Comparing webcam-based eyetracking with normal eyetracking in a value-based decision-making task

Thomas Rijpstra (t.j.rijpstra@student.rug.nl)
Thesis
University of Groningen, The Netherlands
s1712969

Abstract

Everyday we are presented with simple choices such as the choice between an apple or a candy-bar as a snack. Value-based decision-making research tries to investigate which factors contribute to the decision-making process for such choices. Previous studies found that visual attention plays an important part in value-based decision-making, and that this process can be quantitatively and accurately modeled using computational models such as the Drift-Diffusion Model. Using eye-gaze data, the internal process of the value comparison of options can be accurately modeled for both binary and trinary choices. However, these studies all rely on usage of an expensive standalone eyetracker for collecting eye-gaze data and are therefore conducted in a clinical setting, limiting the applications of this type of research.

In this thesis a methodology was developed for using a commodity webcam to extract information about eye-gaze using the OpenFace framework and inferring the intrinsic value of options in a value-based decision-making task. An experiment was conducted (n = 17) in which binary simple choice was investigated using Hierarchical Drift-Diffusion Models augmented with gaze information simultaneously collected from a webcam and eyetracker. By exposing the connection between gaze and intrinsic value, predictions can be made about upcoming decisions using only information collected through the webcam, thus providing an intrinsic measure of item preference using commonly available hardware.

We concluded that value-based decision-making research is possible using a webcam instead of an eyetracker, and that similar conclusions are reached irregardless of the gaze collection method. The methodology developed in this thesis enables researchers to conduct value-based decision-making research through online questionnaires (provided that a webcam is available) allowing for larger sample sizes while maintaining low research costs.

Introduction

Most behavioral studies are carried out on Western university students participating for student credit or a small financial compensation, but a growing body of evidence suggests that findings from this population do not necessarily apply to humanity as a whole (Henrich, Heine, & Norenzayan, 2010a). This heavily researched population is described as WEIRD (Western, educated, industrialized, rich and democratic) and while research on it has yielded many interesting findings, its validity as a fair representation of the generic human population is challenged by recent developments in various fields suggesting differences in basic cognitive and motivational processes between industrialized and small-scale societies and western and non-western societies. For example, substantial perceptual differences have been shown among populations by Segall, Campbell, and Herskovits (1966) using a Müller-Lyer illusion. They found that members of industrialized societies perceived a greater illusion than small-scale societies, due to the visual system having adapted more to recurrent features (e.g. "carpentered corners") in the local visual environment. This shows that even a basic cognitive process such as visual perception is susceptible to environmental changes and that particular cultural evolutionary trajectories can substantially affect cognitive processes (Henrich, Heine, & Norenzayan, 2010b). Another example of these differences is that both analytical and holistic reasoning strategies are employed differently in western and non-western societies: while both strategies are available in all normal adults, western societies tend to rely more on analytical reasoning (i.e. perceiving an object as separate from its context and focusing on its attributes) in contrast to non-western societies which rely more on holistic reasoning (i.e. perceiving an object in its context or field as a whole and focusing on its relations to other objects in that field) (Nisbett, 2010; Peng & Nisbett, 1999). This reliance on different reasoning strategies in turn affects cognitive processes such as attention: Chua, Boland, and Nisbett (2005) found that gaze patterns differ for Americans and East-Asians when attending a visual scene. Americans gaze at focal objects longer than East Asians, who in turn gaze at the background more than Americans. These cognitive and motivational differences could have large consequences in the field of decision making, as both cognitive and motivational processes are common elements in theories on decision making. As most research on decision making is carried out on the WEIRD population due to economical and practical reasons, it could very well be that many insights in this field cannot be properly generalized to other populations.

