Judgment and Decision Making (2023), **Vol. 18**:e34 1–23
doi:10.1017/jdm.2023.31

EMPIRICAL ARTICLE

Connecting the dots: Nonlinear patterns in the presence of symbolic and nonsymbolic numerical standards

Roland Imhoff[1,2] and Paul Barker[1]

[1]Social Cognition Center Cologne, University of Cologne, Cologne, Germany and [2]Social and Legal Psychology, Johannes Gutenberg University Mainz, Mainz, Germany

Corresponding author: Roland Imhoff; Email: roland.imhoff@uni-mainz.de

Received: 9 March 2023; **Revised:** 23 May 2023; **Accepted:** 17 July 2023

Keywords: anchoring; contrast; enumeration; numerical representations

Abstract
Much like other social and nonsocial evaluations, estimates of numerical quantities are susceptible to comparative influences. However, numerical representations can take either a nonsymbolic (e.g., a grouping of dots) or a symbolic numerical form (e.g., Hindu–Arabic numerals), which each produce comparative biases in opposite directions. The current work takes a fine-grained curve fitting approach across a wide range of values to the investigation of their potential nonlinear influence in the context of a numerical estimation task. A series of 3 experiments ($N = 1{,}613$) showed how nonsymbolic standards produce linear contrastive patterns (Study 1), whereas symbolic numerical anchors show a cubic assimilative effect with a leveling off in strength for more extreme standards (Study 2). Sequential contrast from the previous patterns and assimilation to the previous response were found to be present and additive in the presence of both representations but were larger in absence of the symbolic numerical anchors (Study 3).

1. Introduction

Imagine, you have the task to estimate the number of political supporters at the presidential inauguration ceremony of Donald Trump. As virtually any judgment, your exact estimate will be susceptible to potentially irrelevant influences such as a more or less arbitrarily chosen comparison standards (e.g., supporters on Barack Obama's 2009 inauguration). Importantly, this standard could be presented as a merely numerical symbolic representation (i.e., the number 1.8 million) or non-symbolically as a pictorial representation of the crowd. While the former has been well-established to prompt assimilative anchoring effects (the judgment moves closer to the 1.8 million), the latter is more likely to produce contrast (Trump's crowd looks small when compared visually). The present research sought to revisit these different comparative effects in light of recent developments in social comparison research highlighting the importance of nonlinear relationship (Barker and Imhoff, 2021). For social judgments, the strength of assimilation decreases with increased extremity of a comparison standard, ultimately leading to contrast effects, a data pattern that can be modeled by a positive linear in combination with a negative cubic effect (Barker and Imhoff, 2021). Although it is still unclear if such nonlinearity extends to these more basic comparative judgments, theoretical accounts predict both the contrastive effects, reported for nonsymbolic numerical representations, as well as the anchoring effects, for symbolic numerical representations, to produce these complex patterns. The current work will investigate these possibilities for both of these representations of numerical standards as well as their sequential dependencies on previous responses and standards, offering additional insights into the various processes that underlie both phenomena.

To facilitate the flow of our argument, let us provide an illustrative example. Imagine a situation of a protest and a counter-protest (not an infrequent occurrence in most European cities). As a police officer, your task is to estimate the number of protestors in order to allocate a proportional number of accompanying police forces. The comparative question underlying our research is now: How will the size of the counter protest and the modality of receiving this information affect your estimate? Will the number of protestors seem larger next to a small counter-protest and smaller next to a huge one? Will it make a difference if you see the 2 protests as clusters of people or whether a colleague provides you with the number of counter-protestors in a numerical, symbolic way?

Many of our perceptions and judgments rely heavily on the context surrounding the stimuli we are about to judge. In the famous Ebbinghaus illusion, for instance, the central circle is consistently overestimated or underestimated based on whether the circles surrounding it are smaller or larger. Similar contrast effects, where one's experience or estimation is moved away from the comparative state, have been found in both simultaneous and sequential magnitude judgments across various domains (e.g., Cordes et al., 2014; Jesteadt et al., 1977; Preston, 1936). The reason for these contrastive patterns has generally been attributed to shifts in the internal representation of the stimuli themselves, resulting from the inhibition of the areas surrounding the representational center of the standards that produces this contrastive force (Levine and Grossberg, 1976). Indeed, more recent work suggests that this may be a fundamental constraint to the information processing capacity of humans in general (Carandini and Heeger, 2012). Thus, the same crowd of protestors might seem comparatively small next to a big counter-protest.

In addition to these contrastive effects related to the visual display of quantities, written numbers can affect estimates in a completely different manner. In essence, numerical representation can take 2 forms. One is nonsymbolic, an actual representation of the given quantity, be it inauguration guests, protestors, or dots on a screen. The other one is a symbolic representation, as is the case when we represent quantities with Hindu–Arabic numerals. Humans appear not only to have distinct ways to code nonsymbolic and symbolic representations in memory (Roggeman et al., 2007), but these different representations have decisively different comparative effects on numerical estimates. In contrast to the nonsymbolic numerical representations that produce contrastive effects when used as a comparison standard, symbolic representations have been found to robustly produce assimilative biases (Furnham and Boo, 2011). Both symbolic numerical representations provided externally as well as the own responses given in symbolic numerical form in previous trials can act as a reference point to which the current estimate will assimilate (Mochon and Frederick, 2013). These effects are widely described as 'anchoring' effects, in reference to the 'anchoring and adjustment' heuristic first proposed for this phenomenon by Tversky and Kahneman (1974). In these insufficient adjustment explanations, people are assumed to anchor to a symbolic numerical representation as a reference (e.g., the value given in an initial comparative question asking whether the target is higher or lower than it, but also self-generated or merely incidentally displayed anchor values; Critcher and Gilovich, 2008; Epley and Gilovich, 2001) after which they adjust this value until they reach the first plausible response for the judgment stimuli. Assimilation occurs as this adjustment is typically stopped too early to fully eliminate the influence of the anchor, that is, the direction from which the adjustment starts. Thus, seeing a moderately sized protest and learning that the counter-protest is comprised of 150,000 people will lead to a downward adjustment until a window of plausible values of protestors is entered (from the upper end). If the counter-protest is said to involve only 20 people, adjustment will go upward until it enters the same window of plausible values, but this times from the lower end. Thus, the same initial window of plausible values (e.g., 2,000 to 5,000 people) will create drastically different estimates depending on whether the adjustment start from a low anchor (here: 20) or a high anchor (here: 150,000).

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

A more recent explanation for the anchoring effect has been given in the scale distortion theory (Frederick and Mochon, 2012; Mochon and Frederick, 2013), which proposes that the symbolic numerical values simply distort the mapping of the actual estimate on an internal response scale via a contrastive process similar to the one reported for nonsymbolic numbers. More specifically, the presence of a lower/higher symbolic numerical anchor will cause the symbolic numerical estimate, which would have seemed reasonable if no anchor were present, to now seem too high/low due to a contrastive effect. As a result, the now distorted mapping of the unchanged representation of the stimuli will fall on a lower/higher value than it would have otherwise been associated with. To make it more concrete: When you see a protest and judge it to be roughly 4,000 people, this number will seem ridiculously small if you learn that the counter-protest had 150,000 people (a contrast effect). You thus calibrate your response to be bigger. Your belief about the target size has not changed but your use of the numerical scale—what you consider to be an adequate indicator of a small, medium, or large quantity has.

1.1. Nonlinear effects on numerical judgments

A lot of research on contrast as well as assimilation effects compares the influence of either a target judgment in the absence versus presence of a comparison standard or the effect of a high versus a low comparison standard. Such designs with 2 conditions to compare can only uncover linear effects (either contrast or assimilation). It is plausible, however, that the influence of a comparison standard is not just a linear function of its size, but might either attenuate with increasing size (as would be expected for insufficient adjustment where the span of the adjustment process should not affect the width of the window of plausible estimates) or even reverse. An example of the latter effect has recently been provided in the field of comparative judgments involving social stimuli (Barker and Imhoff, 2021). Specifically, estimates of traits (e.g., trustworthiness) from faces assimilated to moderate upward or downward comparison standards presented next to the target stimulus, but once these comparison standards became more extreme, the target judgments contrasted away from these. A fine-grained manipulation of the standards allows to uncover such cubic effects required to adequately model either an attenuation of assimilative effects or even a reversal to contrastive tendencies. Similar sensitivity to the extremity of the comparative stimulus has on also been reported in sequential contrast effects in, for example, attractiveness ratings (Cogan et al., 2013).

As alluded to above, the anchoring effects that dominate when symbolic numerical representations are used as comparisons standards are, in fact, predicted to produce nonlinear effects by some theoretical accounts. For instance, the insufficient adjustment process described earlier clearly suggests a nonlinear relationship, where anchors outside the plausible range of values will produce similar sized shifts in the final judgments as participants will adjust toward the first plausible value no matter the size of the anchor (Tversky and Kahneman, 1974). Indeed, extreme anchors have been reported to produce the same sized effect as more moderate standards implying this nonlinear relationship (Chapman and Johnson, 1994; Mussweiler and Strack, 2001).

1.2. The present research

The current work will attempt to measure these potential nonlinear patterns in sequential judgments in the presence of context stimuli (standards). This allows for judgments to be influenced not only by the present standard but also by the standard presented in the previous trial and their own response to it. In this general approach, we will investigate the effects associated with the 2 distinct comparative processes related to numerical representations: one caused by the current and previous nonsymbolic numerical stimuli, associated with contrastive effects, and the other related to the external symbolic numerical stimuli and previous responses, leading to assimilation. In addition, these different representations of the numerical stimuli and their associated comparative processes may run in parallel, mutually suppress one another, or lead to one form dominating during the formation of the judgment.

