderate to high ERS levels give non-extreme responses more deliberately,

the results do not support the notion that ERS is associated with low cognitive effort of the respondent.

In contrast, acquiescent and mid responding show very similar patterns of response processes. First, on Level 1, we found positive main effects of agree and mid responses, indicating these responses go along with longer response times. Second, there were no main effects of the ARS and MRS traits, indicating that across responses, differences in respondents' ARS and MRS levels did not explain differences in response times. Third, disordinal interactions were found for acquiescent and mid responding indicating that responses that are in line with the respective response style are faster than responses that contradict the response style.

Knowles and Condon (1999) reported that response times were faster when ARS-respondents agreed with the item and based their argumentation on a dual-process theory of acquiescence. According to this theory, people either agree with the item instantly without investing any effort, or follow a normal processing route including comprehension, reconsideration, and decision phases that require more time and effort. We were able to replicate the finding by Knowles and Condon for high ARS levels. At the same time, our data showed a similar pattern for respondents with low ARS levels who were faster when they disagreed (see Figure 3 and Figure A1 in Appendix A). This is a clear contradiction to a dual-process theory with a unipolar conceptualization of acquiescence where the absence of acquiescence means moderate responding (Knowles & Condon, 1999; see also Plieninger & Heck, 2018). The results rather suggest a bipolar acquiescence construct where respondents with low levels of acquiescence tend to disagree with the item more easily, while respondents with high levels of acquiescence tend to agree with items independent of item content. The same process seems to hold for MRS: the disordinal cross-level interaction for MRS indicates that mid responses are slower for low MRS levels, and may be faster for high MRS levels compared with directed responses. Hence, we replicated the effect that low MRS trait levels lead to higher response times when giving a midpoint response (Kulas & Stachowski, 2009) and extended this effect by differentiating between areas of significance for different MRS levels (see Figure 4 and Figure A1 in Appendix A).

Speed–Distance Hypothesis. The analyses of response styles and response times have shown that not only personality traits but also response styles follow the speed–distance hypothesis: the more likely a response is for a certain respondent, the faster he or she gives this type of response. The results suggest that the effect on response times is due to a higher confidence in the response when respondents follow their response tendency (i.e., self-schemata) which guides responses and decreases response times (Germeroth et al., 2015; McIntyre, 2011). In contrast, giving a response that is contrary to the respondent's response style level increases difficulty and therefore leads to longer response times (Dunn et al., 1972; Kuiper, 1981).

The speed–distance hypothesis is a robust theory with precise predictions in many fields besides personality research, for example, with regard to emotion clarity (Arndt et al., 2017), in signal detection theory (Maddox et al., 1998) or value research (Bilsky et al., 2013). With the present investigation, we further extend the application of the speed–distance hypothesis and present evidence for its validity in the area of response styles.

Strengths and Limitations of the Current Analysis

The strength of this study is the comparison of three different data sets that consist of different item types, samples, and response category numbers. The fact that results are highly consistent across the three data sets is even more remarkable given the differences between the data sources. Study 1 only used heterogeneous items that refer to different content domains and therefore is ideal to measure response styles (De Beuckelaer et al., 2010; Greenleaf, 1992). However, the sample size with $N = 161$ respondents and $I = 39$ items is sufficient, but not abundant. Since Study 2 combined items of two personality scales with heterogeneous items, it is well suited to measure response styles, while at the same time being generalizable to applied settings on the basis of the two personality scales. However, as a six-category scale was used, MRS cannot be measured in this study. Study 3 contained homogeneous items assessing five different traits from organizational psychology. Since intercorrelations between items were moderate (mean absolute correlation: $r = .23$ in contrast to $r = .11$ in Studies 1 and 2), response styles can be measured across the different content scales (see also Wetzel & Carstensen, 2017, for a discussion on response style measurement across scales). Study 3 demonstrates that the results obtained in Study 1 and Study 2 are generalizable to applied measurement contexts. Besides the applied context in which the study was conducted, the main advantages of this data set is the large sample size, the high variances of response style traits, and a large power. Overall, the high consistency of effects between these different data sources underpins the results' robustness, stability, and generalizability.

At first sight, Figures 3 and 4 indicate a restricted range of observed ARS and MRS trait levels. However, the observed variances are comparable to values reported in the literatures, where ERS variance is typically largest (see, e.g., Böckenholt, 2012; Böckenholt & Meiser, 2017; Plieninger, 2017; Plieninger & Heck, 2018). Furthermore, ARS levels close to 0 or 1 would indicate 0 or 100%, respectively, agree responses which is extremely unlikely. With respect to the response style correlations, the observed values indicate certain relationships that are comparable with other studies. In more detail, ERS and ARS had zero to moderately positive correlations (Study 1: $r = .06$; Study 2: $r = -.08$; Study 3: $r = .40$), ERS and MRS were negatively

associated (Study 1: $r = -.49$; Study 3: $r = -.66$) and ARS and MRS were negatively correlated (Study 1: $r = -.48$; Study 3: $r = -.62$; see, e.g., Böckenholt & Meiser, 2017; Bolt et al., 2014; Plieninger, 2017; Plieninger & Heck, 2018, for similar relationships). Hence, we find that response styles are related to each other to a certain extent, such that, for example, a preference for extreme categories may come with an avoidance of the middle category. Some of these relationships are substantial, but, at the same time, do not indicate a redundancy or linear dependency among response styles.