One specific insight in the area of decision making concerns how the brain makes simple choices, such as the choice between an apple or a Mars candy-bar as a snack. Most researchers agree that the brain first assigns a value to all available options, and selects the best choice by comparing these values. The intrinsic value attributed to each option can be affected by many factors, originating internally (e.g. blood-sugar level, craving for chocolate) and externally (e.g. other people's choices, price of the item). The perceived difference between intrinsic values of options can be implicitly (e.g. by investigating gaze-patterns) or explicitly (e.g. by asking subjects to rate each item on a fixed scale) measured and is expressed as 'value preference' (Rangel, Camerer, & Montague, 2008; Rangel & Hare, 2010).

A relatively simple paradigm for investigating the influence of value preference on decision making processes is the value-based decision-making task. In this task, subjects are first asked to rate a selection of items (e.g. consumer products such as candy-bars) on their subjective value. Next they are presented with a series of choices between the rated items, varying in difficulty based on the difference in rating between the presented items. For each trial, the chosen item, time to choose (response time) and gaze information (using an eyetracker) is collected to provide insight into the underlying decision making process. The collected behavioral data (e.g. response times, eye-gaze) can be used as input for computational models such as the Drift Diffusion Model (DDM) (Ratcliff, 1978) to test competing theories of decision making on their predictive power.

Drift Diffusion Models enable decision making researchers to parametrize the decision making process by offering an estimation of how evidence accumulation, represented by drift rate parameter (α), and response caution, represented by decision threshold parameter (θ), contribute to the decision process. These parameters are acquired by modeling individual accuracy and response times (RT) distributions. A higher drift rate represents greater accumulation of evidence and results in better accuracy and shorter RT, while a higher decision threshold represents increased response caution and results in better accuracy and longer RT (Ratcliff & McKoon, 2008). Ultimately, these simplified parameters provide insight in the underlying cognitive processes associated with value-based decision making.

Many studies have investigated the effects of value preferences in simple choice using (parts of) the paradigm described above (Armel, Beaumel, & Rangel, 2008; Krajbich, Armel, & Rangel, 2010; Krajbich & Rangel, 2011). These studies found that visual attention plays an important part of in the decision making process, and that this process can be quantitatively and accurately modeled using computational models such as the DDM. Manipulation of visual attention positively affects choice likelihood for appetitive items, while it negatively affects choice likelihood for aversive items (Armel et al., 2008). Using eye-gaze data, the internal process of the value comparison of options can be accurately modeled for both binary (Krajbich et al., 2010) as trinary (Krajbich & Rangel, 2011) choices.

Using the techniques described above, Cavanagh, Wiecki, Kochar, and Frank (2014) found that both eye gaze and pupil dilation reflect latent dissociated decision processes (i.e. both reflect internal decision processes which do not directly influence each other). They found that eye gaze dwell time, additionally to perceived stimuli value (Krajbich et al., 2010), has a direct influence on the rate of evidence accumulation as modeled by the drift-rate parameter in the DDM, while changes in pupil dilation can be used as a predictor for the decision threshold parameter, which represents the minimal required evidence for making a choice. These findings show that valuable information about the decision making process is reflected in the eyes, and suggest that it should be possible to infer the intrinsic value of options by collecting eye-gaze data, thus paving the way for an implicit method of measuring item preference.

Still, the findings by Cavanagh et al. (2014); Krajbich et al. (2010) are based on WEIRD subjects and therefore suffers from the same over-generalization of its conclusions to humans as a whole. To circumvent this, researchers would have to target a more representable set of subjects without damaging the economic feasibility of their research (i.e. the recruitment process should be roughly as expensive as it is now). While the internet could be used to target a much wider range of individuals, it is paramount that the collected data are valid. However, the collection of eye-gaze data poses another problem: since very few people have an eyetracking device at home, another method for collecting eye-gaze data using more commonly available hardware should be used when conducting decision making research using the internet.

Various endeavors have been made to use commonly available webcams as an input source for eyetracking (Sewell & Komogortsev, 2010; Skovsgaard, Agustin, Johansen, Hansen, & Tall, 2011; San Agustin et al., 2010). San Agustin et al. (2010) showed that eye-typing (i.e. accurately entering characters into a computer using gaze and eyeblinks) is possible using a low-cost (infrared) webcam. Skovsgaard et al. (2011) assessed and compared the performance of a similar setup against two commercial eyetrackers and found that their webcam-based gaze-tracker has a performance comparable to the two commercial devices. While these studies show the promise of webcam-based eyetrackers as a means of cheaply collecting eye-gaze data, both studies required modifications to the webcam such as adding an infrared light which thereby prohibit any large-scale real-world applications.