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

To investigate all these possibilities, response patterns and their sequential dependencies will be modeled precisely in a comparative dot enumeration task. Initially, this task will include only the nonsymbolic visual stimuli (Study 1), then the symbolic numerical estimates will also be included (Study 2), and finally both patterns will be compared in a single investigation (Study 3). We also explored the data for interactions of concurrent and sequential effects, but only report these in Appendix B, where we also explain why these may be artifacts of the sequential design. All studies report full descriptions of the determination of sample size, data exclusion, manipulations, and measures used. All sample sizes were determined before data collection. Furthermore, all anonymized raw[1] and aggregated data, additional analysis details and supplementary materials for all studies can be found on the Open Science Framework page at https://osf.io/cmw9r/. Studies 1 and 2 were not pre-registered, Study 3 was pre-registered at https://aspredicted.org/vt7xm.pdf.

2. Study 1

In this first study, the comparative pattern associated with nonsymbolic numerical representations will be investigated along with its dependency on the previous comparative stimulus and the previous response given. Based on the existing literature on magnitude estimation (Mo, 1971; Preston, 1936), we expected numerical judgments of the number of dots also to contrast away from the standards that are presented alongside the judgment image (for a similar reasoning, see also Cordes et al., 2014). This effect could take a purely linear form or may show signs of reduced comparative effects for the more extreme standard as reported for social stimuli (Cogan et al., 2013), which would be reflected in a cubic function with a negative linear term and a positive cubic function (Barker and Imhoff, 2021). In addition, previously seen nonsymbolic numerical representations are similarly expected to result in additive contrastive patterns, whereas previous *responses* are expected to form an anchor to which subsequent judgments assimilate.

Using our illustrative example from before, the same protest might seem smaller next to large counter-protest (compared to a tiny counter-protest) as well as after seeing a particularly large (versus small) protest. The numerical estimate, however, will also be influenced in the direction of previously given estimates (i.e., the estimated size will be larger after estimating 10,000 than after estimating 1,000 as these will serve as anchors to start an adjustment process).

2.1. Method

2.1.1. Participants

In lack of a clear effect size estimate, the sample size for this study was based on a previous investigation into the nonlinear comparative relationships among social stimuli with an identical granularity of comparison standards and also z-standardized estimates as dependent variable (Barker and Imhoff, 2021). In that study, a cubic effect as small as $B = 0.002$ was detected to be significantly different from zero using a sample of 160 participants. As such, 160 German speakers were recruited on campus at the University of Cologne to participate in the study for a monetary reward of 10 euros. This sample consisted of 43.8% females and was aged between 17 and 57 years ($M = 23.26$, $SD = 5.00$).

2.1.2. Dot judgment task (DJT)

To gain a large number of estimates, the dot judgment task (DJT) was created. In this task, participants were asked to estimate the number of dots present in a random pattern of white dots on a black background (300 × 300 pixels). These judgment patterns always included between 490 and 510 dots. A substantially large number was used to avoid counting of individual dots. Alongside each judgment

[1] Data about participants' study area and gender are not included for lab studies nor are any timestamps for all studies to ensure the complete anonymity of participants.

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

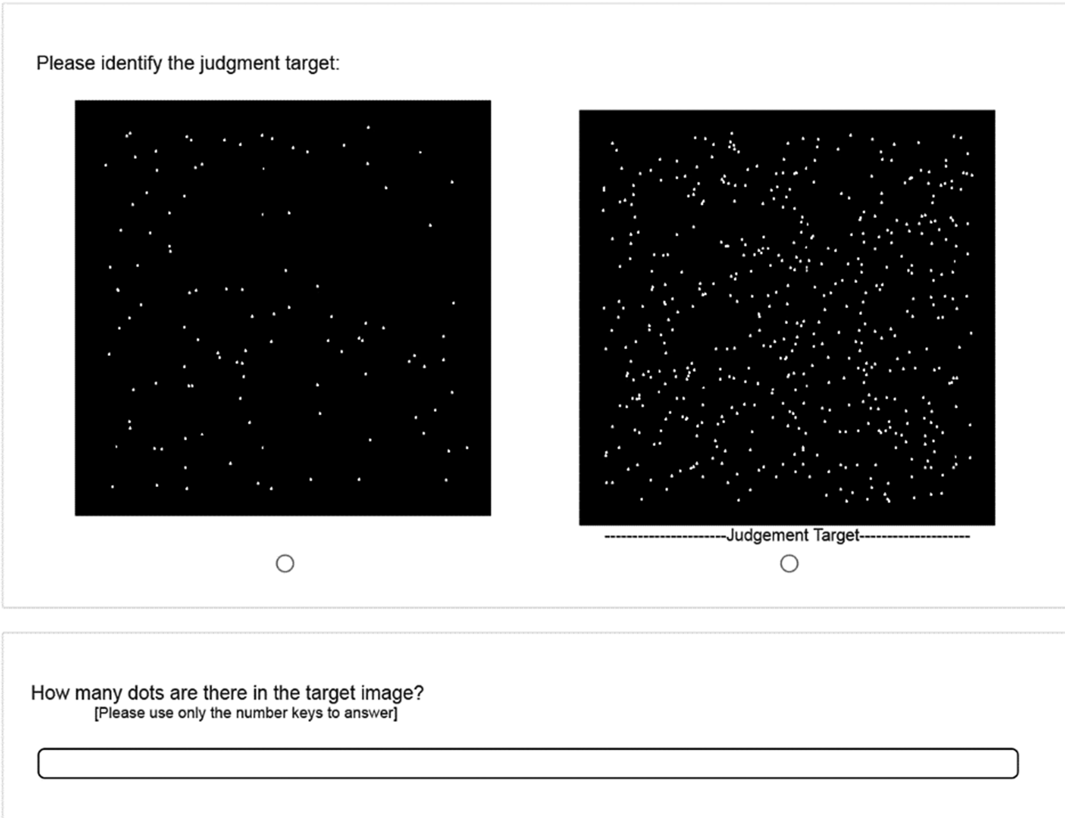

Figure 1. *Example trial for which high values (contrast away from small comparison standard are expected).*

pattern, a second unique comparison pattern was presented (in the same resolution) that included up to 400 more or fewer dots than the judgment pattern in increments of 10 dots, resulting in 81 extremity steps (Figure 1). Both images remained on the screen until a response was given. To ensure that participants judged the correct pattern, they were asked to first identify the pattern labeled as the 'judgment target' prior to giving their estimate. They could indicate this by pressing either 'Z' for the left (on the German keyboards used this was the letter 'Y' in the same location) or 'C' for the right pattern. After this, participants were gave their estimate in an open-ended fashion with no time constraints. In addition to preventing confusion, the accurate identifying of the judgment pattern was also used as an attention check to exclude non-informative responses later on. Participants completed 4 blocks of the 81 trials resulting in 324 total estimates.

2.1.3. Additional measures

Basic demographics such as age, sex, and education were recorded at the start of the study. The Iowa-Netherlands Comparison Orientation Scale (INCOM; Gibbons and Buunk, 1999) was administered at the end of the study for exploratory purposes to see if propensity to engage in social comparisons may also affect these nonsocial comparisons. The INCOM scale includes 11 items (α = 0.78), 2 of which are reverse coded, that are averaged to create the final INCOM score with higher scores reflecting a stronger disposition to engage in social comparisons in daily life. However, the measure was not found to produce interesting findings and will not be further reported in this paper, but the raw data are available in the online materials.

2.1.4. Procedure

Participants were recruited on campus at the University of Cologne were fully informed regarding the general procedure of the study and data storage policy before giving their consent and taking part in the study. Initially, the basic demographics were recorded, after which the DJT was fully explained. Two practice trials were presented and participants were given the opportunity to ask additional questions regarding the task. If no questions remained, the main batch of 324 trials were completed in random order after which participants were debriefed and given their compensation. The study lasted approximately 30 min on average.

2.1.5. Data treatment

All nonnumeric and empty responses were removed and made up 0.1% of trials. Another 5.3% of trials were marked by a failed attention check and were removed. The remaining scores were z-transformed separately per participant to account for personal differences in response ranges. Z-scores above 3 or below −3, or instances where no z-score could be calculated (e.g., because participants always gave the same response and SD was thus zero), were removed, which was the case for 1.8% of trials. Due to co-occurrences of these criteria a total of 7.2% of the original trials were not used in the analyses. Two participants have no longer offered any informative trials, leaving 158 participants with data suitable for use in the main analysis.

In this and all following studies, the main analyses were conducted in mixed models regression using a restricted maximum likelihood estimation (REML; using the lme4 package; Bates et al., 2015) run in R (Version 4.0.4). In addition to the fixed effects of interests, we included random intercepts for subject (all studies) and stimulus set (only Study 1), as well as random slopes per subject for all fixed effects. When these full models failed to converge or led to singular fit, we removed random effects until the model converged and showed no singular fit. We only report these models below. To determine the confidence intervals and p-values for separate fixed effects in this and all following studies, Satterthwaite's approximations were used to estimate the appropriate degrees of freedom (using the lmerTest package; Kuznetsova et al., 2017; and the parameters package in R; Lüdecke et al., 2020).

