Process Evidence Regarding the Anchoring and Adjustment Bias: New Methodologies and Non-Numeric Domains

Gaurav Jain, Dhananjay Nayakankuppam, Gary J Gaeth

Tippie College of Business, University of Iowa

Status: Under Review at Journal of Experimental Psychology - General

Author Note

Gaurav Jain, Department of Marketing, Tippie College of Business, University of Iowa; Dhananjay Nayakankuppam, Department of Marketing, Tippie College of Business, University of Iowa; Gary J Gaeth, Department of Marketing, Tippie College of Business, University of Iowa.

Correspondence concerning this article should be addressed to Gaurav Jain, Department of Marketing, Tippie College of Business, University of Iowa, Iowa City, IA 52240. Email: gaurav-jain@uiowa.edu

Abstract

Anchoring and Adjustment is a ubiquitous heuristic process in judgment and decision making. Although there is clear evidence that the anchor biases final estimates, there is disagreement about the process individuals use to arrive at the final estimate. To this end, with the help of four studies, we examine how individuals look for a response to a question asked, after being influenced by the anchor. Our results indicate a process wherein individuals search for an answer by testing plausible answers; the search being biased by the anchor question. We show that this search is dominated by adjustments to adjacent possible responses implicating a search process constrained by selective accessibility. Additionally, our use of relatively low-order-cognition anchoring contexts (e.g., perceptual anchoring) adds to the literature by demonstrating anchoring and adjustment bias in non-numeric domains.

Keywords: anchoring, adjustment, selective-accessibility, estimation, auditory-search, tactile-search

Process Evidence Regarding the Anchoring and Adjustment Bias

Since Tversky and Kahneman's (1974) landmark paper, the heuristics and biases paradigm has been a productive area of inquiry that has revealed a lot about human cognition and decision-making. Similar to optical illusions that illuminate processes underlying visual information processing systems, heuristics and biases yield insights into the systems that generate judgments.

Tversky and Kahneman identified 3 heuristics in their seminal paper – representativeness, availability and, anchoring & adjustment. Research into the first two has generated considerable convergence of opinion regarding mechanisms that govern them (see for example, Schwarz, 1998). The anchoring and adjustment heuristic, however, has remained more opaque (Gilbert 2002). The phenomenon is reliable and easy to demonstrate but there is less agreement, and more debate, about the underlying processes (Furnham & Boo 2011).

Robust Effects of Anchor on Judgment

Anchoring and adjustment has influenced judgments in various domains such as legal decisions (Englich & Mussweiler 2001), negotiations (Galinsky & Mussweiler 2002), consumer purchases (Wansink, Kent, & Hoch 1998), general trivia (Tversky & Kahneman 1974), population frequencies (Rottenstreich & Tversky 1997), social inferences (Tamir & Mitchell 2013); performance judgments (Thorsteinson et al 2008) and, probability estimates (Plous 1989). The effects of anchors have been shown using contextual (Epley & Gilovich 2001) as well as non-contextual anchors (Tversky & Kahnman 1974). While estimating weight, 20 pounds is a contextual anchor but the same number derived from the last two digits of the participants' telephone number is a non-contextual anchor. Studies have used external anchors provided by the experimenter (Jacowitz & Kahneman, 1995) as well as other sources such as a wheel of fortune (Tversky & Kahneman 1974) or a die (Englich et al 2006) or drawing a number from a hat (Cervone & Peake 1986), and other studies have shown the effects of internal anchors such as the last two digits of participants’ own social security numbers (Ariely, Lowrenstiein & Prelec 2003). Anchoring effects have been shown for implausible anchors as well. For instance, Strack and Mussweiler (1997) asked one group of participants whether Mahatma Gandhi died before or after the age of 9 and asked the other group of participants whether Mahatma Gandhi died before or after the age of 140. Though, both the anchors were clearly implausible, the guesses of the first group were still significantly lower than those of the second.

Proposed mechanisms

In Tversky and Kahneman’s original demonstration of this phenomenon, (see also Epley and Gilovich 2001, 2004, 2006; Epley, Keysar, Van Boven, & Gilovich 2004; Quattrone 1982), the anchor was seen to bias subsequent judgments, prompting the conclusion that adjustments away from the anchor were insufficient. However, Mussweiler and Strack (1999, 2000, 2001b) have advanced an account that would appear to subsume the anchoring and adjustment phenomenon under an accessibility account. Specifically, they argue that individuals engage in confirmatory testing for the anchor which makes information consistent with the anchor more available in the later judgment, resulting in the subsequent judgments assimilating towards the anchor. In other words, adjustment is not a necessary step in this account. Rather, an anchor biases the *construction* of the eventual judgment by making some information more accessible than others. Epley and Gilovich (2001, 2006) have argued that this is inconsistent with the influence of implausible, self-generated or patently random anchors and argue for the continued role of insufficient adjustment.

A third proposed account, often referred to as the ‘scale distortion theory’ (Frederick & Mochon 2012; Mochon & Frederick2013) is less relevant to our discussion here. It suggests that the ‘true answer’ looks extreme relative to the anchor that was just considered, prompting an adjustment towards the anchor. In other words, it proposes a contrast effect as driving the adjustment. While this differs in the proposed mechanics from the traditional adjustment paradigm, it still incorporates the idea of an adjustment and as such can be subsumed under the adjustment explanation for our current purposes.

The extant data cannot resolve this question. What is observed empirically in a typical anchoring and adjustment experiment are final estimates that are biased by answering an initial question that makes one particular anchor salient – that is, judgments assimilate towards an anchor. This phenomenon is perfectly consistent with the idea of insufficient adjustment from the anchor (or adjustment towards the anchor as proposed by the scale distortion account) but it is also consistent with the idea of an answer constructed from information, some of which was made preferentially accessible by answering the anchor question. Moderating variables could help to resolve this debate but finding one that does so unequivocally has been problematic. Creative accounts and retreats can be generated for each moderating variable affecting construction processes as well as adjustment processes. For example, it has been argued that implausible anchors generating anchoring effects is incompatible with the construction account. However, as Mussweiler and Strack (2001a) have argued, an implausible anchor may generate adjustment away from the anchor till a plausible anchor is reached but upon reaching the realm of plausibility, positive hypothesis testing still occurs and the final judgment is thus still constructed. This retains the adjustment component of the process, but it is relegated to explaining movement from implausibility and is divorced from the anchoring phenomenon.

We propose a more nuanced explanation. Specifically, we suggest that the anchoring and adjustment phenomenon represents searching an ordered space for a response. Given the anchor is usually rejected as the final judgment, and roundly rejected in the case of implausible anchors, the absence of a resultant contrast effect, rather than the observed assimilation effect, suggests that what is being observed need not (always) be the construction of a judgment. We posit that the process resembles searching for an answer in the related space, which is typically a number space. This search, however, can be biased by information rendered accessible in assessing the anchor, much as Mussweiler and Strack (1999, 2001b also, Chapman and Johnson 1999; Strack and Mussweiler 1997) have argued – see also Simmons, LeBoeuf & Nelson (2010) for an argument that the two competing theories can operate simultaneously. Thus, the search process resembles adjustment (as argued by Epley and Gilovich 2001, 2006) but the adjustment itself is influenced by the confirmatory testing of the anchor that Mussweiler and Strack (1999; 2001b) argue for. In other words, "search for a response" may be a more accurate description rather than adjustment – individuals are searching a hypothesis space in the wake of having rejected an anchor and the cognitive and affective processes created in dealing with the anchor constrains the search in predictable ways.

This begs the questions: How would search be biased? We argue that the bias would be in the range of search. Confirmatory hypothesis testing is likely to generate an affective response, that is, a 'feeling' of the answer being near at hand. So, we should see search constrained to the space close to the anchor, in what we term 'the adjacent possible'. If presented with an implausible anchor, individuals may take a large "leap" in the direction of an expected answer. Then this potential search region would contain responses that look reasonable compared to the implausible anchor one encountered a moment ago, it should again generate a 'feeling' of the answer being close at hand and constrain search. In any case, we should see individuals searching more intensively in 'the adjacent possible' space rather than more far-flung areas of the hypothesis space barring the first step away from an implausible response option. This is in contrast to the more normative rule of search, which is shown to be more efficient than a sequential search (Knuth 1973) wherein one would establish the boundaries and work inwards from the boundaries by halving the search space with each subsequent hypothesis.