The positive main effect that respondents with high ERS levels take more time to respond is a result of an exploratory study and contradicts previous assumptions and findings in the literature (Aichholzer, 2013; Casey & Tryon, 2001; Krosnick, 1999). The result suggests that high ERS levels may be associated with an increased rather than decreased cognitive effort, but more studies are necessary to test and corroborate this effect.

A major challenge when analyzing response times is the noise that is inherent in the data (Fazio, 1990; Lo & Andrews, 2015; Ratcliff, 1993). With our multilevel modeling approach, we were able to separate variance components in response times that are due to specific responses (Level 1), respondents' response style traits (Level 2) and their cross-level interactions. Before the main analyses, we made several choices to preprocess response time data, such as excluding responses to initially omitted items when respondents where redirected to the survey page, responses correcting previously given responses, and response time outliers.[5] Across all preprocessing steps, we paid careful attention to use procedures that are well embedded in the response times literature connected to rating scale responses (Bassili & Fletcher, 1991; Höhne & Schlosser, 2018; Mayerl & Urban, 2008; Mulligan et al., 2003), and applied the same procedures in all three studies.

In this analysis, response times served as indicators of response processes, for example, of spontaneous or deliberate response modes. However, response times are not pure process measures. When interpreting changes in response times, one must be aware that implications are based on assumptions on the relation of response times and cognitive processes. The relation of response times and cognitive processes are substantiated by evidence in the literature (see, e.g., Lo & Andrews, 2015), but remain presumed associations as processes themselves are always unobserved.

Directions for Future Research

This research project opens up new areas for future research. While we focused on extreme, acquiescent, and mid responding as response tendencies that occur in rating scale measurement, other response biases such as social desirable or careless responding exist (Andersen & Mayerl, 2017;

Dunn et al., 1972; Ellingson et al., 2001; Meade & Craig, 2012). These may similarly be analyzed with respect to response times. Besides, other process measures, such as eye-tracking and mouse-tracking or even fMRI and EEG measures, could provide useful insights into cognitive processes in rating scale usage. Similar to response times, these process measurement methods differentiate between spontaneous and deliberate response processes, but may also provide information on the guidance of attention, such as whether respondents reread a question or encounter difficulties in the response mapping process (Franco-Watkins & Johnson, 2011; Kamoen et al., 2011; van Hooft & Born, 2012). fMRI and EEG measures may additionally provide insights into physiological correlates of response speed.

The relation of response styles and response times might inform the measurement of content traits and of response biases. So far, response times inform the measurement of personality traits and increase, for example, test information (Ferrando & Lorenzo-Seva, 2007; Ranger & Ortner, 2011). However, our results show that response times are not only indicators of the cognitive process with regard to the content trait but also with regard to response styles. Thus, response times measure several processes: response processes related to the content of the items as well as processes underlying response style usage. Future research should evaluate the potential to improve measurement of personality variables when incorporating response style as well as response time information.

While our main goal was to describe how response styles manifest themselves in the response process, the findings presented in this article may lead to further investigations providing practical guidance for applied measurement situations. With this regard, response times may be analyzed using data originating from experiments in which certain assessment characteristics are manipulated. For instance, one could vary the number of response categories or present items in random order to assess differences in response style effects on response times in different measurement settings. Furthermore, information on item characteristics (see supplementary material available online) may be used to further examine effects of item features on response times.

Furthermore, we did not only collect response times for each item but also for each mouse click that the respondent made on the survey page. Thus, collecting response times in such a way may provide useful information in test construction and item selection. Response time data of this kind allows one to evaluate whether responses to specific items were changed more often, or whether reversed-coded items are difficult to process cognitively. Furthermore, changing a given response may be an indicator of high deliberation and motivation of the respondent with regard to the survey. Hence, future research could evaluate whether such correction of responses may be negatively related to careless responding (Meade & Craig, 2012).