2.2. Results

Our first model tested the influence of simultaneously presented standards. We thus predicted the estimated number of dots for the actual judgment target by the actual number of presented dots (centered and divided by 10, thus ranging from −1 to +1)[2], and the relative number of dots presented in the adjacent standard (ranging from −400 to +400 in 81 steps). To allow not only for linear effects of the relative number of dots in the standard, we included the linear, as well as the quadratic and cubic term of this relative number (of more or fewer dots than the target) as fixed effects. We also included random intercepts per participant and per stimulus set, as well as random slopes of all 3 polynomials per participant. A linear effect of the number of actually presented dots indicated that participants were indeed sensitive to the actual judgment target, $t(46141.87) = 2.156$, $\beta = 0.010$, 95% CI [0.001, 0.020], $p = .031$, suggesting that an increase in 10 dots in the target corresponded to an increase in estimated dot number of 0.01 z-scores. Speaking to strictly linear contrasts, results yielded only a negative linear effect, $t(158.08) = -9.29$, $\beta = -15.43$, 95% CI [−18.71, −12.15], $p < .001$, with all other orthogonal polynomial terms $p > .61$ (Figure 2). Thus, there was no indication that the overall contrast effect was reduced for the more extreme relative numbers of dots, as would have been indicated by a significant positive cubic effect.

[2]Note that we did not plan (or preregister in Study 3) the inclusion of this predictor, as we were not optimistic that participants would be sensitive to such subtle variations. During the review process, however, it became apparent that effects of standards and prior trials would be much more informative if they were shown to operate in addition to an actual and serious estimation of the target. We thus included the number of dots in the target image as an additional predictor in all analyses. Analyses without this predictor (as planned/ pre-registered for Study 3) yielded results identical not in terms of exact estimates, but in terms of inferences one would draw from them.

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

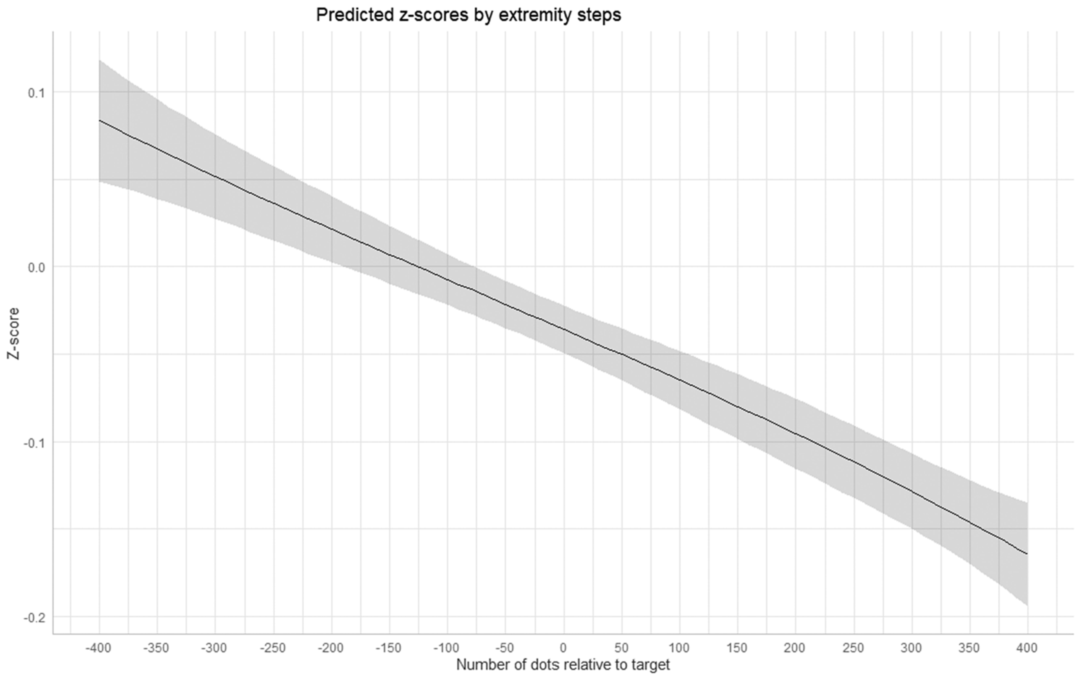

Figure 2. *Predicted Z-scores at each extremity step and 95% CI (created with the ggeffects package, Lüdecke, 2018).*

Following this main analysis for simultaneously presented standard, we followed up with analyses adding potential sequential effects. Specifically, the response given in the previous trial was suspected to have an assimilative influence, as participants might use it as an initial anchor and (insufficiently) adjust from there. Likewise, the relative number of dots in the previous standard might exert an influence of a contrastive nature. As these 2 lagged predictors were dependent (as shown in the initial analyses where there was an effect of relative standard on response within a trial), we investigated their respective influence both in competition to each other in a combined model (reported below) as well as in separate analyses that avoid potential issues with collinearity between the 2 lagged predictors. As these latter separate analyses did not yield the same estimates but suggested identical inferences to be drawn from them, we reported them in Tables A1 and A2.

We predicted participants' *z*-transformed responses by the number of dots in the target, the relative number in the current standard, as well as the lagged *z*-transformed responses and lagged relative number of dots from the previous trial as fixed effects. We dropped all random slopes as their inclusion created singular fit. Results showed that all effects reached significance (Table 1). More precisely, participants were above-chance accurate in guessing the number of dots, larger responses in previous trials were associated with larger responses in the current trial, while the simultaneous contrast effect remained unaffected and of a similar size. The comparison standard in the previous trial did create an incremental contrast effect.

2.3. Discussion

The study showed a robust contrast effect of a simultaneously present standard. Although participants were above chance accurate in estimating the number of presented dots in the target, the larger (smaller) the number of dots presented in the allegedly irrelevant and spatially separated context stimulus, the lower (higher) the numbers of estimated dots in the judgment target. Turning to our police officer,

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

Table 1. *All fixed effects and related statistics from mixed model analysis controlling for lagged response in Study 1.*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.035 [−0.046; −0.024]	0.005	−7.04	14.86	<.001
Dots	0.012 [0.003; 0.021]	0.005	2.58	42650.67	.010
x	−16.11 [−17.30; −14.92]	0.605	−26.61	46086.45	<.001
x^2	0.084 [−1.103; 1.271]	0.606	0.14	46086.07	.890
x^3	−0.464 [−1.650; 0.721]	0.605	−0.77	46085.00	.442
Lagged response (LR)	0.246 [0.240; 0.252]	0.002	87.11	46125.14	<.001
Lagged extremity (LX)	−0.161 [−0.185; −0.138]	0.012	−13.64	46086.95	<.001

Note: Dots are the actual number of dots in the target image, standardized and divided by 10, thus ranging from −1 to +1. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). The lagged response (LR) is the response given in a previous trial (z-standardized). The lagged extremities were rescaled to aid model conversion so that a 1-unit increase represents a step of a 1,000 dots. Random intercepts per participant and stimulus set.

they will be able to give a rough estimate of the number of protestors, but a larger number of counter-protestors will push their estimate down.

Given that target and standard were presented in 2 clearly separate (albeit adjacent) boxes, this finding contrasts with previous investigations that found spatial separation of dot patterns to attenuate all contrastive effect (Cordes et al., 2014). Although we have no definite information of what produced these diverging results, one candidate might be the number of dots to-be-estimated (here: ~400; ~20 in Cordes et al., 2014). Thus, the judgment in the current study might have been experienced as more challenging and thereby leading participants to rely more heavily on the surrounding stimuli to form their judgments. Importantly, however, our results do not suggest that participants merely relied on the contrastive influence of these surroundings as they had no other signal to rely on. On the contrary, they were above-chance accurate in calibrating their estimates to the nuanced differences in actual number of dots in the to-be-judged target (between 490 and 510). Furthermore, the contrastive pattern was found to be completely linear with no signs that the more extreme standards attenuated the effect in any way, as was reported for facial attractiveness (Cogan et al., 2013), allowing speculation whether nonlinearity in contrast effects is specific to the social domain.

In addition to the simultaneous contrast from the comparative stimulus in the current trial, evidence also emerged that the previous stimulus provided an additive contrastive effect in line with the expectations. At the same time, the previous response was found to exert a strong assimilative anchoring effect, with higher estimates in the previous trial predicting higher estimates in the current trial. The fact that this effect of (self-generated) symbolic number representation worked in direct opposition (assimilation) to the simultaneous effect of nonsymbolic representations (contrast) may indeed indicate that the 2 modes of numerical representation seem to independently bias judgments in distinct ways. However, in the current investigation, the symbolic anchors in the previous trials were self-generated responses given by participants, whereas the nonsymbolic standards are externally presented. It may be that offering external symbolic anchors alongside the nonsymbolic numerical standards would lead participants to only consider one mode of numerical processing over the other. The next study, therefore, used the same paradigm, but included informative anchors underneath the comparative dot patterns.

3. Study 2

In this second study, exact symbolic numerical anchors were included underneath the image of the standard, offering immediate nonsymbolic and symbolic numerical information while the estimate is given. Thereby, the current study assessed if the visual information in the nonsymbolic representation still exerted their contrastive influence, or whether anchoring to the symbolic numerical representations

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

dominated in the responses. We also investigated potential nonlinearity of the resulting pattern, as was done for the contrastive patterns. Any insufficient adjustment type process would predict that the anchoring effects would level off as the anchor value becomes more extreme (Mussweiler and Strack, 2001). We expected to see this pattern expressed as a cubic function with positive linear and a slight negative cubic term.

The scale distortion theory (Frederick and Mochon, 2012) on the other hand does not directly predict this leveling off, but a purely linear effect in which the sequential anchors would be expected to work in an additive way (Mochon and Frederick, 2013). Finally, the study also probed if the sequential contrast from previous standards still occurred when concrete symbolic numerical representations are available within each the trial. In other words, the study was designed to provide insight whether a clear numeric anchor will overrule the influence of the previous trial.

3.1. Method

3.1.1. Participants

The number of participants for this study was increased as a shortened online format was implemented. The necessary sample size of 800 participants was again determined using simulations based on the effect sizes and parameters found in previous work in the social comparative domain that used a similar paradigm (Barker and Imhoff, 2021). Therefore, 803 participants recruited via Prolific Academic completed the study for a reward of £1.25. This final sample consisted of 59% females and was aged between 18 and 74 years ($M = 35.48$, $SD = 13.31$).