Domain

Most studies of anchoring and adjustment rely upon the number line as the vehicle for response– that is, anchors and adjustments are numerical. While the number line has excellent properties that encourage continuous response metrics, it is subject to the criticism of relying upon higher order cognition. LeBouef and Shafir (2006) showed that participants' estimates of the length of a drawn line, the weight of a cup full of pennies, and the volume of music were biased by their earlier encounters with a line of a particular length, a cup of a particular weight and the volume of a sound heard initially. Oppenheimer, LeBoeuf, and Brewer (2007) find that individuals evaluated the temperature in Honolulu to be lower when they were made to draw short lines versus when they were made to draw long lines. But in these studies, the final estimates were still a numeric measurement of a target entity calling for higher order metric-based cognitions from the individuals making the estimates. That is, a perception still has to be cognitively represented and mapped onto a number line. The more complex the cognition, the more difficult it is to assess the mechanism of adjustment and thus we decided that there would be an advantage to detecting the adjustment process in lower-order judgments (judgments closer to perception). Doing this also permitted the use of more discreet observations of adjustment that are not subject to the usual caveats of self-reports or inferences, such as interruption of processing, veracity of introspective access, etc.

Predictions based on the alternative accounts

The traditional adjustment based perspective would argue that individuals would begin at the anchor and then start adjusting. Although silent on the specifics of adjustment, presumably individuals would 'take a number of equal sized steps' in a particular direction. Importantly, they would stop far too early, generating an 'insufficient adjustment' explanation for the anchoring phenomenon. The 'scale distortion' account would specify a move from the anchor to the answer followed by adjustment away from the answer towards the anchor.

On the other hand, the accessibility and construction based perspective would argue that no adjustment takes place but there would be positive hypothesis testing in answering the anchor question. The subsequent judgment would be constructed, the construction being dominated by information made accessible in answering the prior anchor question. So there should be no adjustment (unless presented with an implausible anchor, in which case, one would adjust to the realm of the plausible, after which there is still a positive hypothesis testing which takes place). As such, from the realm of the plausible (which is reached by the presentation of an anchor or by adjustment), there should be one move to the constructed answer. In other words, there should be one move from a plausible anchor or a series of moves from an implausible anchor followed by one move to the constructed final answer. In any case, these moves should be dominated by the *large last adjustment*.

Our account suggests that search is constrained by hypothesis confirmation and would be dominated by movement to the 'adjacent possible' which would result in a large number of small steps. The search would start with few large steps and rapidly converge to a large number of small steps whose size is of a diminishing, negative exponential form.

In other words, these testable differences in the predictions made by the three theorized processes motivate our decision to observe search processes – that is, how individuals consider potential final answers in the wake of having answered an anchor question.

Summary

In sum, we examine the search process in the anchoring & adjustment paradigm. We propose that selective accessibility of information consistent with the anchor (caused by the positive hypothesis testing of the anchor) biases the search process such that there is an increased search in 'the adjacent possible' space. The three different proposed processes generate testably different predictions. In order to eliminate the impact of higher order cognitions associated with numerical response scales, we examine anchoring and adjustment in relatively low-order-cognition domains and assess search surreptitiously. We report original empirical evidence in support of these hypotheses across four studies, all of which use alternative modalities and one of which employs eye tracking to add depth to our understanding of the adjustment process.

Study 1

Study 1 utilized the sense of touch as the stimulus and response domain; we investigated anchoring and adjustment mechanisms in the domain of textures. We utilized sandpapers of varying degrees of coarseness. This is a domain that is much further divorced from the number line when compared to line lengths etc. that have been used in the past. Participants were exposed to a particular reference grade (roughness) of sandpaper first. Subsequently, they were shown a continuum of sandpapers of various grades arranged on a white board and they were asked if one of them (the anchor, marked with a red pointer) was of the same coarseness as the reference they had seen earlier. They were then asked to estimate which of the various grades on the table was the same as the reference they had seen earlier. A camera was used to record their interactions with the various grades of sandpaper placed on the table. The recording was reviewed and scored at a later time.

Method

Individuals who were employees or students at a large university were invited to participate in the study through a university approved mass email. Sixty-eight individuals agreed to participate and were randomly assigned to one of two conditions in a single-factor, between-subjects design.

After signing a consent form, participants were shown a folder inside which was a piece of sandpaper. Instructions on the folder asked them not to open the folder themselves. They were requested to close their eyes and ask the research assistant to open the folder for them. They were then instructed to feel the sandpaper inside the folder with their eyes closed and were informed that they would be asked questions about the same later. They were asked to tell the research assistant to close the folder before opening their eyes. Subsequently, they proceeded to another room where 16 pieces of sandpaper of various levels of coarseness were arranged in decreasing order of coarseness on a poster board placed on a table. One of the 16 sandpapers had two pointers, one red and one blue, above it and served as the anchor. Participants were asked to judge whether this sandpaper with the pointers above it was of the same coarseness as the one they had felt earlier in the previous room. If so, they could leave the blue pointer where it was, else they were instructed to move the blue pointer to the sandpaper they judged to be the same as they had felt earlier. The red pointer could not be moved. Participants were told that they could touch all the sandpapers as many times as they wanted in the process of making their judgment. A camera was used to record their interactions with the various sandpapers.

The sixth sandpaper sample in the set was actually identical to the one they had felt earlier in the previous room. However, the red pointer which served as the anchor was placed over either the second sandpaper (which was coarser) or over the tenth sandpaper (which was finer). Initially the blue pointer was placed with the red pointer as well, but, as mentioned above, it could be moved by the participants. Thus this was a single factor (coarseness of the anchor) between-subjects design with two levels. Both anchors were four positions away from the "correct" answer.

Analysis and Results

The camera recording was transcribed to code for the order in which participants touched the various sandpapers before making their final judgment. For the purposes of the analysis, the coarsest sandpaper was coded 1 and the smoothest was coded 16.

We first considered the influence of the anchor on the final judgment and found that the basic anchoring and adjustment phenomenon was replicated. Participants who started with the coarser (no. 2) anchor judged a much coarser sandpaper (M = 5.15) to be similar to the target sandpaper they had touched earlier, compared to participants who started with the finer (no. 10) anchor (M = 7.14), ($F(1, 66) = 27.69$, $p < .001$, $\eta_p^2 = .296$).

As per our position, search would be constrained to the 'adjacent possible' indicating that the number of unit sized steps would be more than the number of non-unit sized steps. Moreover, participants would take bigger steps in the beginning converging to smaller steps as they closed in on a final response. Since the data are nested (steps taken nested within participants) and there are independent variables at both levels (e.g. anchor at the level of the individual, number of prior steps at the level of the steps taken), the use of multi-level modelling permits parsing the variance at both levels (participant level heterogeneity and step level) and utilize appropriate statistical tests and effects.

Essentially, we modelled whether the size of the nth step taken by a participant is influenced by the value of n. The size of the step taken was modelled as a function of the number of prior steps taken, with random effects at the individual level to allow for heterogeneity (variation in intercepts (or starting points) and slopes (the rate at which step size varied as a function of steps)). Table 1 provides the fit statistics for the various models that were tested. The best fitting model allowed for heterogeneity in intercepts and slopes (hardly surprising given individuals were given different anchor values). With this model formulation, analyses revealed that, consistent with our prediction, step size did decrease as a function of number of prior steps. Table 2 provides the results for the best fitting model and figure 1 provides the regression plot for the same.