Conclusion

Our analyses have shown that agree and midpoint responding follow a joint cognitive process that is qualitatively different from extreme responding: respondents need more time to give agree and midpoint, but not extreme, responses and respondents with high ERS traits take more time to respond, while this is not the case for respondents with high ARS and MRS traits. However, extreme, acquiescent, and midpoint response styles accelerate response times when the given response is in line with the latent response style trait. This finding indicates that when respondents follow their response styles, their self-schemata guide and therewith accelerate item responses as proposed by the speed–distance hypothesis. Our analyses suggest that every respondent employs some type of response tendency when reacting to a rating scale and that the area of a "neutral" response is actually quite small. The joint result of our studies may furthermore guide future developments in designing testing situations to improve psychological assessment.

Appendix A

Magnitude of the Cross-Level Interaction Effect on Response Times

Figure A1 illustrates the predicted magnitude of the effect of response type (extreme, agree, and mid responses) on response times in the three data sets. On the x-axis, we see the percentage change in response times when the respondent gives an extreme response compared to a non-extreme response (upper row), an agree response compared to a non-agree response (middle row) or a mid response compared to a directed response (lower row), hence given a certain response style trait level. When this effect is negative, response times decrease when the respondents gives a certain response (extreme, agree, or midpoint); when it is positive, response times increase when the respondent gives a certain response. On the y-axis, we see the frequency of respondents in the sample for whom this effect takes place. For example, in the upper row the second bar from the right in Study 1 indicates that for more than 50 respondents in the sample, response times decreased by approximately 3% when giving an extreme response compared to a non-extreme response.

We can see that for extreme responses, the effect is negative for the whole sample (see also Figure 2). In contrast, for agree responses, negative as well as positive effects have occurred (see the disordinal interaction in Figure 3); the same applies to mid responses where response times increased for a majority of the sample, but also decreased for a subsample, when giving a mid response (see the disordinal interaction in Figure 4).
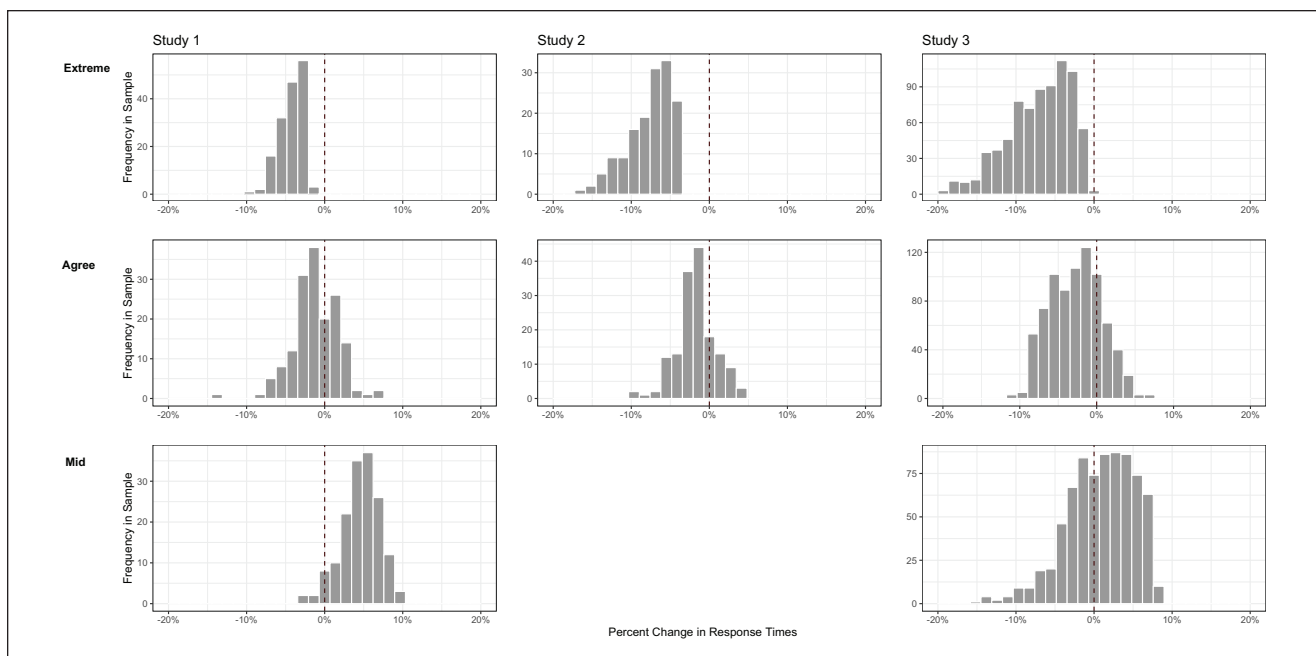

Figure A1. Histogram plots illustrating cross-level interaction effects in terms magnitude of the impact of response type (extreme, agree, and midpoint) on response times on the respondent level in the three studies; *x*-axis shows the percentage change in response times for one respondent, hence given a certain response style trait level, when giving an extreme compared to a non-extreme (upper row), an agree compared to a non-agree (middle row), or a midpoint compared to a directed (lower row) response, *y*-axis shows the frequency of occurrence of this effect in the analysis sample.