Both normal commercial eyetrackers as the experiments by San Agustin et al. (2010); Skovsgaard et al. (2011) use infrared-cameras with two (internal or external) infrared light-sources to produce a glint reflection on the eyes. Using these glints, the exact position and angle of the pupils can be triangulated as the distance between the two light-sources is known. Furthermore, the usage of an infrared camera produces a less noisy image than produced by a commodity webcam which captures images in red, blue and green channels on a higher resolution and lower framerate. In other words, a commodity webcam collects less relevant data in the same timeframe as infrared cameras and cannot provide a frame of reference to a known spatial location.

Instead of relying on heuristics, such as the position and angle of the pupils in relation to glints, a commodity webcam-based eyetracking solution should rely on other techniques which are more resilient to noisy data to enable its usage in a large-scale real-world application.

Sewell and Komogortsev (2010) tried to tackle this challenge by developing a webcam-based eyetracker which employs a simple perceptron-style neural network for estimating gaze coordinates from a greyscale image of the pupil. In their set-up, a RGB-image from an unmodified webcam is preprocessed so that a greyscale image of the pupil remains. This greyscale image is used as input to the neural network which is trained using supervised learning. Target outputs are provided by a calibration round in which subjects look in turn at 48 different targets on the screen. After this training, subjects confirmed the responsiveness of the webcam-based eyetrackers. While Sewell and Komogortsev (2010) have shown the feasibility of a webcam-based eyetracking powered by a neural network, they still rely on a preprocessing of webcam images using heuristics (e.g. pupil location using head tracking) which exposes a weak-point in their methodology to noisy data. Indeed, Sewell and Komogortsev (2010) reported numerous calibration failures due to a failure to find the pupil in the preprocessing step. They suggest numerous ways to improve performance: image preprocessing should be improved by employing a more robust method of pupil location which corrects for head position and rotation; the neural network should be pre-trained on a large set of subjects and fine-tuned during calibration to decrease training time; and a higher resolution camera should be used to improve the quality of input for the neural network.

Since the research by Sewell and Komogortsev (2010), enormous progress has been made in the field of machine learning through the discovery of deep learning (Krizhevsky, Sutskever, & Hinton, 2012). As a result, many steps necessary for the development of a webcam-based eyetracker are now relatively easy, such as reliably recognizing faces directly from images using open-source frameworks such as OpenFace (Baltrušaitis, Robinson, & Morency, 2016) and feature reduction and detection using techniques like Deep-Convolutional Neural Networks. This could offer a solution for the problems encountered by Sewell and Komogortsev (2010), as feature reduction and detection is handled by the convolutional neural network, thus eliminating the need for the researcher to identify and implement preprocessing algorithms.

However, the usage of deep learning comes with a price: the computational resources required for training are much higher than for a simple perceptron network and a substantial dataset with enough labelled training data is needed (Goodfellow, Bengio, & Courville, 2016). Thus the approach employed by Sewell and Komogortsev (2010), in which incremental training is done images as they come available, is not possible.

Fortunately, the recently developed OpenFace framework (Baltrušaitis et al., 2016) is equipped which various pre-trained deep-neural networks which enable both facial landmark detection and tracking (Baltrušaitis, Robinson, & Morency, 2013) and eye gaze tracking (Wood et al., 2015).

In this study we will develop a methodology for using a commodity webcam to extract information about eye-gazes using the OpenFace framework, and use its output to infer the intrinsic value of options in a value-based decision-making task as described by Krajbich et al. (2010); Cavanagh et al. (2014). By exposing this connection between gaze and intrinsic value, predictions can be made about upcoming decisions using only information collected through the webcam, thus providing an intrinsic measure of item preference using commonly available hardware.