<Insert tables 1 and 2 about here>

<Insert Figure 1 about here>

For purposes of exposition, we aggregated the data to the level of the individual and analyzed the frequency of steps of various sizes. An analysis of the frequency of steps of various sizes as a within-participant factor revealed that participants took many more unit-sized steps (i.e., steps to adjacent possibilities) ($M_{\text{number of steps of size 1}} = 17.69$) than all non-unit sized steps combined ($M_{\text{number of steps >size 1}} = 2.26$), ($F(1, 67) = 250.25$, $p < .001$, $\eta_p^2 = .789$). In other words, the adjustment process is dominated by sampling the adjacent possibilities rather than exploring the range of the scale.

Study 1 Discussion

Study 1 allows us to draw the following conclusions. 1) Anchoring and adjustment extends to non-numeric domains. Judgments of granularity of sand paper are more basic perceptions and suggest the anchoring and adjustment bias does not depend on a numeric representation of the stimulus or response scale, 2) There appears to be a process at work that is consistent with the idea of searching a hypothesis space (and the search process is biased by the anchor) within which individuals consider other plausible answers. 3) This process is dominated by movement to the adjacent responses that are plausible, rather than the more normatively suggested range contraction, i.e., establish outer bounds and work inwards in small steps rather than the incremental step in a particular direction approach that is seen here.

It should be noted that Study 1 showed a relatively large number of adjustments, most of which we have noted are unit sized. It is arguably almost costless to look at or feel a particular square. Also, participants may have been trying to familiarize themselves with the scale (to see if it was linear, etc.) which may have increased the number of unit sized steps. We attempt to address this issue more directly in the next two studies. In Study 2, we introduce the notion of a cost to the adjustment process and in Study 3, we eliminate the role of scale familiarization by permitting participants to get familiar with the scale beforehand.

Study 2

The procedure for Study 2 was substantially the same as Study 1, except that participants were now informed that while they could touch as many pieces of sandpaper as they wanted to, their task was to provide as accurate an answer as possible while minimizing the number of touches. It was explained to participants that touching a piece of sandpaper a second time after touching other sandpapers would be counted as a separate touch.

Method

Individuals who were employees or students at a large university were invited to participate in the study through a university approved mass email. Forty-four individuals agreed to participate and were randomly assigned to one of two conditions in a single-factor, between subjects design. The procedure was identical to Study 1, with the one change that participants were asked to be as accurate as possible while minimizing the number of touches.

Analysis and Results

We first assessed the influence of the anchor on the final judgment. The classic anchoring and adjustment effect was again replicated as participants who had the coarse anchor (sandpaper no. 2) chose a coarser sandpaper as their final answer (M = 5.52) compared to those who had a smooth anchor (sandpaper no. 10) (M = 6.90), ($F(1, 42) = 6.06$, $p = .018$, $\eta_p^2 = .126$).

With the use of multi-level modelling, for similar reasons as in study 1, we once again checked whether the size of steps decreases when individuals are closer to deciding on the final response. Table 3 provides the fit statistics for the various models that were tested. Again, the best fitting model allowed for heterogeneity in intercepts and slopes. With this model formulation, analyses revealed that, consistent with our prediction, step size did decrease as a function of number of prior steps. Table 4 provides the results for the best fitting model and figure 2 provides the regression plot for the same.

<Insert tables 3 and 4 about here>

<Insert Figure 2 about here>

To assess the adjustment process, we conducted an analysis of the frequency of various sized steps on the data aggregated to the level of the individual (as in Study 1), which revealed that participants used a significantly higher number of unit sized steps (M = 1.98) compared to all non-unit sized steps combined (M = 1.11), (F(1, 41) = 11.096, p = .002, η_p^2 = .205).

Study 2 Discussion

Note that the basic anchoring and adjustment phenomenon, and more specifically, the results of Study 1, were replicated even when the participants were asked to minimize the number of touches. It should also be noted that this was arguably a relatively subtle manipulation and participants may have utilized visual cues indicating coarseness, in addition to touches (and our measure of adjustment does not capture that). However, it could be argued that while we are interpreting 'touches' as evidence of search, it may partially indicate 'scale exploration'. As such, participants' final answer might still be constructed; study 3 addresses this concern.

Study 3

Study 3 was run for two main reasons. First, though the previous studies converge on demonstrating a preference for unit sized adjustments (i.e., exploring the adjacent response space as a potential answer rather than the distant responses), it can be argued that in Study 1 and 2 we are misinterpreting touches as an unambiguous indicator of a search process, while it may actually indicate scale exploration. It would be more convincing to completely eliminate the need for obtaining familiarity with the scale. To this end, we gave participants the opportunity to first familiarize themselves with the scale. A second reason for Study 3 was to demonstrate the phenomenon in an additional, numeric free, perceptual domain, namely in auditory perception.

A tonal scale was presented on the computer, consisting of 21 unlabeled buttons, each of which produced a pure sinusoidal tone (i.e., no timbre) of increasing frequency ranging from 900 Hz to 3900 Hz. Participants were allowed to familiarize themselves with the scale by pressing any of the buttons as many times as they wanted – the average number of buttons pressed during this familiarization period was 13.71 button presses, providing some face validity to the idea that individuals did expend effort familiarizing themselves with the scale. Subsequently, they heard a reference tone of 1600 Hz. They were then shown the tone scale but with one of the buttons enlarged and were instructed to press the enlarged button, which served as the anchor, and judge whether it matched the reference tone they had just heard. No other button at this stage was functional except the enlarged button. After this judgment, they were instructed to judge which button produced a tone that matched the reference tone they had heard earlier. At this point, all the buttons worked including the enlarged button and they could press any button any number of times.

Method

Individuals who were employees or students at a large university were invited to participate in the study through a university approved mass email. Sixty-eight participants agreed to take part in a single factor between subject design study where approximately half of the participants were provided the tonal scale with the fourth button enlarged and the remaining were provided the tonal scale with the fifteenth button enlarged. The correct answer was the ninth button.

Analysis and Results

The anchoring and adjustment phenomenon extended to the aural domain as well. The participants provided with a lower frequency anchor picked a significantly lower tone as the final response (M = 8.38) compared to participants provided with a higher frequency anchor (M = 10.31), ($F(1, 67) = 5.24$, $p = .025$, $\eta_p^2 = .074$).

The order in which participants pressed buttons, while deciding on the correct response, was recorded and analyzed. Every button tried by the participant counted as a step and the size of the step from the previous button was recorded.

To assess the adjustment process, we conducted an analysis of the frequency of various sized steps on the data aggregated to the level of the individual. Analysis of the frequency of steps of various sizes as a within-subject factor revealed a significantly larger number of unit sized steps (M = 6.19) compared to all non-unit sized steps combined (M = 1.58), ($F(1, 68) = 106.51$, $p < .001$, $\eta_p^2 = .614$).

The number of button presses in the scale familiarization stage ranged from 2 to 51. A further analysis using this measure as an indicator of the amount of scale familiarization yielded a significant interaction ($F(1,66) = 3.050$, $p = 0.005$, $\eta_p^2 = .112$) with the focal effect. The number of non-unit sized steps did not differ as a function of the amount of practice. The number of unit sized steps did differ as a function of practice. A spotlight analysis was performed at 1 SD above and below the mean. Examining the number of non-unit sized and unit sized steps 1SD above and below the mean amount of practice confirmed that there was no difference in the number of non-unit sized steps (1.72 at Mean – 1SD Vs 1.45 at M + 1SD; $t(66)= -.56$, $p > 0.250$), while there was a difference in the number of unit sized steps (5.09 at Mean – 1SD Vs

7.28 at M + 1SD; $t(66) = 2.53$, $p = 0.014$). Note, however, that this difference is in the opposite direction to the scale familiarization account. The scale familiarization account would argue that individuals take unit sized steps to familiarize themselves with the scale and thus there should be a decrease in the number of unit sized steps with increased practice. Said differently, it is the individuals with less practice (and thus, less familiarity with the scale) who would be inclined to explore the scale during the judgment phase of the task. What was observed, however, was exactly the opposite, with increased practice resulting in a greater number of unit sized steps (std. b = .297). Thus, the scale familiarization account is clearly incompatible with these data and is unlikely as an explanation for the results of the prior studies.