Acknowledgments

Declaration of Conflicting Interests

Funding

ORCID iDs

Mirka Henninger https://orcid.org/0000-0003-4676-2361
Hansjörg Plieninger https://orcid.org/0000-0002-4416-300X

Supplemental Material

Supplemental material for this article is available online.

Notes

1. See https://osf.io/gqb4y/
2. However, 13 participants were redirected to a previous page in Study 1, 28 participants were redirected in Study 2. A majority of participants initially omitted one or two items on the previous page, two participants omitted all items of the Honesty–Humility scale. In Study 3, no participants were redirected to previous pages.
3. In line with the tradition of item response theory modeling that is commonly used to account for response styles (e.g., Böckenholt, 2012; Wetzel & Carstensen, 2017), we specified persons as random and items as fixed effects instead of using a model with crossed-random effects.
4. Furthermore, we used the packages splitstackshape, gtools, stringr, dplyr, and tidyr for data management (Mahto, 2018; Warnes et al., 2018; Wickham, 2018; Wickham et al., 2018; Wickham & Henry, 2018) as well as gridExtra and ggplot2 for plotting (Auguie, 2017; Wickham, 2016).
5. As a robustness check, we excluded all responses to items that received more than one click (rather than keeping the first, spontaneous response). The pattern of the estimates remained unchanged, but two effects were no longer significant. This is attributable to losing power when the sample size is reduced, which is corroborated by the fact that significance was not affected in Study 3, which had the largest power. The effects which were no longer significant were the interaction effect for ERS in Study 1 and the Level 1 ARS effect in Study 2.

References

Abad, F. J., Sorrel, M. A., Garcia, L. F., & Aluja, A. (2018). Modeling general, specific, and method variance in personality measures: Results for ZKA-PQ and NEO-PI-R. *Assessment*, *25*(8), 959-977. https://doi.org/10.1177/1073191116667547

Aichholzer, J. (2013). Intra-individual variation of extreme response style in mixed-mode panel studies. *Social Science Research*, *42*(3), 957-970. https://doi.org/10.1016/j.ssresearch.2013.01.002

Akrami, N., Hedlund, L. E., & Ekehammar, B. (2007). Personality scale response latencies as self-schema indicators: The inverted-U effect revisited. *Personality and Individual Differences*, *43*(3), 611-618. https://doi.org/10.1016/j.paid.2006.12.005

Andersen, H., & Mayerl, J. (2017). Social desirability and undesirability effects on survey response latencies. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *135*(1), 68-89. https://doi.org/10.1177/0759106317710858

Arndt, C., Lischetzke, T., Crayen, C., & Eid, M. (2017). The assessment of emotional clarity via response times to emotion items: Shedding light on the response process and its relation to emotion regulation strategies. *Cognition and Emotion*, *32*(3), 530-548. https://doi.org/10.1080/02699931.2017.1322039

Auguie, B. (2017). *gridExtra: Miscellaneous functions for "grid" graphics* (R Package Version 2.3) [Computer software]. https://cran.r-project.org/package=gridExtra

Austin, E. J., Deary, I. J., & Egan, V. (2006). Individual differences in response scale use: Mixed Rasch modelling of responses to NEO-FFI items. *Personality and Individual Differences*, *40*(6), 1235-1245. https://doi.org/10.1016/j.paid.2005.10.018

Bassili, J. N., & Fletcher, J. F. (1991). Response-time measurement in survey research: A new method for Cati and a new look at nonattitudes. *Public Opinion Quarterly*, *55*(3), 331-346.

Bassili, J. N., & Scott, B. S. (1996). Response latency as a signal to question problems in survey research. *Public Opinion Quarterly*, *60*(3), 390-399. https://doi.org/10.1086/297760

Bauer, D. J., & Curran, P. J. (2005). Probing interactions in fixed and multilevel regression: Inferential and graphical techniques. *Multivariate Behavioral Research*, *40*(3), 373-400. https://doi.org/10.1207/s15327906mbr4003_5

Bilsky, W., Borg, I., Janik, M., & Groenen, P. (2013). Children's value structures: Imposing theory-based regional restrictions onto an ordinal MDS solution. In *Proceedings of the 14th Facet Theory Conference* (pp. 25-40). https://dspace.uevora.pt/rdpc/bitstream/10174/10933/1/Facet%20Theory%20-%20Proceedings%2014FTC.pdf

Böckenholt, U. (2012). Modeling multiple response processes in judgment and choice. *Psychological Methods*, *17*(4), 665-678. https://doi.org/10.1037/a0028111

Böckenholt, U. (2017). Measuring response styles in Likert items. *Psychological Methods*, *22*(1), 69-83. https://doi.org/10.1037/met0000106