3.1.2. DJT-41

As in the previous experiment, each trial consisted of a target image (indicated by “judgment target”) for which the number of dots had to be estimated horizontally next to a standard image (both images 300×300 pixel) without any time constraints. For half of the trials, the target image was on the left and the standard on the right, for the other half it was vice versa. In order to rule out confusion regarding which image to estimate, participants had to click a radio button to indicate what they thought the target image was (trials with wrong clicks were excluded). Different from the previous study, the nonsymbolic dot pattern of the standard was now accompanied by the symbolic numeric information displaying the actual number of dots presented in this standard (“This image contains [*between 93 and 901*] dots.”). A visual display of the trials can be found on OSF.

For better use online and to reduce strain on participants, the DJT was shortened to include just 41 trials per participant. In order to keep a similar granularity in measurement steps, however, 2 counterbalancing conditions were created so that half of the participants were randomly assigned to either estimate 41 dot patterns of around 500 dots (±10 dots) next to standards that varied between 400 dots above or below the target image in intervals of 20 dots. The second group judged 40 similar dot images in the presence of standards at each 20-dot-interval between 390 dots below or above the target image as well as one in which the both images had an equal number of dots. In effect, this counterbalancing offers the same 81-step precision in measurement that was also included in the previous study.

3.1.3. Additional measures

Again, some basic demographics were recorded, including age, sex, and education, although the INCOM was no longer administered in this study. An additional item was included at the start of the survey to check for correct image display by asking participants to describe what they saw on a simple image containing 3 shapes of different colors. A small traditional anchoring task, closely modeled on the task and one of the questions in Strack and Mussweiler’s work (1997), was also included after the main task for exploratory purposes not directly related to this line of research. This task and its analysis are described in detail in the additional materials available online. After this, an open-ended item was included to gauge participant’s suspicions regarding the goal of the current study. Finally, a new

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

data-quality item was included in which participants were asked to indicate if their responses were made in earnest or not in order to increase the quality of the data in the final analyses. Participants were guaranteed to not be affected in any way based on their responses to this item, but that it would help the researchers in their investigation if they responded honestly. Responses were given on a Likert scale ranging from "Definitely do not use my data" (1) to "Definitely use my data" (4). Responses of 2 or lower were taken as an indication of poor data-quality and used for the exclusion of participants from the analyses.

3.1.4. Procedure

Participants were fully informed about the data storage policy, procedure, and their rights before they were asked for their consent to take part in the study. After the image display check and demographics items participants completed 2 practice trails to get familiarized with the procedure of the main task. Randomly allocated to 1 of the 2 counter balancing conditions, participants completed the main batch of 41 trials and afterward the traditional anchoring task. After this, the suspicion check item and the data-quality item were presented before participants were debriefed, thanked and given their compensation. The study lasted about 10 min.

3.1.5. Data treatment

Nonnumeric and empty responses made up 1% of trials while 0.7% showed a failed attention check leading to their exclusion. This time, responses that exceeded the maximum or minimum standard value (lower than 100 or higher than 900) were also removed (0.5%) to increase data quality, as such estimates are unlikely to be honest estimates as a clear reference was available. The remaining scores were z-transformed separately per participant to account for personal differences in response ranges. Z-scores above 3 or below −3, or instances where no z-score could be calculated were removed (e.g., because participants always gave the same estimate, resulting in a denominator of zero), which was the case for 0.2% of trials. Due to co-occurrences of these criteria a total of 2.4% of the original trials were not used in the analyses. In addition, 12 participants indicated their data quality was low and should not be included in the analysis. Combined the exclusion criteria left a total of 790 participants with data suitable for use in the main analysis.

3.2. Results

Again, the main analysis took the form of a mixed model regression with REML that included fixed effects for the number of dots present in the target, as well as the standard (and additionally given as a numeric anchor below the standard) in hundreds and modeled up to the third polynomials term. We included random slopes of the 3 polynomials as well as the actual number of dots for participants to account for participant level variation in patterns (no random intercept was estimated due to singular fit issues). As in Study 1, participants were above-chance accurate in detecting subtle differences in the judged target stimulus (main effect of dots; Table 2). In addition, both the linear and cubic term of the comparison standard reached significance in this analysis (Table 2), describing a pattern of assimilation in which the participants' z-transformed estimates increased as the standards became more removed up until a difference of around 300 dots was reached, after which estimates remained stable (Figure 3). Thus, the provided symbolic anchor yielded an exclusively assimilative influence (leveled off toward extreme values) that completely overrode the contrastive influence of the (still present) nonsymbolic representation.

As in the previous study, we conducted a second analysis adding 2 lagged predictors from the previous trials (the relative extremity of the standard as well as participants' self-generated and standardized responses), as well as random slopes for all fixed effects per participant. This model showed an increased fit overall compared to the base model, $\chi^2(4) = 393.93$, $p < .001$. The results suggested that in addition to the (basically unaltered) effects reported above, there are again 2 opposing influences from the previous trial. While a high response in the previous trial increased the estimate,

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

Table 2. *All fixed effects and related statistics from mixed model analysis (Study 2).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.004 [−0.012; 0.006]	0.005	−0.79	27856.53	.429
Dots	0.074 [0.058; 0.090]	0.008	9.03	325.42	<.001
x	45.05 [39.93; 50.17]	2.608	17.28	740.10	<.001
x^2	0.623 [−2.090; 3.336]	1.382	0.45	743.06	.652
x^3	−8.210 [−10.395; −6.025]	1.113	−7.38	755.11	<.001

Note: Dots are the actual number of dots in the target image, standardized and divided by 10, thus ranging from −1 to +1. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). Random slopes of dots and the polynomial terms of x per participant.

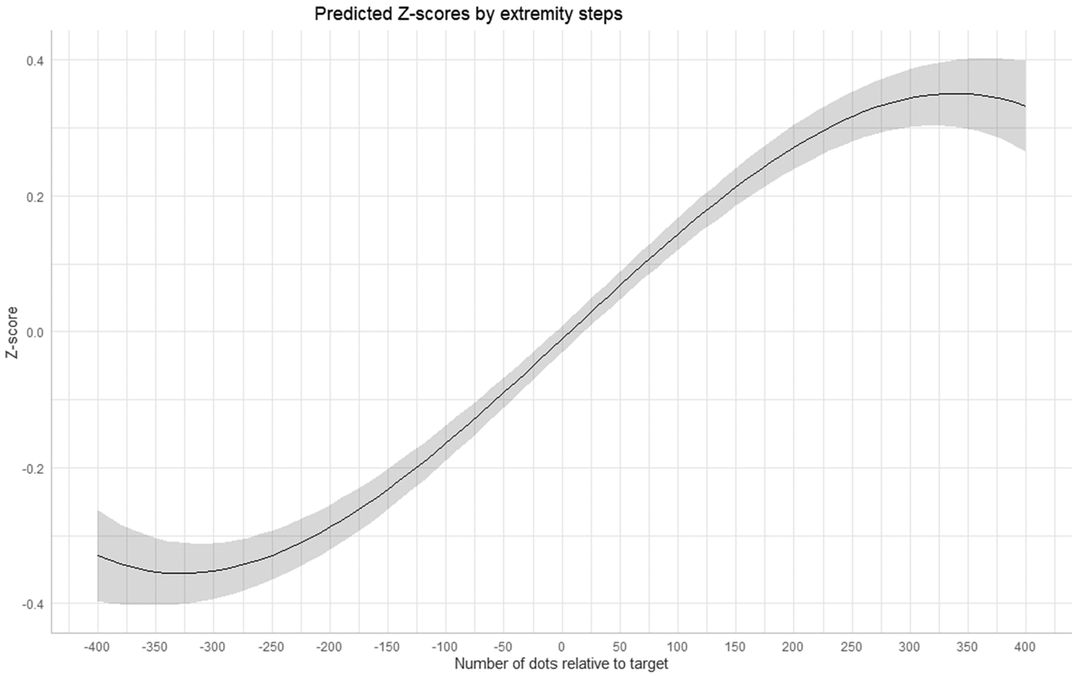

Figure 3. *Predicted intercept adjusted marginal Z-scores at each extremity step and predicted 95% CI.*

a high numerical standard (accompanied by a symbolic anchor) decreased the estimate. This is noteworthy, as the symbolic representation had an assimilative influence for current standards, but a contrastive one for the lagged standards from the previous trial.

As a caveat, however, the assimilative influence of symbolic numerical standard in the same trial was substantially larger (about 5 times as large) than the contrastive influence of the nonsymbolic representations in Study 1. Thus, both lagged predictors show more overlap in this than the previous study, allowing for the possibility of (spurious) suppression effects due to collinearity (Table 3). Separate analyses yielded that all significant main effects remained significant and in the same direction when lagged response and lagged extremity were entered in separate models (Tables A3 and A4).