We will assess the reliability of this methodology by simultaneously collecting data from the webcam and a commercial eyetracker and comparing the single-trial influence of each on the DDM parameters decision threshold and drift rate.

The experiment will be carried out in an up-to-date internet browser so that results from this study can be interpreted in the same context as an internet-conducted experiment, thus paving the way for decision making studies conducted through the internet while collecting eyetracking data and maintaining economic feasibility. This should enable researchers to target a more representative set of subjects than the commonly used members of the WEIRD population.

Method

Participants

20 right-handed, healthy individuals with normal ($n = 16$) or corrected-to-normal ($n = 4$) vision participated in the study. Participants were recruited from the social circle of the experimenter (e.g. friends, family, colleagues) and received no compensation for participation. Informed consent was obtained digitally prior to the start of the experiment. Three participants were excluded due to near-chance performance on easy trials($M_{\text{score}} = 55.9\% \pm 0.9\%$), an indication of insufficient task comprehension or motivation. In total 17 data sets were analyzed ($M_{\text{age}} = 30$; $SD_{\text{age}} = 8.7$; range 23 - 57; female 9; male 8). All participants had attended university-level education.

Design

A within-subject design was used in which binary-choice on rated stimuli was compared to value-preference from a separate rating task. The decision process was modelled using reaction times and visual fixations to determine what influenced the decision process. Visual fixations were acquired using both a commercial eyetracker and a novel webcam-based approach.

Participants performed a rating task in which they rated 50 stimuli, followed by a binary value-based decision task (VBDT) in which the choice difficulty was manipulated using values from the rating task. During the decision task participants were recorded with the integrated webcam, and information regarding eye-movements was collected using a commercial eyetracker. Dependent variables were the number of fixations and fixation durations collected per area of interest (AOI), reaction times (RT), choice accuracy and the parameters of the hierarchical drift diffusion models fitted to the behavioral data.

Stimuli

The stimuli consisted of 50 images of food items, categorized as snacks ($n = 43$) or fruits ($n = 7$). The images were full-color on a white background, 400 by 400px and were obtained from a national supermarket website without permission. All stimuli are listed in appendix A.

Experiment

The study was conducted at various locations, such as a living room and office, and all instructions were solely communicated through on-screen text to simulate the conditions respondents to an online questionnaire generally encounter.

The researcher did not interact with participants during the experiment, other than calibrating the eyetracker before the decision task. The general flow of the experiment is shown in figure 1a.

All participants completed the experiment in less than 20 minutes.

Questionnaire

Participants first completed a questionnaire regarding age, gender, education, dexterity and vision, followed by the Eating Restraint Scale (ERS) (Polivy, Herman, & Warsh, 1978) from which the subscales Concern for Dieting (CD) and Weight Fluctuation (WF) were used to check for participants with abnormal eating habits (van Strien, Breteler, & Ouwens, 2002; van Strien, Herman, Engels, Larsen, & van Leeuwe, 2007), as abnormal eating habits could potentially influence the decision making process in ways out of scope of this thesis. On the WF-subscale, three participants ($n = 3$) scored more than three SD higher than normal weighted females in the study by van Strien et al. (2007), indicating a potential abnormal eating habit. One of these participants was excluded from analysis due to near-chance performance on easy trials. Almost all participants ($n = 16$) scored more than three SD higher on the CD-subscale than normal and overweight females in van Strien et al. (2007). This should be kept in mind while interpreting results from this thesis, as this could potentially indicate abnormal eating habits in almost all participants.

Rating Task

The ERS was followed by the Rating Task. Participants were asked to rate 50 randomly presented stimuli using a scale from -100 (hate it) to 100 (love it). The center of the scale (0) was labeled as neutral and was used to distinguish between aversive (rated < 0) and preferred (rated > 0) items.