Böckenholt, U., & Meiser, T. (2017). Response style analysis with threshold and multi-process IRT models: A review and tutorial. *British Journal of Mathematical and Statistical Psychology*, *70*(1), 159-181. https://doi.org/10.1111/bmsp.12086

Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response style. *Applied Psychological Measurement*, *33*, 335-352. https://doi.org/10.1177/0146621608329891

Bolt, D. M., Lu, Y., & Kim, J.-S. (2014). Measurement and control of response styles using anchoring vignettes: A model-based approach. *Psychological Methods*, *19*(4), 528-541. https://doi.org/10.1037/met0000016

Borsboom, D., Kievit, R. A., Cervone, D. P., & Hood, S. B. (2009). The two disciplines of scientific psychology, or: The disunity of psychology as a working hypothesis. In J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra, & N. Chaudary (Eds.), *Developmental process methodology in the social and developmental sciences* (pp. 67-89). Springer. https://doi.org/10.1007/978-0-387-95922-1

Cabooter, E. (2010). *The impact of situational and dispositional variables on response styles with respect to attitude measures* [Doctoral Dissertation, Ghent University]. Ghent University Archive. https://biblio.ugent.be/publication/4333765/file/4427719

Callegaro, M., Yang, Y., Bhola, D. S., Dillman, D. A., & Chin, T.-Y. (2009). Response latency as an indicator of optimizing in online questionnaires. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, *103*(1), 5-25. https://doi.org/10.1177/075910630910300103

Casey, M. M., & Tryon, W. W. (2001). Validating a double-press method for computer administration of personality inventory items. *Psychological Assessment*, *13*(4), 521-530. https://doi.org/10.1037/1040-3590.13.4.521

Couch, A., & Keniston, K. (1960). Yeasayers and naysayers: Agreeing response set as a personality variable. *Journal of Abnormal and Social Psychology*, *60*(2), 151-174. https://doi.org/10.1037/h0040372

De Beuckelaer, A., Weijters, B., & Rutten, A. (2010). Using ad hoc measures for response styles: A cautionary note. *Quality & Quantity*, *44*(4), 761-775. https://doi.org/10.1007/s11135-009-9225-z

Dunn, T. G., Lushene, R. E., & O'Neil, H. F. (1972). Complete automation of the MMPI and a study of its response latencies. *Journal of Consulting and Clinical Psychology*, *39*(3), 381-387. https://doi.org/10.1037/h0033855

Eid, M., & Rauber, M. (2000). Detecting measurement invariance in organizational surveys. *European Journal of Psychological Assessment*, *16*(1), 20-30. https://doi.org/10.1027//1015-5759.16.1.20

Eisenberger, R., Huntington, R., Hutchison, S., & Sowa, D. (1986). Perceived organizational support. *Journal of Applied Psychology*, *71*(3), 500-507. https://doi.org/10.1037//0021-9010.71.3.500

Ellingson, J. E., Smith, D. B., & Sacket, P. R. (2001). Investigating the influence of social desirability on personality factor structure. *Journal of Applied Psychology*, *86*(1), 122-133. https://doi.org/10.1037//0021-9010.86.1.122

Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 74-97). Sage.

Fekken, G. C., & Holden, R. R. (1992). Response latency evidence for viewing personality traits as schema indicators. *Journal of Research in Personality*, *26*(2), 103-120. https://doi.org/10.1016/0092-6566(92)90047-8

Fekken, G. C., & Holden, R. R. (1994). The construct validity of differential response latencies in structured personality tests.

Canadian Journal of Behavioural Science, *26*(1), 104-120. https://doi.org/10.1037/0008-400X.26.1.104

Ferrando, P. J., & Lorenzo-Seva, U. (2007). An item-response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement*, *31*(6), 525-543. https://doi.org/10.1177/0146621606295197

Fladerer, M., Kugler, S., & Misterek, L. (2019). *Leading 'us' for better health: A multilevel exploration of identity leadership and burnout*. Manuscript in preparation.

Franco-Watkins, A. M., & Johnson, J. G. (2011). Decision moving window: Using interactive eye tracking to examine decision processes. *Behavior Research Methods*, *43*(3), 853-863. https://doi.org/10.3758/s13428-011-0083-y

Fuhrman, R. W., & Funder, D. C. (1995). Convergence between self and peer in the response-time processing of trait-relevant information. *Journal of Personality and Social Psychology*, *69*(5), 961-974. https://doi.org/10.1037/0022-3514.69.5.961

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly*, *73*(2), 349-360. https://doi.org/10.1093/poq/nfp031

Germeroth, L. J., Wray, J. M., & Tiffany, S. T. (2015). Response time to craving-item ratings as an implicit measures of craving-related processes. *Clinical Psychological Science*, *3*(4), 530-544. https://doi.org/10.1177/2167702614542847

Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, *56*(3), 328-351. https://doi.org/10.1086/269326

Grimm, S., & Church, A. (1999). A cross-cultural study of response biases in personality measures. *Journal of Research in Personality*, *33*(4), 415-441. https://doi.org/10.1006/jrpe.1999.2256

Gross, S. R., Holtz, R., & Miller, N. (1995). Attitude certainty. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude strength: Antecedents and consequences* (pp. 215-245). Lawrence Erlbaum.