3.3. Discussion

The addition of symbolic numerical anchors within the DJT paradigm produced a complete flip in comparative patterns. Instead of contrastive influences, standards now produced a purely assimilative effect in line with the anchoring literature. Furthermore, the pattern reflected the nonlinear association

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

Table 3. *All fixed effects and related statistics from mixed model analysis including lagged response and lagged standard extremity (Study 2).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.002 [−0.011; 0.007]	0.005	−0.37	25633.14	.710
Dots	0.075 [0.059; 0.091]	0.008	9.03	391.78	<.001
x	43.38 [38.29; 48.46]	2.589	16.78	738.54	<.001
x^2	0.378 [−2.310; 3.067]	1.369	0.27	744.41	.782
x^3	−7.654 [−7.799; −5.509]	1.093	−7.01	755.05	<.001
Lagged response (LR)	0.090 [0.079; 0.102]	0.005	15.79	785.55	<.001
Lagged extremity (LX)	−0.278 [−0.324; −0.232]	0.023	−11.89	797.67	<.001

Note: Responses are z-standardized within participant to adjust for different response scales and detect outliers. Dots are the actual number of dots in the target image, standardized and divided by 10, thus ranging from −1 to +1. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). The lagged response (LR) is the response given in a previous trial (z-standardized). The lagged extremities (LX) are the actual number of dots presented in the previous trials, rescaled to aid model conversion so that a 1-unit increase represents a step of a 1,000 dots. Random slopes per participants for all fixed effects.

that was predicted by previous work and an anchoring and adjustment type mechanism (Chapman and Johnson, 1994; Mussweiler and Strack, 2001), with the assimilative strength leveling off for extreme standards. This is remarkable as there was no longer any apparent effect of the concurrent nonsymbolic information. Although the police officer in our leading example contrasted their numerical estimate of protestors away from the larger counter-protest this effect completely flipped to assimilation once symbolic information (the actual number of counter-protestors) entered the picture.

However, the sequential anchoring effects proved to be additive and here, the nonnumerical information continued to exert a contrastive effect. Spelling this complex pattern out suggests that the current estimate was assimilated to the concurrent standard and the symbolic (numerical) response given in the previous standard independently, but still contrasted away from the nonsymbolic image in the previous trial. One perceptual explanation may be that, although the symbolic standard is no longer taken into account, the nonsymbolic visual information causes an inhibitory force that suppresses the expression of the current target stimulus. Thus, the contrast from the nonsymbolic numerical standards seems to not be easily disregarded even when clear objective symbolic numerical estimates are present and dominate the judgment process.

Although the concurrent effects in Studies 1 and 2 were clearly in diametrically opposed directions (strong linear contrast in Study 1, even stronger assimilation leveled off toward extreme anchors in Study 2), directly comparing them is constrained by the fact that participants were not randomly allocated to either of the 2 conditions. In addition, Study 1 was a lab study with a large number of trials, Study 2 was considerably shorter. Although none of these differences suggest themselves to explain the different result, we conducted a final, pre-registered study to directly compare these 2 conditions (and their sequential effects).

4. Study 3

As different populations, methods, and contexts were used in Studies 1 and 2, the sizes of the different effects are not directly comparable. To offer a more formal test of these differences, our final study thus combined both to replicate the findings reported so far and explore any differences in sequential effect between the 2 conditions in a pre-registered study (https://aspredicted.org/vt7xm.pdf). Based on the previous findings, we expected a purely negative linear contrastive effect when only nonsymbolic numerical representations were presented as dot patterns, whereas the inclusion of symbolic numerical anchors below the images would reverse the effect, leading to a purely assimilative force that levels out for more extreme standards as seen in a cubic effect with a positive linear and negative cubic function.

4.1. Method

4.1.1. Participants

As the size of both effects was found to be substantially larger than those found in the social comparative domain previously (Barker and Imhoff, 2021), simulations were conducted based on the effect sizes and variance parameters found in the previous 2 studies. The results showed that only 650 participants would be needed to offer over 80% power to detect the smaller simultaneous contrast effect in absence of a numerical anchor. Therefore, 650 participants were recruited on Prolific for a reward of £1.25 and consisted of 59.7% females and was aged between 18 and 80 years ($M = 36.58$, $SD = 13.20$).

4.1.2. Procedure

The overall procedure was identical to Study 2 with only one alteration: It was manipulated between subjects (orthogonal to the counter-balancing factor) whether symbolic numerical anchors informed about the number of dots in the standard or not. All else was identical.

4.1.3. Data treatment

As was the case in the previous study, trials with null responses or only nonnumeric symbols (0.2% of trials), trials in which the attention check was failed (0.8%), and response that exceeded the maximum standard value (6.4%) were removed for the analysis. Remaining scores were z-transformed per participant and truncated above 3 and below −3 (0.3%). A total of 7.7% of all trials were excluded by these criteria. Finally, 12 participants indicated that their data were of low quality and should be excluded, leaving 626 participants' data usable in the analyses.[3]

4.2. Results

Two mixed models regression with ML estimation were implemented with the number of dots as well as polynomials up to the third degree for the standard extremity steps as fixed effects and the orthogonal random slopes for the 3 polynomial terms of standard extremity per each participants. However, one of the models included the condition variable as a fixed effect along with all its interactions. Comparing these 2 models revealed that the anchor condition indeed significantly influenced the dot estimates that participants gave, $\chi^2(4) = 205.55$, $p < .001$, in opposite directions and in line with the predictions (Figure 4). Specifically, the condition moderated the linear, as well as the cubic effect of standard extremity (Table A5). Separate models per condition yielded only a negative linear (contrast) effect of standard extremity in the condition with only nonsymbolic information, as in Study 1 (Table 4). Adding a symbolic anchor flipped this pattern to a positive linear and negative cubic effect, as in Study 2 (Table 5).

To investigate the potential variation of the sequential anchoring effects in the different conditions, a cubic model for the extremity steps with the condition variable, the number of target dots, the lagged response, lagged extremity and all interactions with the condition variable was fitted using ML on trials that included valid lagged responses. This model was then compared to a reduced model that did not include the interaction between the condition variable and lagged response or extremity. This analysis showed that the inclusion of the interaction effects significantly improved the model, $\chi^2(2) = 713.83$, $p < .001$, indicating that the sequential effects differed between conditions. To gain a more detailed insight into the main effects and their sequential dependencies, separate models using REML for both the anchor and non-anchor conditions were modeled.

Fully replicating Study 1, in the no-anchor condition (exclusively nonsymbolic information), the concurrent standard produced a contrast effect while the lagged response served as an assimilative

[3] Both the maximum value and z-score exclusion criteria resulted in significantly more removed trials in the no-anchor condition than in the anchor condition. However, the main analysis still provided substantially similar results when these criteria were not enforced.

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

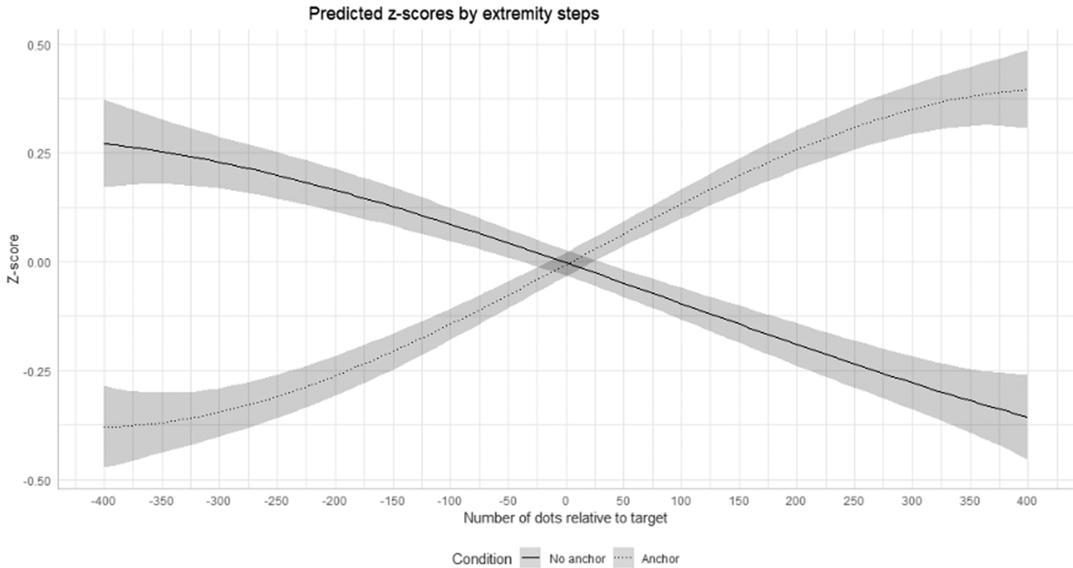

Figure 4. *Predicted intercept adjusted marginal z-scores at each extremity step and predicted 95% CI.*

Table 4. *All fixed effects and related statistics from mixed model analysis in the no-anchor condition.*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.003 [−0.029; 0.024]	0.014	−0.22	765.03	.827
Dots	0.048 [0.018; 0.078]	0.015	3.10	10421.24	.002
x	−0.089 [−0.114; −0.064]	0.013	−7.04	299.26	<.001
x^2	−0.003 [−0.007; 0.001]	0.002	−1.46	267.12	.147
x^3	0.000 [−0.002; 0.003]	0.000	−0.44	291.73	.663

Note: Dots are the actual number of dots in the target image, standardized and divided by 10, thus ranging from −1 to +1. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). Random slopes for polynomials of standard extremity.

Table 5. *All fixed effects and related statistics from mixed model analysis in the anchor condition.*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.011 [−0.040; 0.0190]	0.015	−0.71	535.03	.479
Dots	0.099 [0.074; 0.124]	0.013	7.77	12510.00	<.001
x	0.137 [0.107; 0.168]	0.015	8.97	332.80	<.001
x^2	0.001 [−0.004; 0.006]	0.002	0.46	327.44	.648
x^3	−0.002 [−0.004; −0.001]	0.001	−2.50	335.10	.013

Note: Dots are the actual number of dots in the target image, standardized and divided by 10, thus ranging from −1 to +1. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). Random slopes for polynomials of standard extremity.

anchor and the lagged extremity was associated in the direction of an additional contrast effect (Table 6).