Study 3 Discussion

Results suggest that participants' search strategies are dominated by explorations in the adjacent possible of the hypothesis space. Note also that individuals explored the scale even though they had familiarized themselves with the scale. While we would not argue that individuals develop infallible auditory memory of the scale, the reluctance to explore further reaches of hypothesis space supports the existence of search in the 'adjacent possible'.

The design of Study 1 and 2 suffered from the limitation that what we interpreted as 'search' might have really just been participants familiarizing themselves with the scale in order to find out where their 'constructed' answer fits. Since participants were given a prior opportunity to familiarize themselves with the scale, Study 3 corrects this, and renders moot the issue of confabulating 'search' and 'scale familiarization'.

Study 4

Study 4 utilized shades of grey as the stimuli. This affords an added advantage, apart from being another domain that is not dependent on the number line, by allowing us to utilize eye-tracking as a process measure of the search process. While participants were performing the focal task, an eye tracker was utilized to track their eye fixations. Thus, we could track the sequence of eye fixations, which allowed us a covert way of examining whether there was a tendency for small adjustments to the ‘adjacent possible’ rather than large steps in exploring the hypothesis space.

Method

Individuals who were employees or students at a large university were invited to participate in the study through a university approved mass email. One hundred and three individuals agreed to participate and were randomly assigned to one of four “anchor” conditions in a single-factor, between-subjects design. 26 shades of grey from near-black (G1) to near white (G26) were utilized for the response scale. An ordered array of 26 shades of grey from near-white to near-black was created on the computer. Participants were shown a shade of grey that served as an anchor, and were asked if that was the shade of the moon. They were then shown all 26 shades and asked to pick the shade that, in their opinion, was the same as that of the moon. Shades of grey were developed by additive mixing of the three primary colors (R, G, B) such that (0, 0, 0) corresponded to black. The next lighter shade utilized (10, 10, 10) and so on to the lightest shade of grey (250, 250, 250).

After signing a consent form, participants were seated at a computer with an eye tracker. A five-point calibration was performed to calibrate the eye tracker to the particular individual’s eyes. Participants were then shown an anchor which, depending on the anchor condition they were in, was either G2 (almost black), G15 (dark grey), G19 (light grey) or G24 (near white). They were then asked whether, in their opinion, this was the same shade of grey as that of the moon. They were then shown all the 26 shades of grey and were asked to indicate which shade of grey they thought to be the shade of the moon.

Binocular eye movement data were recorded using a noninvasive Tobii X2-60 Eye Tracker that records the accuracy of eye movements to 0.4° of visual angle. The eye tracker measures visual scanning by computing the pupil-corneal reflection at a sampling rate of 60 Hz (i.e. 60 gaze data points are collected per second for each eye) based on the reflection of near-infrared light from the cornea and pupil.

Analysis and Results

The basic anchoring and adjustment phenomenon was replicated ($M_{anchor=G2} = 6.536$, $M_{anchor=G15} = 17.85$, $M_{anchor=G19} = 20.00$, $M_{anchor=G24} = 22.75$; $F(3, 99) = 6.98$, $p < .001$, $\eta_p^2 = .175$). Participants final ratings of the moon's shade were significantly impacted by the anchor they had been given.

The eye tracking data was utilized to examine the adjustment process. The pattern of fixations was coded to indicate how often individuals moved from each shade of grey to adjacent shades of grey or more distant shades of grey. Thus, each individual could take step sizes ranging from 1 (to the adjacent shade of grey) to a maximum step size of 25.

As per our position the number of unit sized steps would be more than the number of non-unit sized steps. Moreover, participants would take bigger steps in the beginning converging to smaller steps as they closed in on a final response. Again, since the data are nested (steps taken nested within participants) and there are independent variables at both levels (e.g. anchor at the level of the individual, number of prior steps at the level of the steps taken), the use of multi-level modelling permits parsing the variance at both levels (participant level heterogeneity and step level), utilizing the appropriate statistical tests and testing for the crucial cross-level interaction.

Just like in study 1 and 2, we modelled whether the size of the nth step taken by a participant was influenced by the value of n. The size of the step taken was modelled as a function of the number of prior steps taken, with random effects at the individual level to allow for variation in intercepts (or starting points) and slopes (the rate at which step size varied as a function of steps). Table 5 provides the fit statistics for the various models that were tested. The best fitting model allowed for heterogeneity in intercepts and slopes (hardly surprising given individuals were given different anchor values). With this model formulation, analyses revealed that, consistent with our prediction, step size did decrease as a function of number of prior steps. Table 6 provides the results for the best fitting model and figure 3 provides the regression plot for the same.

<Insert tables 5 and 6 about here>

<Insert Figure 3 about here>

For purposes of exposition, we aggregated the data to the level of the individual and analyzed the frequency of steps of various sizes. This analysis revealed that participants took significantly more unit sized steps (M = 7.52) compared to all larger than unit sized steps combined (M = 6.34) (F(1, 102) = 7.49, p = .007, η_p^2 = .068). This was not moderated by the anchor. Figures 4 and 5 provide the heat maps and the order of fixations across the different conditions.

Figure 4 illustrates the order in which a typical participant fixates on different parts of the scale when the anchor was G24, G19, G15 or, G2. All circles represent the fixations, their size represents the duration of the fixation and the numbers inside the circles represent the order of the fixation. It can be seen that the step sizes decreased as the participants proceeded with their search for the response. For instance, in Figure 4 panel (b), which illustrates the fixation made by the participant when the anchor was G 19, about 36 fixations are recorded. The large steps can be seen in initial fixations, primarily before fixation 15, but not for the latter half of fixations.

Figure 5, showing the heat maps, illustrates the duration of a typical participant's fixations on different parts of the grey scale when the anchor was G24, G19, G15 or, G2. In Fig. 2. Panels (a), (b) and (c), where the anchors were G24, G19 and, G15 respectively, it can be seen that the fixation duration was maximum near the anchor. However, as expected and as seen in panel (d), this was not the case when the anchor was G2 as it was a highly implausible anchor.

<Insert figures 4 and 5 about here>

The eye tracking data permit one more set of analyses to test a further aspect of our prediction. Note that we have argued that search in the 'adjacent possible' is due to a feeling of the answer being close at hand, thanks to the positive hypothesis testing one engages in. Such a feeling is likely to display itself in longer gaze duration at potential hypotheses as one assesses whether it could be the answer. Thus, we would predict that longer gaze duration at an area of interest (AOI) indicating greater consideration of the hypothesis should be followed by smaller movements (i.e., a greater tendency to move to the 'adjacent possible'). A multi-level model confirmed this to be the case – the gaze duration time at an AOI predicted the size of the next step (b = -0.002, F (1, 1367) = 8.36, $p < .001$). In other words, a long gaze duration (indicative of an increased feeling of the answer being close at hand) was followed by a greater tendency to move to the adjacent possible.

It should also be noted that the previous step size also predicted gaze duration (b = -14.17, F (1, 1406) = 25.12, p < .001), suggesting that smaller step sizes presumably served as a signal of a potentially likely answer, which should exacerbate positive hypothesis testing tendencies.

General Discussion

There has been disagreement about the mechanisms underlying anchoring and adjustment. While some have argued for the role of insufficient adjustment (Epley & Gilovich 2004, 2005; Simmons et al 2010; Tversky & Kahneman 1974), others have proposed the accessibility based construction account (Chapman & Johnson 1999; Mussweiler & Strack 1999, 2001a; Strack & Mussweiler 1997). No resolution has emerged. To this end, in a series of four studies, we obtain process evidence by examining how individuals look for a response to a question asked after being influenced by the anchor.

Our results indicate a search process dominated by adjustments to *adjacent possible* responses, implicating a search process constrained by selective accessibility. The predictions generated by the extant proposed process accounts do not appear capable of explaining the data reported here.