Decision Task

After rating all stimuli, participants were presented with 250 trials in which two stimuli were presented. Participants were instructed to choose their preferred item in each trial. A choice was made by pressing the 'Q' or 'P' key for left or right item respectively. Participants were instructed to use both index fingers for pressing the response keys. After 250 trials the experiment ended. All participants completed the task within 12 minutes ($M_{time\ on\ task} = 480s \pm 90s$).

Participants were instructed to answer as fast and accurately as possible.

Trials Figure 1b shows a graphical representation of one trial. Each trial initiated with a centered fixation cross for 500ms, followed by a blank screen for 50ms. Afterwards, two stimuli were presented equally spaced from the horizontal center of the screen until the participant's response, for a maximum of 3000ms. If the participant failed to respond in that time frame, the trial was labeled as invalid and feedback was presented for 2000ms urging to "answer more quickly". A trial was marked correct if the chosen stimulus was the highest rated of the presented pair, and incorrect otherwise. Trials were both stimuli had the exact same rating were marked as neutral.

Conditions Choice difficulty was manipulated by presenting stimuli-pairs with a similar (i.e. difficult) or dissimilar (i.e. easy) rating as supplied by the participant in the rating task. Trials where the stimuli-pair rating difference was less than 10 points were labeled as difficult, other trials as easy. Overall, participants were presented with more easy than difficult trials ($P_{\text{easy}} = 58.2\% \pm 7.9\%$).

Trials were also categorized on stimuli-aversion (i.e. do the stimuli have a positive or negative rating) and on stimuli-similarity (i.e. are both stimuli from the same stimuli category). Overall, participants were presented with more non-aversive ($P_{\text{aversive}} = 24.6\% \pm 13.3\%$) and similar ($P_{\text{similar}} = 76.6\% \pm 5.3\%$) trials.

Materials

The study was performed on a single Apple 13" Macbook Pro Retina (late 2014 model) placed on a pedestal to put the screen at eye height. A generic USB keyboard and mouse were connected to enable user input.

Eyetracker Eyetracking data was collected during the decision task using an EyetechDS VT2 Mini (EyetechDS, 2014), placed below the screen. The sampling rate was 80Hz. Calibration occurred before the decision task using the QuickGlance software bundled in the Quicklink SDK (EyetechDS, 2016).

Webcam Webcam data was collected during the decision task using the integrated webcam (720p Facetime HD). The recorded videos are full color with a resolution of 1280 x 720px at 30 frames per second.

Analysis

RStudio (RStudio, 2012) was used to carry out all statistical analyses, with the exception of the HDDM analyses. Linear mixed effect models were created and fitted using the 'lme4' package (Bates, Mächler, Bolker, & Walker, 2015) and the corresponding p-values where computed using the 'lmerTest' package (Kuznetsova, Bruun Brockhoff, & Haubo Bojesen Christensen, 2016). Degrees of freedom for fitted LME models were calculated with the 'pbkrtest' package (Halekoh & Højsgaard, 2014) using Kenward-Roger approximation (Kenward & Roger, 2009).

All non-responses were omitted from analysis.

Eyetracking Eyetracking fixations were calculated using the 'saccades' package (von der Malsburg, 2015). X and Y coordinates where represented as percentages of the screen.

The ‘saccades‘ package can be tuned using the lambda parameter. Best results (in reference to Krajbich and Rangel (2011)) were obtained using $\lambda = 5$. Figure 14a shows a graphical representation of eyetracker fixations during the experiment.

Webcam Webcam images were analyzed using OpenFace (Baltrušaitis et al., 2016), from which gaze- (Wood et al., 2015) and head tracking-data (Baltrušaitis et al., 2013) were obtained. Gaze vectors ($openface_{x,y,z}$) of both eyes (left, right) obtained through OpenFace, were converted to screen pixel coordinates ($gaze_x$ and $gaze_y$ for X and Y pixel coordinate respectively) using the webcam's focal length (f_{base}) and sensor size (cx and cy) as shown in equations 1 till 4 using values shown in table 1.