In the anchor condition, all linear effects observed in Study 2 were replicated. The symbolic representation of the concurrent standard overrode the nonsymbolic representation in yielding an assimilation effect. This assimilative effect was additive to the exerted by the response given in the previous trial. The lagged extremity, however, again produced a reliable contrast effect (Table 7).

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

Table 6. *All fixed effects and related statistics from mixed model analysis including lagged response and lagged standard extremity in the no-anchor condition (Study 3).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.001 [−0.025; 0.025]	0.012	−0.11	728.88	.911
Dots	0.058 [0.026; 0.082]	0.014	4.16	9674.60	<.001
x	−0.094 [−0.118; −0.072]	0.011	−8.29	284.61	<.001
x^2	−0.003 [−0.007; 0.000]	0.002	−1.76	263.57	.080
x^3	0.000 [−0.002; 0.002]	0.001	0.23	284.36	.822
Lagged response (LR)	0.362 [0.353; 0.385]	0.015	24.39	289.43	<.001
Lagged extremity (LX)	−0.082 [−0.129; 0.011]	0.037	−2.21	273.93	.028

Note: Responses are z-standardized within participant to adjust for different response scales and detect outliers. Dots are the actual number of dots in the target image, standardized and divided by 10, thus ranging from −1 to +1. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). The lagged response (LR) is the response given in a previous trial (z-standardized). The lagged extremities were rescaled to aid model conversion so that a 1-unit increase represents a step of a 1,000 dots. Random slopes for the 3 polynomials of x, lagged response and lagged extremity.

Table 7. *All fixed effects and related statistics from mixed model analysis including lagged response and lagged standard extremity in the anchor condition (Study 3).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.008 [−0.032; 0.016]	0.012	−0.66	12512.14	.506
Dots	0.100 [0.069; 0.131]	0.016	6.29	354.74	<.001
x	3.134 [0.116; 0.151]	0.009	14.84	12535.81	<.001
x^2	0.001 [−0.003; 0.004]	0.002	0.36	12577.32	.722
x^3	−0.003 [−0.004; −0.001]	0.001	−3.02	12600.25	.003
Lagged response (LR)	0.064 [0.044; 0.084]	0.010	6.31	353.52	<.001
Lagged extremity (LX)	−0.274 [−0.349; −0.199]	0.038	−7.22	350.95	<.001

Note: Responses are z-standardized within participant to adjust for different response scales and detect outliers. Dots are the actual number of dots in the target image, standardized and divided by 10, thus ranging from −1 to +1. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). The lagged response (LR) is the response given in a previous trial (z-standardized). The lagged extremities were rescaled to aid model conversion so that a 1-unit increase represents a step of a 1,000 dots. Random slopes for dots, lagged response, and lagged extremity.

4.3. Discussion

Confirming the findings from the previous studies, the current results show a purely linear contrastive influence of the nonsymbolic numerical standards that turns into an assimilative and cubic pattern once symbolic numerical anchors are included underneath the standards. Furthermore, the sequential anchoring effect (to the self-generated response), although present in both conditions, was smaller when the symbolic anchors were present. This may be a result of the increased information that the objective anchors provide, which has been known to reduce the magnitude of sequential effects by narrowing the range of probable values (Mori, 1998; Ward, 1979; Wilson et al., 1996). The opposite was true for the sequential contrast effect that was substantially larger in the condition with symbolic information and descriptively in the expected direction but not significantly different from zero anymore in the condition with only nonsymbolic information. The third pre-registered study thus largely corroborated the results of the first 2 studies and thus solidifies a rather complex and rich picture of influences on sequential numerical judgments.

5. General discussion

Across 3 studies, we consistently find that participants contrast their numerical estimates away from a (purely nonsymbolic) comparison standard that is either displayed adjacent to the target stimulus

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

or in temporal proximity (the target in the previous trial). As soon as a symbolic representation, a numerical anchor, however, is provided this redundant information completely overrides the contrastive influence of the concurrent standard and yields assimilation to the numerical anchor instead. The same is not true for the contrastive influence of the previous target that still yields a contrastive influence. A self-generated numerical anchor, the response given in the previous trial, exerts an additional and independent assimilation to that anchor. We observe evidence for all these influences independently and in parallel to an above-chance accuracy in estimating the subtle variation (490 to 510) in number of dots in the target image.

Initially, Study 1 showed that robust contrastive effects do occur for spatially separated nonsymbolic numerical representations in a comparative context, despite previous work in this area reporting that spatial separation of dot patterns can attenuate this influence (Cordes et al., 2014). In addition, this contrastive pattern was found to be completely linear regardless of the extremity of the comparison patterns that were presented alongside the target. These findings, thereby, do not support the presence of nonlinear relationships for nonsymbolic numerical patterns that have been previously described in the domain of facial attractiveness (Cogan et al., 2013) and earlier work on magnitude estimations of weights (Sarris, 1968).

However, when symbolic numerical anchors were provided alongside the patterns in Study 2, theses association became purely assimilative in nature. Furthermore, the association took on a slightly cubic effect signaling an upper boundary to the assimilative strength for more extreme standards. The pattern of a positive linear and a negative cubic relation mimics recent findings in the field of social comparisons, with the exception that in the current studies very extreme standards do not produce manifest contrast (Barker and Imhoff, 2021). These findings are compatible with previous work describing the effects of extreme anchors (Chapman and Johnson, 1994; Mussweiler and Strack, 2001), and predictions based on insufficient adjustment accounts (Tversky and Kahneman, 1974). People form an impression of a plausible range of estimates and correct either upward or downward from the given numerical anchor. They stop this correction as soon as the range of plausible estimates is entered. More extreme anchors will then only result in an extended adjustment, not in a shifted range of plausible estimates. An alternative explanation might be that extreme numerical symbolic values bear less informative weight; and thus, the contrasting force of the nonsymbolic representation wins through. Future research presenting either symbolic and nonsymbolic (as in the current study) or only symbolic information can test these 2 explanations.

Additionally, the insufficient adjustment accounts, in which respondents stop adjusting at the boundary of the range of plausible values, do not offer a suitable explanations for the additive sequential anchoring effects that were also found. The scale distortion account, on the other hand, does predict an additive anchoring effect based on the fact that the contrast effects that are ultimately responsible for the distortion of the response scale and eventual assimilative effects themselves work in an additive way (Mochon and Frederick, 2013). In this way, the additive sequential anchoring effects, which can work both to strengthen and to dampen the simultaneous anchoring effect can be explained by a single mechanism.

Future research might also pit the different numerical representation directly against each other. In the current work, the numerical anchor always provided accurate information about the number of dots in the standard (and still completely reversed the influence). Future research might provide numerical anchors inconsistent with the actually represented standard. Although this might hurt the credibility of the anchor and introduce issues of deception (not present in the current research) it could also lead to deeper understanding in how the anchors interact and influence the magnitude estimation.

An interesting note on the complex interaction between externally presented and self-generated anchors is that the strong assimilation found toward the externally presented simultaneous anchor disappeared completely in the sequential trials. Instead, although assimilation was found toward the self-generated anchor of the previous response, the previous standards exerted a clearly contrastive effect even when the symbolic anchor was present.

Therefore, it seems the influence of the previous symbolic anchor is almost completely disregarded in sequential trials, while the previous response is still used to form the next judgment. One reason for this may be that external anchors are often evaluated in a more effortful manner and given extra weight when credible and in agreement with one's own self-generated anchor (Dowd et al., 2014). Indeed, previous work has often emphasized the possible different relevance of self-generated and externally presented anchors (Epley and Gilovich, 2001). This could explain why only the most salient external anchor is considered for each judgment and why it seems to be most influential when it is in the same direction as the self-generated anchor, especially when it may itself not be considered a highly credible estimate (e.g., very extreme and, therefore not similar to the target).

More research would be needed to parse which of these is most the most plausible cause for the complex assimilative patterns found for the simultaneous external and self-generated anchors. Nevertheless, the fine-grained continuous approach of the current paradigm already highlights the complex and nonlinear nature of these symbolic numerical representations compared to the purely linear contrastive influence of nonsymbolic representations. Furthermore, the sequential contrast effects from the previous standards in both conditions (albeit only descriptively there in Study 3) shows that both representations may influence numerical estimations in a relatively independent fashion.

Coming back to our introductory example, the police officer's estimate of the number of protestors will likely be subject to many other influences than just their actual number. Their estimate will contrast away from an adjacent counter-protest (i.e., increase in the presence of a small counter-protest and decrease in presence of a large counter-protest) if they see it. If a colleague just gives them the number of counter-protestors via radio (in a symbolic numerical expression), assimilation will prevail. This fascinating dissociation of the influence acts in addition to the influence of the previous shift. Officers will contrast their estimate away from yesterday's protest but assimilate to their numerical estimate from yesterday.

Our works thus has shown how numerical standards, even when spatially separated from targets, can simultaneously produce assimilative and contrastive biases in numerical estimates depending on whether they are presented symbolically or non-symbolically. Importantly, their influence was highly systematic and robust, even though—from a strictly rational standpoint—the standards were always completely irrelevant to the task at main (estimating the number of dots in the target).

Data availability statement. All data files, analyses, and additional materials related to this article can be found at https://osf.io/cmw9r/.

Funding statement. This project was supported by a DFG grant (IM147/3-2) awarded to R.I. as part of the DFG Research Unit "Relativity in Social Cognition" (FOR 2150).

Competing interest. The authors have no competing interests to declare.