The selective accessibility account suggests that individuals construct the answer based on the information that is accessible to them – hence, the account suggests that there should be one large step towards the final answer (or steps from an implausible answer to the range of plausibility followed by one large step to the final response). As such, this is incompatible with the data reported here, where individuals appear to consider various alternatives before settling on an answer.

The insufficient adjustment account suggests that individuals take steps towards the final response but remains silent about the characteristics of the steps. Importantly, as per this account, individuals stop far too early, generating the ‘insufficient adjustment’ explanation for the anchoring phenomenon. This account would generate a prediction of individuals taking multiple steps, presumably equal sized, and then stopping due to the claimed effortful nature of the steps. In short, this account does predict steps but it is silent about the nature of these steps – nothing in the specifications of this theory would generate the ‘large number of small steps’ seen in the data. Thus, at the very least, these data help to specify this theory with more precision. More importantly, seeing that the prediction of ‘search in the adjacent possible’ is derived based upon the selective accessibility account, it helps to also explain the underlying process generating this tendency. The fact that both processes operate jointly also helps to explain why either process on its own is sufficient to generate an anchoring and adjustment effect – setting up conditions for either process is sufficient. The contribution of these data is in demonstrating that the two accounts also influence each other. The selective accessibility experience affects search and subsequent adjustment. In other words, while either account can individually explain the anchoring and adjustment phenomenon of final judgments assimilating to an anchor, the melding together also generates predictions about the nature of cognitive processes involved in getting from the anchor to the final judgment – that is, we obtain much greater specificity about the underlying psychological processes involved in the anchoring and adjustment heuristic. That is, arguably, the main contribution of these data.

Further, the methods and evidence add to the scarce literature examining anchoring and adjustment in non-numeric modalities. Finally, note that we have advanced an account based on a ‘feeling that the answer is near.’ While these data do not help to explicate the role of this feeling, the data are consistent with the ‘feelings as information’ account (Schwarz & Clore 1983). Current explorations of ‘feelings as information’ have been confined to the impact of feelings on judgments and evaluations. Our findings suggest that feelings can bias earlier stages of decision-making such as search processes.

One other point is worth mentioning. Study 4 demonstrated that longer gazes, indicating greater consideration of a potential answer, resulted in smaller movements in the subsequent step. Another analysis also revealed that smaller steps led to longer subsequent gazes, suggesting that an increased tendency to move to the adjacent possible might have served to create a sensation of getting closer to the potential answer. These two mechanisms create a mutually reinforcing feedback cycle (a plausible answer as evidenced in longer fixation times, resulting in smaller subsequent movements, which in turn increased the feeling of the next candidate answer being plausible). Such a mechanism seems a plausible candidate to support the idea of insufficient adjustment originally proposed in the heuristic. This warrants further scrutiny.

References

Ariely, D., Loewenstein, G., & Prelec, D. (2003). Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences. *Quarterly Journal of Economics*, 118, 73-105.

Cervone, D. & Peake, P. K. (1986). Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior, *Journal of Personality and Social Psychology*, 50, 492–501.

Chapman, G. B., & Johnson, E. J. (1999). Anchoring, activation, and the construction of values. *Organizational Behavior and Human Decision Processes*, 79, 1–39.

Englich, B. & Mussweiler, T. (2001). Sentencing under uncertainty: Anchoring effects in the courtroom, *Journal of Applied Social Psychology*, 31, 1535–51.

Englich, B., Mussweiler, T., & Strack, F. (2006). Playing dice with criminal sentences: The influence of irrelevant anchors on experts' judicial decision making. *Personality and Social Psychology Bulletin*, 32(2), 188-200.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the A&A heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, 12, 391–396.

Epley, N., & Gilovich, T. (2004). Are adjustments insufficient? *Personality and Social Psychology Bulletin*, 30(4), 447–460.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological science*, 17(4), 311-318.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology*, 87, 327–339.

Frederick, S. W., & Mochon, D. (2012). A scale distortion theory of anchoring. *Journal of Experimental Psychology: General*, 141(1), 124.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40(1), 35-42.

Galinsky, A. & Mussweiler, T. (2002). First offers as anchors: The role of perspective-taking and negotiator focus, *Journal of Personality and Social Psychology*, 81, 657–69.

Gilbert, D. T. (2002). Inferential correction. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 167–184). Cambridge, England: Cambridge University Press.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21, 1161-1166.

Knuth, D. E. (1973). Sorting and searching. (The art of computer programming, vol. 3) *Addison-Wesley*. Reading, MA, 551-575.

LeBoeuf, R. A., & Shafir, E. (2006). The long and short of it: Physical anchoring effects. *Journal of Behavioral Decision Making*, 19(4), 393-406.

Mochon, D., & Frederick, S. (2013). Anchoring in sequential judgments. *Organizational Behavior and Human Decision Processes*, 122(1), 69-79.

Mussweiler, T., & Strack, F. (1999). Hypothesis-consistent testing and semantic priming in the anchoring paradigm: A selective accessibility model. *Journal of Experimental Social Psychology*, 35, 136–164.

Mussweiler, T. & Strack, F. (2000). The use of category and exemplar knowledge in the solution of anchoring tasks, *Journal of Personality and Social Psychology*, 78, 1038–52.

Mussweiler, T., & Strack, F. (2001a). The semantics of anchoring. *Organizational Behavior and Human Decision Processes*, 86, 234–255.

Mussweiler, T., & Strack, F. (2001b). Considering the impossible: Explaining the effects of implausible anchors. *Social Cognition*, 19, 145–160.

Oppenheimer, D. M., LeBoeuf, R. A., & Brewer, N. T. (2007). Anchors aweigh: A demonstration of cross-modality anchoring and magnitude priming. *Cognition*, 106(1), 13–26.

Plous, S. (1989). Thinking the unthinkable: The effect of anchoring on likelihood estimates of nuclear war, *Journal of Applied Social Psychology*, 19, 67–91.

Quattrone, G. A. (1982). Overattribution and unit formation: When behavior engulfs the person. *Journal of Personality and Social Psychology*, 42, 593–607.

Rottenstreich, Y. & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in Support Theory, *Psychological Review*, 104, 406–15.

Schwarz, N. (1998). Accessible content and accessibility experiences: The interplay of declarative and experiential information in judgment. *Personality and Social Psychology Review*, 2(2), 87-99.

Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: informative and directive functions of affective states. *Journal of personality and social psychology*, 45(3), 513.

Simmons, J. P., LeBoeuf, R. A., & Nelson, L. D. (2010). The effect of accuracy motivation on A&A: do people adjust from provided anchors? *Journal of personality and social psychology*, 99(6), 917.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, 73, 437–446.

Tamir, D. I., & Mitchell, J. P. (2013). Anchoring and adjustment during social inferences. *Journal of Experimental Psychology: General*, 142(1), 151.

Thorsteinson, T. J., Breier, J., Atwell, A., Hamilton, C., & Privette, M. (2008). Anchoring effects on performance judgments. *Organizational Behavior and Human Decision Processes*, 107(1), 29-40.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124 –1131.

Wansink, B., Kent, R. J., & Hoch, S. J. (1998). An anchoring and adjustment model of purchase quantity decisions, *Journal of Marketing Research*, 35, 71–81.