$$gaze_x = \frac{openface_{x,left} + openface_{x,right} * fx}{openface_{z,left} + openface_{z,right}} + cx \quad (1)$$

$$gaze_y = \frac{openface_{y,left} + openface_{y,right} * fx}{openface_{z,left} + openface_{z,right}} + cy \quad (2)$$

$$fx = -2 * cx * \frac{f_{base}}{resolution_x} \quad (3)$$

$$fy = -2 * cy * \frac{f_{base}}{resolution_y} \quad (4)$$

Table 1: Values used in webcam gaze vector transformation.

cx	400
cy	300
f_{base}	150
$resolution_x$	1280
$resolution_y$	720

Screen pixel coordinates were then converted to screen percentages for comparison with eyetracker data. After normalization, clusters in the webcam coordinates were mapped to clusters found in the eyetracker data by applying $kmeans(2)$ and $kmeans(1)$ in x and y coordinates respectively.

Figure 2c shows both webcam and eyetracker coordinates on the x axis after these transformations. While both measurements show similarity, in the webcam coordinates a pattern is clearly visible. To correct for this pattern, two linear models were fitted on the webcam coordinates using head tracking data (rotation and translation coordinates in x, y, z) obtained through OpenFace as shown in equation 5 and equation 6.

$$gaze_x \sim \beta_{Rx,Ry,Rz} * Rotation + \beta_{Tx,Ty,Tz} * Translation \quad (5)$$

$$gaze_y \sim \beta_{Rx,Ry} * Rotation + \beta_{Tx,Ty,Tz} * Translation \quad (6)$$

The residuals of these linear models are hypothesized to be a more accurate representation of eye gaze coordinates. Indeed, figure 2f clearly shows that both webcam and eyetracker data are more similar.

Similar to the eyetracker, fixations were calculated using the ‘saccades‘ package. Best results were obtained using $\lambda = 3.25$.

Figure 14b shows a graphical representation of webcam fixations during the experiment.

Webcam/Eyetracker Alignment Visual inspection showed that webcam and eyetracking data was not fully aligned. This was traced back to an implementation error which failed to log the timestamps during webcam retrieval.

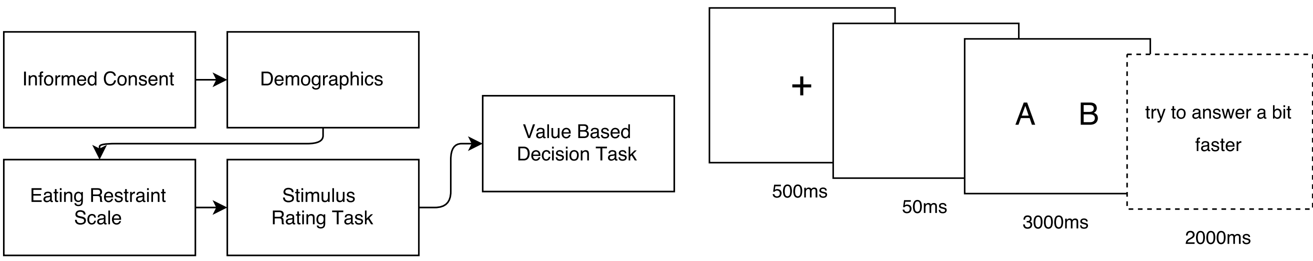
For every participant, this was corrected by right-aligning the webcam data to the eyetracker data using equation 7, fitting a linear model in the form of equation 8 and using the calculated parameters to transform the webcam data using equation 9.