References

Barker, P., & Imhoff, R. (2021). The dynamic interactive pattern of assimilation and contrast: Accounting for standard extremity in comparative evaluations. *Journal of Experimental Social Psychology*, *97*, 104190.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62. https://doi.org/10.1038/nrn3136

Chapman, G. B., & Johnson, E. J. (1994). The limits of anchoring. *Journal of Behavioral Decision Making*, *7*, 223–242. https://doi.org/10.1002/bdm.3960070402

Cogan, E., Parker, S., & Zellner, D. A. (2013). Beauty beyond compare: Effects of context extremity and categorization on hedonic contrast. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 16–22. https://doi.org/10.1037/a0031020

Cordes, S., Goldstein, A., & Heller, E. (2014). Sets within sets: The influence of set membership on numerical estimates. *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 94. https://doi.org/10.1037/a0034131

Critcher, C. R., & Gilovich, T. (2008). Incidental environmental anchors. *Journal of Behavioral Decision Making*, *21*, 241–251.

Dowd, K. W., Petrocelli, J. V., & Wood, M. T. (2014). Integrating information from multiple sources: A metacognitive account of self-generated and externally provided anchors. *Thinking & Reasoning*, *20*, 315–332. https://doi.org/10.1080/13546783.2013.811442

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12*, 391–396. https://doi.org/10.1111/1467-9280.00372

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General*, *141*, 124–133. https://doi.org/10.1037/a0024006

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *Journal of Socio-Economics*, *40*, 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Gibbons, F. X., & Buunk, B. P. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology*, *76*(1), 129–142. https://doi.org/10.1037/0022-3514.76.1.129

Jesteadt, W., Luce, R. D., & Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 92–104. https://doi.org/10.1037/0096-1523.3.1.92

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Levine, D. S., & Grossberg, S. (1976). Visual illusions in neural networks: Line neutralization, tilt after effect, and angle expansion. *Journal of Theoretical Biology*, *61*, 477–504. https://doi.org/10.1007/BF00344744

Lüdecke, D. (2018). ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, *3*(26), 772. doi:10.21105/joss.00772.

Lüdecke, D., Ben-Shachar, M., Patil, I., & Makowski, D. (2020). Extracting, Computing and Exploring the Parameters of Statistical Models using R. *Journal of Open Source Software*, *5*(53), 2445. doi:10.21105/joss.02445

Mo, S. S. (1971). Judgment of temporal duration as a function of numerosity. *Psychonomic Science*, *24*, 71–72.

Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes*, *122*, 69–79. https://doi.org/10.1016/j.obhdp.2013.04.002

Mori, S. (1998). Effects of stimulus information and number of stimuli on sequential dependencies in absolute identification. *Canadian Journal of Experimental Psychology*, *52*, 72–83. https://doi.org/10.1037/h0087282

Mussweiler, T., & Strack, F. (2001). Considering the impossible: Explaining the effects of implausible anchors. *Social Cognition*, *19*, 145–160. https://doi.org/10.1521/soco.19.2.145.20705

Preston, M. G. (1936). Contrast effects and the psychophysical judgments. *The American Journal of Psychology*, *48*, 389–402. https://doi.org/10.2307/1415884

Roggeman, C., Verguts, T., & Fias, W. (2007). Priming reveals differential coding of symbolic and non-symbolic quantities. *Cognition*, *105*, 380–394. https://doi.org/10.1016/j.cognition.2006.10.004

Sarris, V. (1968). Adaptation-level theory: Absolute or relative anchor effectiveness? *Psychonomic Science*, *13*, 307–308. https://doi.org/10.3758/BF03342599

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Ward, L. M. (1979). Stimulus information and sequential dependencies in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 444–459. https://doi.org/10.1037/0096-1523.5.3.444

Wilson, T. D., Houston, C. E., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, *125*, 387–402. https://doi.org/10.1037/0096-3445.125.4.387

Appendix A. Additional analyses

Table A1. *All fixed effects and related statistics from mixed model analysis including lagged response (Study 1).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.031 [−0.036; −0.026]	0.002	−12.38	46110.01	<.001
Dots	0.015 [0.007; 0.023]	0.004	3.59	45956.52	<.001
x	−0.031 [−0.037; −0.025]	0.003	−9.71	158.56	<.001
Lagged response (LR)	0.309 [0.263; 0.340]	0.019	15.50	159.72	<.001

Note: Dots are the actual number of dots in the target image. *x* denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). Random slopes for the standard extremity and lagged response.

Table A2. *All fixed effects and related statistics from mixed model analysis including lagged extremity (Study 1).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.036 [−0.042; −0.031]	0.003	−12.56	47694.77	<.001
Dots	0.011 [0.001; 0.020]	0.005	2.20	47820.67	.028
x	−0.030 [−0.037; −0.024]	0.003	−9.36	158.12	<.001
Lagged extremity (LX)	−0.235 [−0.276; −0.195]	0.021	−11.46	158.13	<.001

Note: Dots are the actual number of dots in the target image. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). Random slopes for the standard extremity and lagged extremity.

Table A3. *All fixed effects and related statistics from mixed model analysis including lagged response (Study 2).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.664 [−0.988; −0.340]	0.165	−4.02	4636.71	<.001
Dots	0.001 [0.001; 0.002]	0.000	4.03	4614.47	<.001
x	0.164 [0.145; 0.183]	0.010	17.10	732.85	<.001
x^2	−0.000 [−0.003; 0.003]	0.002	−0.11	742.32	.910
x^3	−0.005 [−0.006; −0.004]	0.001	−8.13	734.93	<.001
Lagged response (LR)	0.073 [0.062; 0.084]	0.006	12.98	700.42	<.001

Note: Dots are the actual number of dots in the target image. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). Random slopes for the standard extremity polynomials and lagged response.

Table A4. *All fixed effects and related statistics from mixed model analysis including lagged extremity (Study 2).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.736 [−1.055; −0.416]	0.163	−4.52	4938.34	<.001
Dots	0.001 [0.001; 0.002]	0.000	4.53	4915.08	<.001
x	0.163 [0.144; 0.182]	0.010	16.97	731.76	<.001
x^2	−0.000 [−0.003; 0.003]	0.002	−0.11	741.38	.911
x^3	−0.005 [−0.006; −0.004]	0.001	−8.08	734.85	<.001
Lagged extremity (LX)	−0.180 [−0.225; −0.135]	0.023	−7.85	740.77	<.001

Note: Dots are the actual number of dots in the target image. Responses are z-standardized within participant to adjust for different response scales and detect outliers. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). The lagged extremities are the actual number of dots presented in the previous trials, rescaled to aid model conversion so that a 1-unit increase represents a step of a 1,000 dots. Random slopes for the polynomials of the standard.

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

Table A5. *All fixed effects and related statistics from mixed model analysis including lagged response, lagged extremity, condition, and interactions with condition (Study 3).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.004 [−0.024; 0.016]	0.010	−0.44	1174.42	.663
Dots	0.078 [0.060; 0.097]	0.009	8.16	21970.28	<.001
Condition	−0.005 [−0.046; 0.034]	0.020	−0.29	1163.00	.771
x	0.019 [−0.000; 0.039]	0.010	1.93	638.24	.054
x^2	−0.001 [−0.004; 0.002]	0.002	−0.61	616.56	.540
x^3	−0.001 [−0.002; 0.000]	0.001	−1.42	636.89	.156
Lagged response (LR)	0.219 [0.208; 0.230]	0.006	38.53	22447.11	<.001
Lagged extremity (LX)	−0.169 [−0.217; −0.122]	0.024	−6.99	22472.60	<.001
Condition × x	0.235 [0.196; 0.273]	0.020	11.93	618.63	<.001
Condition × x^2	0.004 [−0.002; 0.010]	0.003	1.23	614.82	.219
Condition × x^3	−0.003 [−0.006; −0.001]	0.001	−2.43	613.46	.015
Condition × LR	−0.301 [−0.323; −0.279]	0.011	−26.49	22447.32	<.001
Condition × LX	−0.224 [−0.319; −0.129]	0.048	−4.63	22472.63	<.001

Note: Dots are the actual number of dots in the target image. Responses are z-standardized within participant to adjust for different response scales and detect outliers. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). The lagged extremities are the actual number of dots presented in the previous trials, rescaled to aid model conversion so that a 1-unit increase represents a step of a 1,000 dots. Random slopes for the polynomials of the standard extremity.

Appendix B. Interaction effects

On an exploratory base, we included interaction terms between the sequential effects (lagged response or lagged extremity, in separate models) with the 3 polynomial terms of the concurrent standard. Tables B1 (lagged response) and B2 (lagged extremity) show this for Study 2, whereas Tables B3–B6 refer to Study 3, separate by condition.

The results do not provide a consistent picture. In the no-anchor condition, linear concurrent standard extremity interacts with previous standard extremity in Study 3. This would suggest that higher estimates were given when either both were negative or both were positive. For the anchor condition and Study 2, no clear pattern emerges either. We thus refrained from interpreting these effects.

Table B1. *All fixed effects and related statistics from mixed model analysis including lagged response (Study 2).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	0.001 [−0.019; 0.020]	0.010	0.06	1220.96	.952
Dots	0.013 [0.007; 0.020]	0.003	3.99	4515.25	<.001
x	0.164 [0.145; 0.183]	0.010	17.13	732.74	<.001
x^2	−0.000 [−0.003; 0.003]	0.002	−0.05	741.96	.964
x^3	−0.005 [−0.006; −0.004]	0.001	−8.18	735.29	<.001
Lagged response (LR)	0.062 [0.046; 0.078]	0.008	7.54	1275.64	<.001
x × LR	0.004 [−0.007; 0.015]	0.006	0.74	1743.86	.462
x^2 × LR	0.002 [−0.000; 0.004]	0.001	1.85	749.83	.064
x^3 × LR	0.000 [−0.001; 0.001]	0.001	0.65	991.73	.513

Note: Dots are the actual number of dots in the target image. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). Random slopes for the standard extremity polynomials, lagged response and their interactions.