Tables

Table 1
Fit Statistics for the models tested for study 1

Model	- 2 Res Log Likelihood
Base Model with random intercept	5024.9
STSIZE= β_1 STORD with random intercept	5029.6
STSIZE= β_1 STORD + β_2 STORD^2 with random intercept	5038.8
STSIZE= β_1 STORD with random intercept and linear term	5023.1
STSIZE= β_1 STORD + β_2 STORD^2 with random intercept, linear and quadratic term	5020.0

STSIZE: Size of a particular step
STORD: Order of a particular step

Table 2
Results for the best fitting model for study 1

Effect	**Estimate**	**Standard Error**	**DF**	**t Value**	**Pr > |t|**
Intercept	1.8265	0.1137	67	16.06	< .001
STORD	-0.0468	0.0178	67	-2.63	.011
$STORD^2$	0.0011	0.0001	67	2.10	.040

STORD: Order of a particular step

Table 3
Fit Statistics for the models for study 2

Model	**- 2 Res Log Likelihood**
Base Model with random intercept	451.5
STSIZE= β_1 STORD with random intercept	446.3
STSIZE= β_1 STORD + β_2 STORD^2 with random intercept	447.4
STSIZE= β_1 STORD with random intercept and linear term	441.4
STSIZE= β1 STORD + β2 STORD2 with random intercept, linear and quadratic term	Did Not Converge

STSIZE: Size of a particular step
STORD: Order of a particular step

Table 4
Results for the best fitting model for study 2

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	2.5633	0.2433	37	10.53	<.001
STORD	-0.1844	0.0591	30	-3.12	0.004

STORD: Order of a particular step

Table 5
Fit Statistics for the models tested for study 4

Model	- 2 Res Log Likelihood
Base Model with random intercept	8095.3
STSIZE= β_1 STORD with random intercept	8084.5
STSIZE= β_1 STORD + β_2 STORD^2 with random intercept	8091.8
STSIZE= β_1 STORD with random intercepts and slopes	8082.8
STSIZE= β_1 STORD + β_2 STORD^2 with random intercepts and slopes	8046.7

STSIZE: Size of a particular step
STORD: Order of a particular step

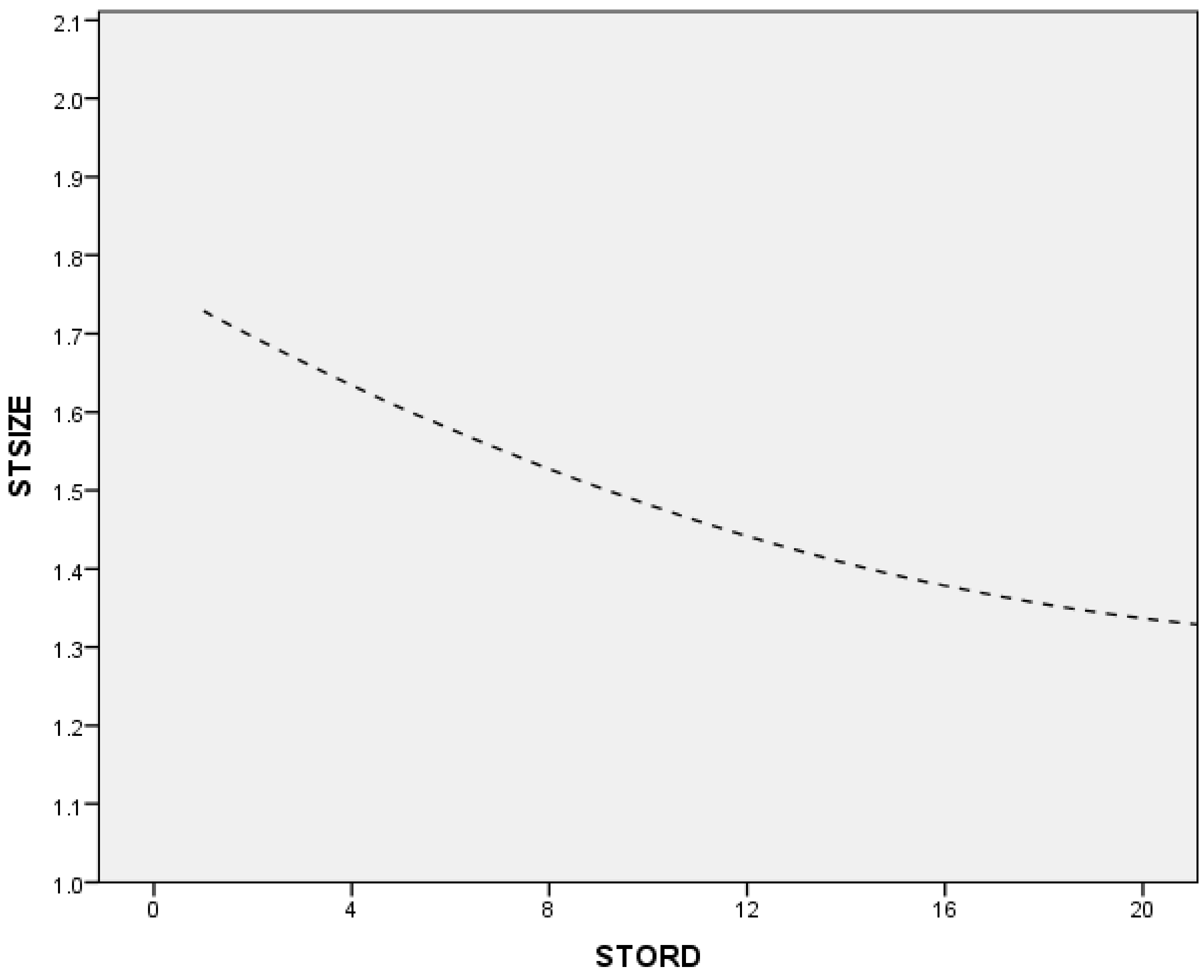
Table 6
Results for the best fitting model for study 4

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	5.303	0.427	102	12.43	< .001
STORD	-0.415	0.094	99	-4.40	< .001
$STORD^2$	0.017	0.005	99	3.63	< .001

STORD: Order of a particular step

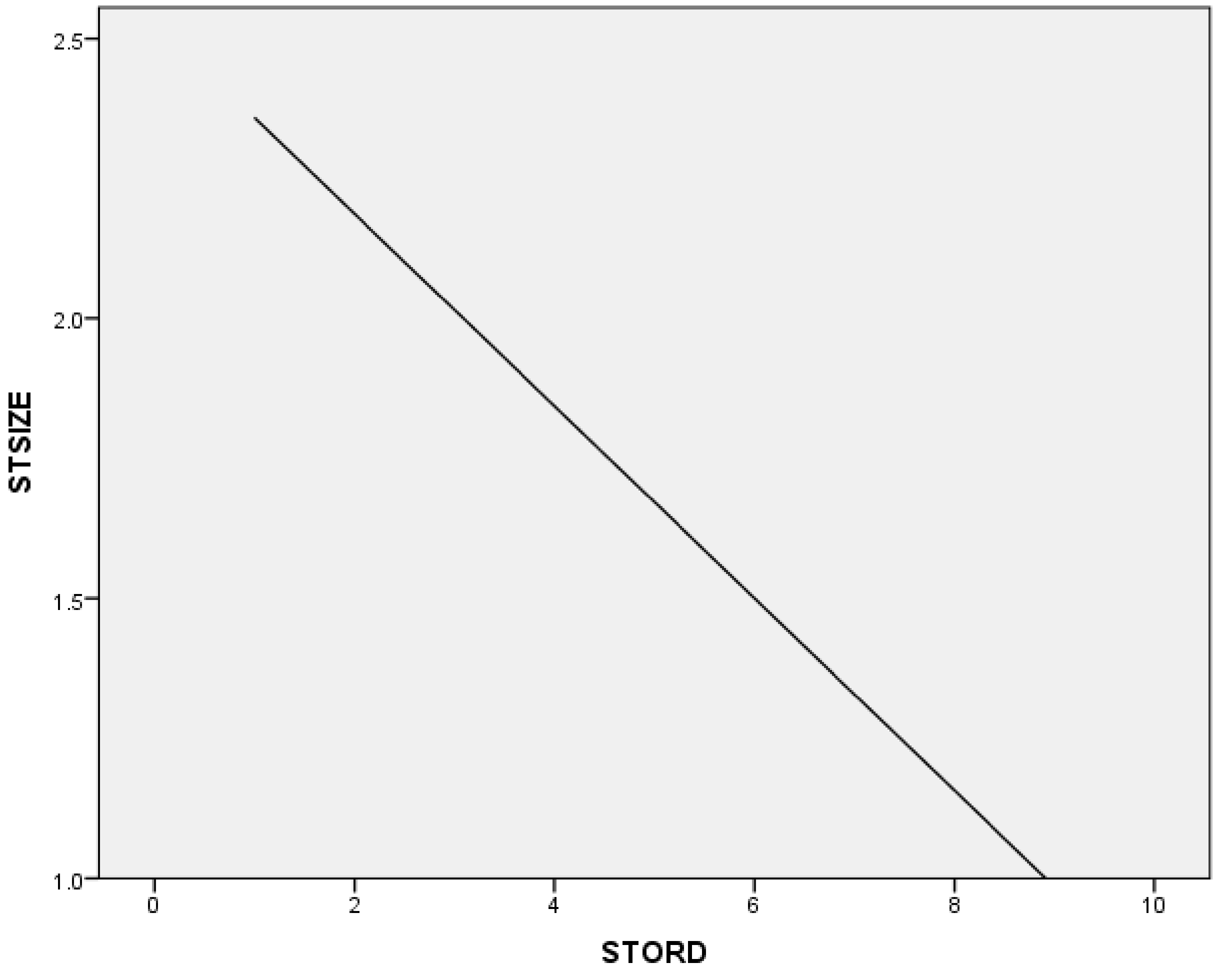
Figures

Fig.1. Regression plot for the best fitting model in study 1.