Hallquist, M. N., & Wiley, J. F. (2018). *MplusAutomation*: An R package for facilitating large-scale latent variable analyses in M*plus*. *Structural Equation Modeling*, *25*(4), 621-638. https://doi.org/10.1080/10705511.2017.1402334

Hamilton, D. L. (1968). Personality attributes associated with extreme response style. *Psychological Bulletin*, *69*(3), 192-203. https://doi.org/10.1037/h0025606

Hanley, C. (1965). Personality item difficulty and acquiescence. *Journal of Applied Psychology*, *49*(3), 205-208. https://doi.org/10.1037/h0022107

He, J., & Van De Vijver, F. J. R. (2013). A general response style factor: Evidence from a multi-ethnic study in the Netherlands. *Personality and Individual Difference*, *55*(7), 794-800. https://doi.org/10.1016/j.paid.2013.06.017

Heck, D. W., & Erdfelder, E. (2016). Extending multinomial processing tree models to measure the relative speed of cognitive processes. *Psychonomic Bulletin & Review*, *23*(5), 1440-1465. https://doi.org/10.3758/s13423-016-1025-6

Hibbing, M. V, Cawvey, M., Deol, R., Bloeser, A. J., & Mondak, J. J. (2017). The relationship between personality and response patterns on public opinion surveys: The Big Five, extreme response style, and acquiescence response style. *International Journal of Public Opinion Research*, *31*(1), 161-177. https://doi.org/10.1093/ijpor/edx005

Höhne, J. K., & Schlosser, S. (2018). Investigating the adequacy of response time outlier definitions in computer-based web surveys using paradata SurveyFocus. *Social Science Computer Review*, *36*(3), 369-378. https://doi.org/10.1177/0894439317710450

Kamoen, N., Holleman, B., Mak, P., Sanders, T., & van den Bergh, H. (2011). Agree or disagree? Cognitive processes in answering contrastive survey questions. *Discourse Processes*, *48*(5), 355-385. https://doi.org/10.1080/0163853X.2011.578910

Kievit, R. A., Frankenhuis, W. E., Waldorp, L. J., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*, Article e513. https://doi.org/10.3389/fpsyg.2013.00513

Knowles, E. S., & Condon, C. A. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology*, *77*(2), 379-386. https://doi.org/10.1037/0022-3514.77.2.379

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, *50*, 537-567.

Kuiper, N. A. (1981). Convergent evidence for the self as a prototype: The "inverted U RT effect" for self and other judgments. *Personality and Social Psychology Bulletin*, *7*(3), 438-443. https://doi.org/10.1177/014616728173012

Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered Likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, *43*(3), 489-493. https://doi.org/10.1016/j.jrp.2008.12.005

Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO Personality Inventory: Two new facet scales and an observer report form. *Psychological Assessment*, *18*(2), 182-191. https://doi.org/10.1037/1040-3590.18.2.182

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology*, *24*(7), 1003-1020. https://doi.org/10.1002/acp.1602

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, Article e01171. https://doi.org/10.3389/fpsyg.2015.01171

Lüdtke, O., Marsh, H. W., Robitzsch, A., & Trautwein, U. (2011). A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy-bias trade-offs in full and partial error correction models. *Psychological Methods*, *16*(4), 444-467. https://doi.org/10.1037/a0024376

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: A new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, *13*(3), 203-229. https://doi.org/10.1037/a0012869

Machunsky, M., & Meiser, T. (2006). Personal need for structure als differenzialpsychologisches Konstrukt in der Sozialpsychologie. *Zeitschrift Für Sozialpsychologie*, *37*(2), 87-97. https://doi.org/10.1024/0044-3514.37.2.87

Maddox, W. T., Ashby, F. G., & Gottlob, L. R. (1998). Response time distributions in multidimensional perceptual categorization. *Perception & Psychophysics*, *60*(4), 620-637. https://doi.org/10.3758/BF03206050

Mael, F., & Asiforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of Organizational Behavior*, *13*(2), 103-123. https://doi.org/10.1002/job.4030130202

Mahto, A. (2018). *splitstackshape: Stack and reshape datasets after splitting concatenated values* (R Package Version 1.4.6) [Computer software]. https://cran.r-project.org/package=splitstackshape

Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: Integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, *44*(6), 764-802. https://doi.org/10.1080/00273170903333665

Mayerl, J. (2013). Response latency measurement in surveys: Detecting strong attitudes and response effects. *Survey Methods: Insights from the Field*, Article e00005. https://doi.org/10.13094/SMIF-2013-00005

Mayerl, J., & Urban, D. (2008). *Antwortreaktionszeiten in Survey-Analysen* [Response times in survey analses] (1. Auflage). VS Verlag für Sozialwissenschaften.