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

Table B2. *All fixed effects and related statistics from mixed model analysis including lagged extremity (Study 2).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	0.001 [−0.019; 0.021]	0.010	0.07	1244.77	.947
Dots	0.015 [0.008; 0.021]	0.003	4.50	4938.09	<.001
x	0.164 [0.145; 0.182]	0.010	16.99	731.64	<.001
x^2	−0.000 [−0.003; 0.003]	0.002	−0.01	741.91	.995
x^3	−0.005 [−0.007; −0.004]	0.001	−8.17	734.89	<.001
Lagged extremity (LX)	0.080 [−0.141; −0.019]	0.032	−2.55	27699.11	.011
$x \times$ LX	0.026 [−0.019; 0.071]	0.023	1.14	27865.89	.251
$x^2 \times$ LX	−0.020 [−0.029; −0.011]	0.004	−4.43	28074.69	<.001
$x^3 \times$ LX	0.002 [−0.002; 0.006]	0.002	0.88	28357.89	.381

Note: Dots are the actual number of dots in the target image. Responses are z-standardized within participant to adjust for different response scales and detect outliers. x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). The lagged extremities are the actual number of dots presented in the previous trials, rescaled to aid model conversion so that a 1-unit increase represents a step of a 1,000 dots. Random slopes are only modeled for the polynomials of the standard extremity due to singular fit if all were included.

Table B3. *All fixed effects and related statistics from mixed model analysis including lagged response in the no-anchor condition (Study 3).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.002 [−0.026; 0.023]	0.012	−0.13	722.58	.901
Dots	0.058 [0.030; 0.085]	0.014	4.14	9676.18	<.001
x	−0.093 [−0.116; −0.071]	0.001	−8.24	283.58	<.001
x^2	−0.003 [−0.007; 0.000]	0.002	−1.73	262.33	.086
x^3	0.000 [−0.002; 0.002]	0.001	0.20	283.47	.842
Lagged response (LR)	0.365 [0.331; 0.399]	0.017	21.23	509.68	<.001
$x \times$ LR	0.011 [−0.006; 0.028]	0.009	1.24	10082.95	.215
$x^2 \times$ LR	0.000 [−0.003; 0.004]	0.002	0.17	9851.57	.914
$x^3 \times$ LR	−0.001 [−0.002; 0.001]	0.001	−0.85	9935.16	.394

Note: x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100).

Table B4. *All fixed effects and related statistics from mixed model analysis including lagged response in the anchor condition (Study 3).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.009 [−0.040; 0.021]	0.015	−0.62	530.44	.534
Dots	0.099 [0.074; 0.124]	0.013	7.65	12066.34	<.001
x	0.134 [0.104; 0.164]	0.015	8.71	332.62	<.001
x^2	0.001 [−0.004; 0.006]	0.003	0.54	327.37	.590
x^3	−0.002 [−0.004; −0.000]	0.001	−2.41	335.91	.017
Lagged response (LR)	0.031 [0.007; 0.055]	0.012	2.58	932.45	.010
$x \times$ LR	−0.000 [−0.016; 0.015]	0.007	−0.01	12304.83	.995
$x^2 \times$ LR	0.004 [0.001; 0.007]	0.002	2.52	12374.98	.012
$x^3 \times$ LR	0.000 [−0.001; 0.002]	0.001	0.23	12514.40	.816

Note: x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100).

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

Table B5. *All fixed effects and related statistics from mixed model analysis including lagged extremity in the no-anchor condition (Study 3).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*p*
(Intercept)	−0.001 [−0.026; 0.025]	0.013	0.05	9913.84	.964
Dots	0.049 [0.018; 0.079]	0.016	3.09	10321.99	.002
x	−0.091 [−0.112; −0.070]	0.011	−8.45	2195.16	<.001
x^2	−0.003 [−0.007; 0.000]	0.002	−1.86	10022.45	.062
x^3	0.000 [−0.001; 0.002]	0.001	0.30	10086.90	.786
Lagged extremity (LX)	−0.351 [−0.464; −0.237]	0.058	−6.05	1102.47	<.001
$x \times$ LX	0.105 [0.025; 0.186]	0.041	2.57	5284.77	.010
$x^2 \times$ LX	−0.007 [−0.022; 0.009]	0.008	−0.84	10221.17	.401
$x^3 \times$ LX	−0.007 [−0.015; 0.000]	0.004	−1.90	10133.21	.058

Note: x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). The lagged extremities were rescaled to aid model conversion so that a 1-unit increase represents a step of a 1,000 dots.

Table B6. *All fixed effects and related statistics from mixed model analysis including lagged extremity in the anchor condition (Study 3).*

Fixed effects	*B* [95% CI]	SE	*t*	df	*P*
(Intercept)	−0.009 [−0.039; 0.021]	0.015	−0.58	528.87	.565
Dots	0.098 [0.073; 0.124]	0.013	7.60	12086.90	<.001
x	0.135 [0.104; 0.165]	0.015	8.73	332.82	<.001
x^2	0.001 [−0.004; 0.006]	0.003	0.50	327.21	.617
x^3	−0.003 [−0.004; −0.001]	0.001	−2.53	336.16	.012
Lagged extremity (LX)	−0.207 [−0.305; −0.108]	0.050	−4.11	556.35	<.001
$x \times$ LX	0.085 [0.017; 0.152]	0.034	2.47	401.15	.014
$x^2 \times$ LX	0.001 [−0.014; 0.016]	0.007	0.14	313.42	.886
$x^3 \times$ LX	−0.002 [−0.009; 0.004]	0.003	−0.63	313.09	.532

Note: x denotes relative standard extremity ranging from −4 to +4 (number of dots relative to neutral target divided by 100). The lagged extremities were rescaled to aid model conversion so that a 1-unit increase represents a step of a 1,000 dots.

A cautionary note also seems prudent with regard to interpreting sequential interaction in the current design as we must consider a potential confounding methodological artifact that arises when extreme anchors are investigated in a repeated measures design. Namely, as the current anchor becomes more extreme in relation to the target, by necessity, the numerical distance from the previous response simultaneously increases (at least on average). Spelled out, the anchor of 900 can have any distance to the previous anchor between 20 and 800. In contrast, an anchor of 500 can maximally have a distance of 400 to the previous anchor. Hence, the sequential distance between the current standard and previous response is a secondary potential influence that is inherently confounded with the extremity manipulation within a repeated measures design.

This can be easily illustrated by plotting the numerical distance between the previous response and the current standard for each trial using the data from the anchoring condition in Study 3, with negative scores indicating the previous response was less extreme or in the opposite direction from the target compared to the standard (Figure B1). The sequential distances show that as standards become more extreme, that is, they are further removed numerically from the target stimuli, they simultaneously become further removed from the previous self-generated response. Therefore, the potential effects of increasing sequential distance between the 2 numerical anchors and the standard extremity itself are fundamentally confounded within the design.

https://doi.org/10.1017/jdm.2023.31 Published online by Cambridge University Press

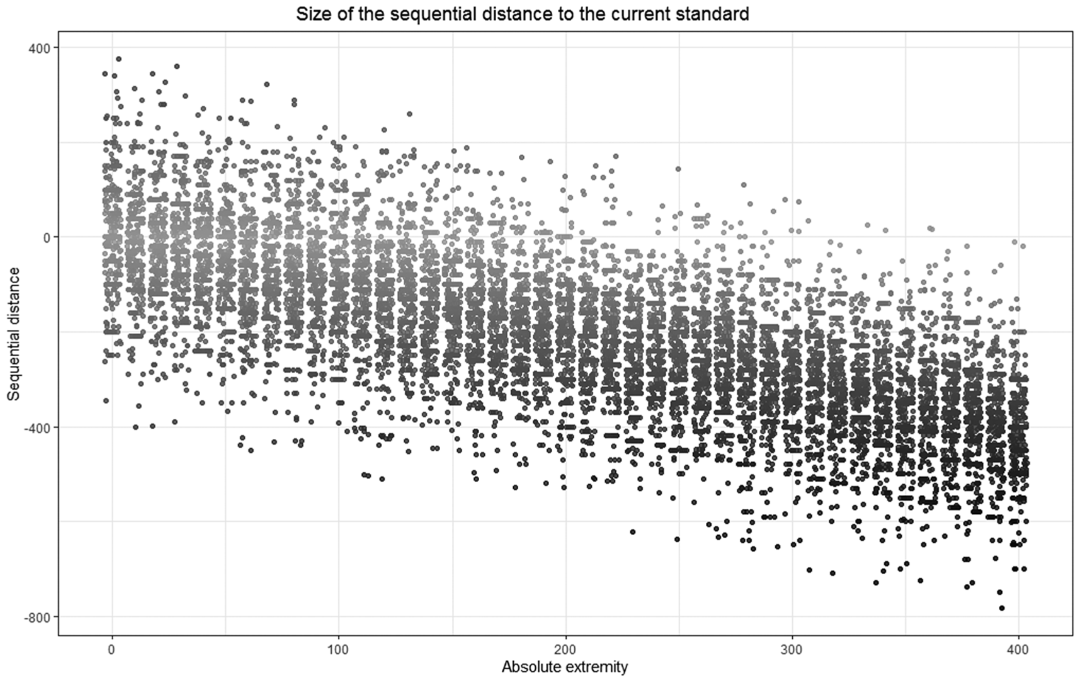

Figure B1. *Scatterplot of the discrepancy between responses in previous trials and the absolute extremity of the current anchor, with darker dots indicating larger sequential distances.*

Cite this article: Imhoff, R. and Barker, P. (2023). Connecting the dots: Nonlinear patterns in the presence of symbolic and nonsymbolic numerical standards. *Judgment and Decision Making*, e34. https://doi.org/10.1017/jdm.2023.31