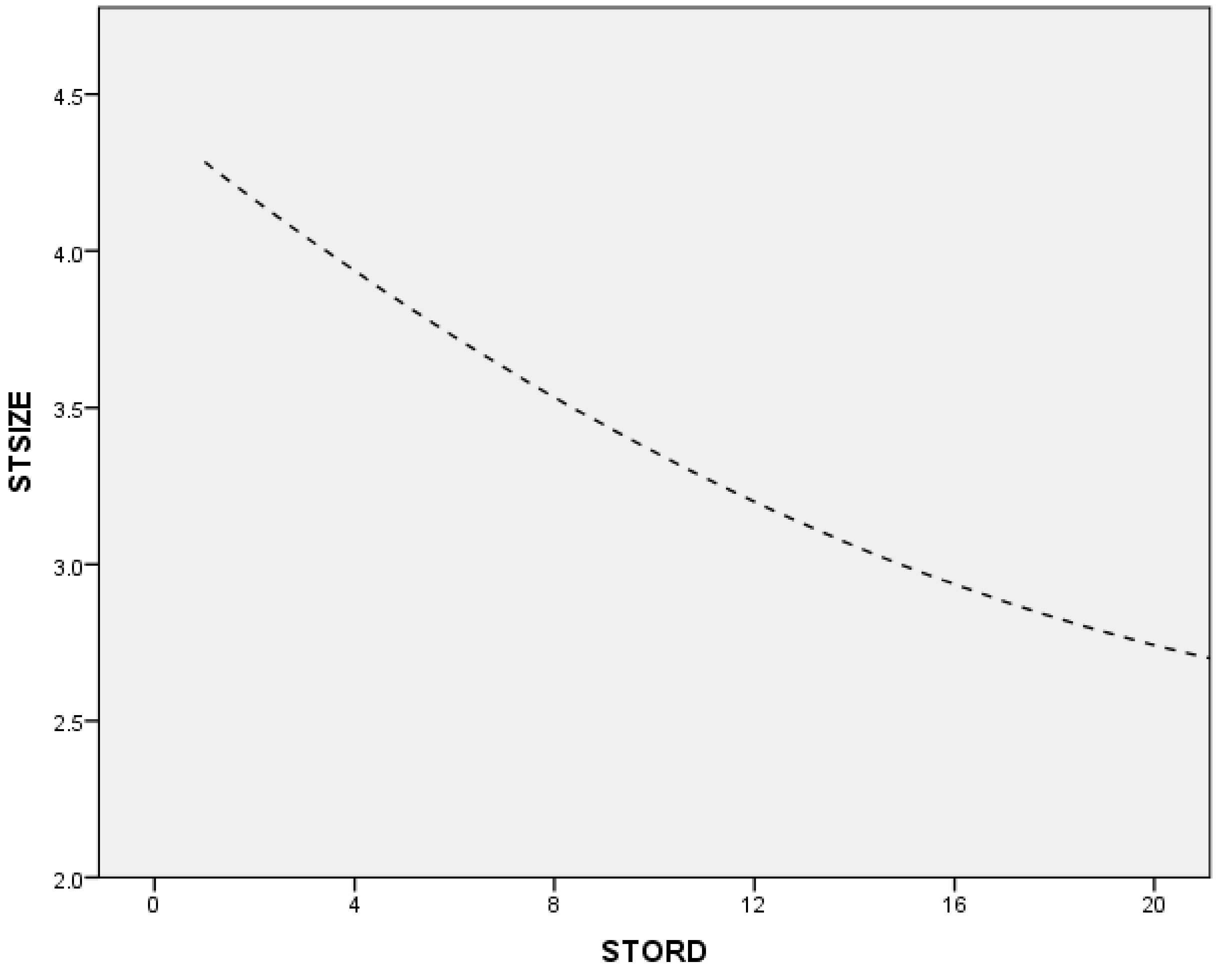
STSIZE: Size of a particular step
STORD: Order of a particular step

Fig.2. Regression plot for the best fitting model in study 2.



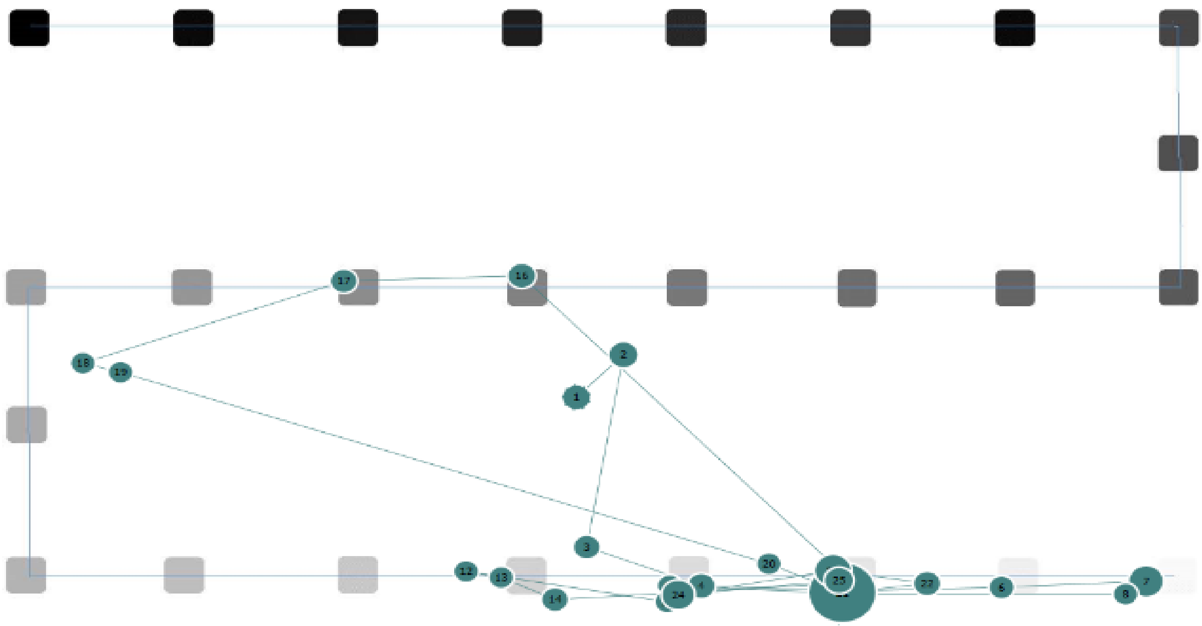
STSIZE: Size of a particular step
STORD: Order of a particular step

Fig.3. Regression plot for the best fitting model in study 4.