McIntyre, H. H. (2011). Investigating response styles in self-report personality data via a joint structural equation mixture modeling of item responses and response times. *Personality and Individual Differences*, *50*(5), 597-602. https://doi.org/10.1016/j.paid.2010.12.001

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437-455. https://doi.org/10.1037/a0028085

Meisenberg, G., & Williams, A. (2008). Are acquiescent and extreme response styles related to low intelligence and education? *Personality and Individual Differences*, *44*(7), 1539-1550. https://doi.org/10.1016/j.paid.2008.01.010

Meiser, T., & Machunsky, M. (2008). The personal structure of personal need for structure: A mixture-distribution Rasch analysis. *European Journal of Psychological Assessment*, *24*(1), 27-34. https://doi.org/10.1027/1015-5759.24.1.27

Moors, G. (2008). Exploring the effect of a middle response category on response style in attitude measurement. *Quality & Quantity*, *42*(6), 779-794. https://doi.org/10.1007/s11135-006-9067-x

Mulligan, K., Grant, J. T., Mockabee, S. T., & Monson, J. Q. (2003). Response latency methodology for survey research: Measurement and modeling strategies. *Political Analysis*, *11*, 289-301. https://doi.org/10.1093/pan/mpg004

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Muthén & Muthén.

Naemi, B. D., Beal, D. J., & Payne, S. C. (2009). Personality predictors of extreme response style. *Journal of Personality*, *77*(1), 261-286. https://doi.org/10.1111/j.1467-6494.2008.00545.x

Neubauer, A. C., & Malle, B. F. (1997). Questionnaire response latencies: Implications for personality assessment and self-schema theory. *European Journal of Psychological Assessment*, *13*, 109-117. https://doi.org/10.1027/1015-5759.13.2.109

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). Academic Press. https://doi.org/10.1016/B978-0-12-590241–0.50006-X

Pfister, M. (2018). *Are extreme and acquiescent response styles related to the implicit self-concept of personality?* [Unpublished bachelor thesis] University of Mannheim.

Plieninger, H. (2017). Mountain or molehill? A simulation study on the impact of response styles. *Educational and Psychological Measurement*, *77*(1), 32-53. https://doi.org/10.1177/0013164416636655

Plieninger, H., & Heck, D. W. (2018). A new model for acquiescence at the interface of psychometrics and cognitive psychology. *Multivariate Behavioral Research*, *53*(5), 633-654. https://doi.org/10.1080/00273171.2018.1469966

Plieninger, H., Henninger, M., & Meiser, T. (2019). *An experimental comparison to the effect of different response formats on response styles*. Manuscript submitted for publication.

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*, *31*(4), 437-448. https://doi.org/10.3102/10769986031004437

R Core Team (2019). *R: A language and environment for statistical computing*. https://www.r-project.org

Ranger, J., & Ortner, T. M. (2011). Assessing personality traits through response latencies using item response theory. *Educational and Psychological Measurement*, *71*(2), 389-406. https://doi.org/10.1177/0013164410382895

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510-532. https://doi.org/10.1037/0033-2909.114.3.510

Riggs, M. L., Warka, J., Babasa, B., Betancourt, R., & Hooker, S. (1994). Development and validation of self-efficacy and outcome expectancy scales for job-related applications. *Educational and Psychological Measurement*, *54*(3), 793-802. https://doi.org/10.1177/0013164494054003026

Rogers, T. B. (1973). Toward a definition of the difficulty of a personality item. *Psychological Reports*, *33*(1), 159-166. https://doi.org/10.2466/pr0.1973.33.1.159

Rollock, D., & Lui, P. P. (2016). Measurement invariance and the five-factor model of personality: Asian international and Euro American cultural groups. *Assessment*, *23*(5), 571-587. https://doi.org/10.1177/1073191115590854

Sarubin, N., Gutt, D., Giegling, I., Böhner, M., Hilbert, S., Krähenmann, O., Wolf, M., Jobst, A., Lena Sabaß, L., Rujescu, D., Falkai, P., & Padberg, F. (2015). Erste Analyse der psychometrischen Eigenschaften und Struktur der deutschsprachigen 10-und 25-Item Version der Connor-Davidson Resilience Scale (CD-RISC) [First analyses of the psychometric properties and structure of the German 10- and 25-item version of the Connor-Davidson Resilience scale (CD-RISK)]. *Zeitschrift Für Gesundheitspsychologie*, *23*(3), 112-122. https://doi.org/10.1026/0943-8149/a000142