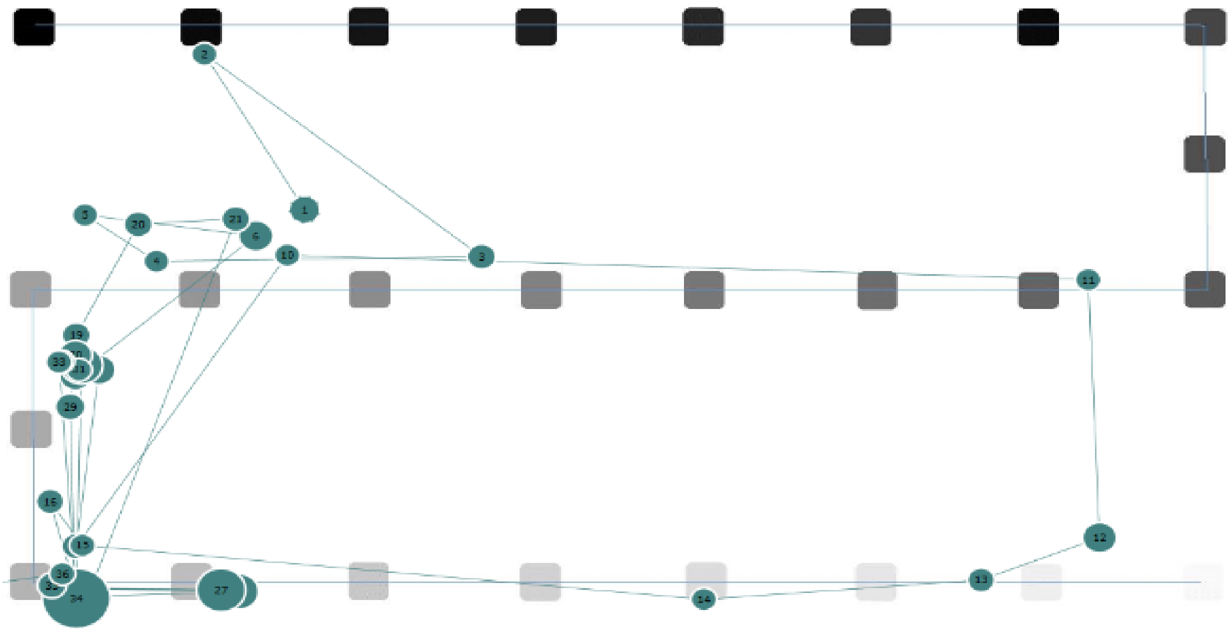


STSIZE: Size of a particular step
STORD: Order of a particular step

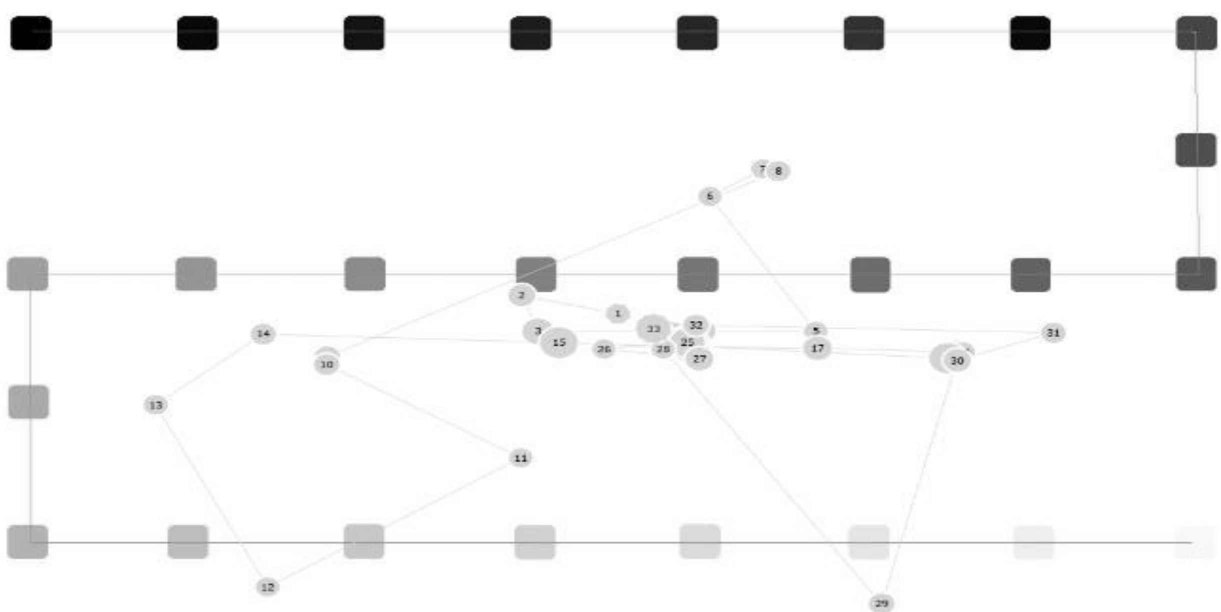
Fig. 4. Order of fixations for a typical participant in different anchor conditions.
Panel (a) illustrates the order of fixations for a typical participant when the anchor was G24. The numbers inside the circles represent the order of the particular fixation and the size of the circle represents the duration of that particular fixation. Panels (b), (c) and (d) illustrate the same when the anchors were G19, G15 and G2 respectively.



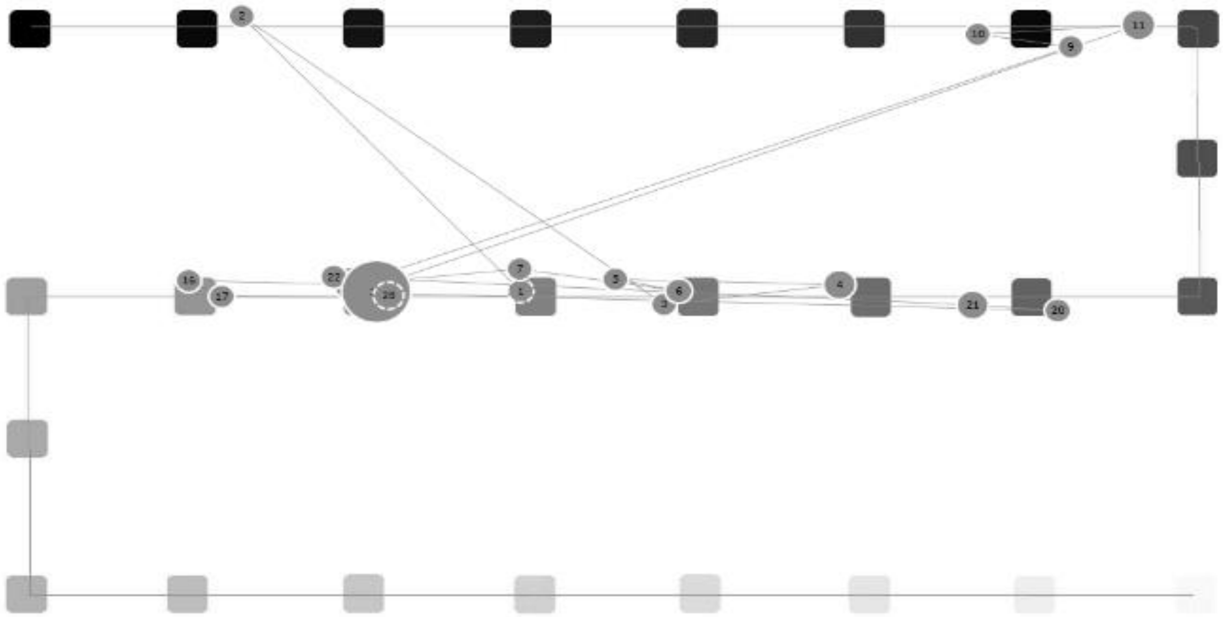
(a)



(b)



(c)



(d)

Fig. 5. Heat maps illustrating the duration of fixations on a particular shade of grey for a typical participant in different anchor conditions.
Panel (a) shows the heat map illustrating the total duration of fixations for a typical participant when the anchor was G24. A shade towards red stands for higher duration than yellow which in turn stands for higher duration than green. Panels (b), (c) and (d) illustrate the same when the anchors were G19, G15 and G2 respectively.