Sauer, C., Auspurg, K., Hinz, T., & Liebig, S. (2011). The application of factorial surveys in general population samples: The effects of respondent age and education on response times and response consistency. *Survey Research Methods*, *5*(3), 89-102. https://doi.org/10.18148/srm/2011.v5i3.4625

Steffens, N. K., Haslam, S. A., Reicher, S. D., Platow, M. J., Fransen, K., Yang, J., Ryan, M. K., Jettena, J., Peters, K., & Boen, F. (2014). Leadership as social identity management: Introducing the Identity Leadership Inventory (ILI) to assess and validate a four-dimensional model. *Leadership Quarterly*, *25*(5), 1001-1024. https://doi.org/10.1016/j.leaqua.2014.05.002

Swain, S. D., Weathers, D., & Niedrich, R. W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research*, *45*(1), 116-131. https://doi.org/10.1509/jmkr.45.1.116

Tourangeau, R., Rasinski, K. A., & D'Andrade, R. (1991). Attitude structure and belief accessibility. *Journal of Experimental Social Psychology*, *27*(1), 48-75. https://doi.org/10.1016/0022-1031(91)90010-4

van der Linden, W. J., Entink, R. H. K., & Fox, J.-P. (2010). IRT parameter estimation with response times as collateral information. *Applied Psychological Measurement*, *34*(5), 327-347. https://doi.org/10.1177/0146621609349800

van Dijk, T. k, Datema, F., Piggen, A.-L. J. H. F., Welten, S. C. M., & van de Vijver, F. J. R. (2009). Acquiescence and extremity in cross-national surveys: Domain dependence and country-level correlates. In A. Gari & K. Mylonas (Eds.), *Quod erat demonstrandum: From herodotus' ethnographic journeys to cross-cultural research: Proceedings from the 19th international congress of the international association for cross-cultural psychology* (pp. 149-158). Pedio Books. https://scholarworks.gvsu.edu/iaccp_papers/51/

van Hooft, E. A. J., & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*(2), 301-316. https://doi.org/10.1037/a0025711

van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, *25*(2), 195-217. https://doi.org/10.1093/ijpor/eds021

Warnes, G. R., Bolker, B., & Lumley, T. (2018). *gtools: Various R programming tools* (R Package Version 3.8.1) [Computer software]. https://cran.r-project.org/package=gtools

Weijters, B., Cabooter, E., & Schillewaert, N. (2010a). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, *27*(3), 236-247. https://doi.org/10.1016/j.ijresmar.2010.02.004

Weijters, B., Geuens, M., & Schillewaert, N. (2010b). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement*, *34*(2), 105-121. https://doi.org/10.1177/0146621609338593

Weijters, B., Geuens, M., & Schillewaert, N. (2010c). The stability of individual response styles. *Psychological Methods*, *15*(1), 96-110. https://doi.org/10.1037/a0018721

Wetzel, E., & Carstensen, C. H. (2017). Multidimensional modeling of traits and response styles. *European Journal of Psychological Assessment*, *33*(5), 352-364. https://doi.org/10.1027/1015-5759/a000291

Wetzel, E., Carstensen, C. H., & Böhnke, J. R. (2013). Consistency of extreme response style and non-extreme response style across traits. *Journal of Research in Personality*, *47*(2), 178-189. https://doi.org/10.1016/j.jrp.2012.10.010

Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 years. *Assessment*, *23*(3), 279-291. https://doi.org/10.1177/1073191115583714

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. http://ggplot2.org

Wickham, H. (2018). *stringr: Simple, consistent wrappers for common string operations* (R Package Version 1.3.1) [Computer software]. https://cran.r-project.org/package=stringr

Wickham, H., François, R., Henry, L., & Müller, K. (2018). *dplyr: A grammar of data manipulation* (R Package Version 0.7.6) [Computer software]. https://cran.r-project.org/package=dplyr

Wickham, H., & Henry, L. (2018). tidyr: Easily tidy data with "spread()" and "gather()" functions (R Package Version 0.8.1) [Computer software]. https://cran.r-project.org/package=tidyr

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, *10*(1), 1-17. https://doi.org/10.1207/s15326977ea1001_1

Wise, S. L., & DeMars, C. E. (2006). An application of item response time: The effort-moderated IRT model. *Journal of Educational Measurement*, *43*(1), 19-38. https://doi.org/10.1111/j.1745-3984.2006.00002.x

Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology*, *21*(1), 51-68. https://doi.org/10.1002/acp

Zhang, C., & Conrad, F. G. (2013). Speeding in web surveys: The tendency to answer very fast and its association with straight-lining. *Survey Research Methods*, *8*(2), 127-135. https://doi.org/10.18148/srm/2014.v8i2.5453