$$t_{webcam} = t_{webcam} + (max(t_{eyetracker}) - max(t_{webcam})) \quad (7)$$

$$t_{eyetracker} \sim t_{webcam} \quad (8)$$

$$t_{webcam} = \beta_1 * t_{webcam} + \beta_0 \quad (9)$$

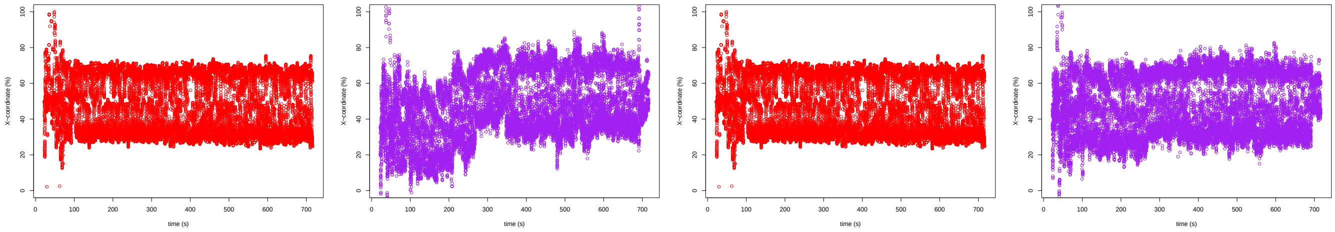
Direct correlation between eyetracker fixations and pure webcam significantly improved with this transformation ($\rho_{before}(3654) = 0.58$, $p < 0.001$; $\rho_{after}(3654) = 0.63$, $p < 0.001$; $p_{\rho_{after} > \rho_{before}} < 0.001$), as well as inter-trial gaze difference (i.e. duration difference between left and right fixations during trials) did ($\rho_{before}(3654) = 0.04$, $p < 0.05$;



(a) Flow of the experiment

(b) Flow of one trial in Value Based Decision Task.

Figure 1: Schematic representation of both the experimental set-up and a single trial in the Value Based Decision Task.



(a) Eyetracker (b) Webcam (d) Eyetracker (e) Webcam

(c) Before head movement correction. (f) After head movement correction.

Figure 2: X-coordinates over time for subject (223), before and after head movement correction.

$\rho_{after}(3654) = 0.28$, $p < 0.001$; $p_{\rho_{after} > \rho_{before}} < 0.001$).

Drift Diffusion Models In Drift Diffusion Models (DDM's) the decision process is modelled as a noisy process which accumulates evidence for one of two competing choices. While DDM's sport various parameters, the most influential are drift rate (i.e. the rate of evidence accumulation) and decision threshold (i.e. the amount of evidence needed to reach a decision) (Ratcliff, 1978).

In this study the Hierarchical Drift Diffusion Model (HDDM) toolbox (Wiecki, Sofer, & Frank, 2013) is used, as it allows for easy inclusion of physiological parameters in model fits and produces "more accurate DDM parameter estimates for individual and groups, particularly given low trial numbers or when assessing coefficients between psychophysiological measures and behavior" (Cavanagh et al., 2014).

Following Cavanagh et al. (2014) models were fitted by drawing 5,000 samples from the posterior, of which the first 200 were discarded as burn-in. 5% of participants were randomly regarded as outliers on each sampling iteration. Regression coefficients were estimated to investigate single trial variations in psychophysiological measures (e.g. eye gaze from eyetracker or webcam), psychological measures (e.g. proportional difference in rating of presented stimuli) and other influences (e.g. are presented stimuli from the same category or different) on model parameters (drift rate, decision threshold).

Multiple HDDM's were fitted to investigate which model of decision making best explains the experimental data. Following Cavanagh et al. (2014), we compare the independent model of decision making (i.e. choice is influenced by gaze and perceived stimulus value independently), attentional DDM (aDDM; i.e. choice is influenced by the interaction of gaze and perceived stimulus value) and hybrid aDDM (i.e. choice is influenced by the interaction of gaze and perceived stimulus value and gaze independently) (Krajbich et al., 2010; Krajbich & Rangel, 2011; Krajbich, Lu, Camerer, & Rangel, 2012).

The best models were chosen using the deviance information criterion (DIC) and visual inspection of the estimated regression coefficients distributions. The DIC value decreases when model likelihood increases and when model complexity decreases, thus when comparing models, the one with a lower DIC value is favored.

Specifications of fitted models Two models were fitted to determine the influence of proportional rating difference and trial similarity (i.e. are presented stimuli from the same category) on both drift rate and decision threshold. Six models (three using eyetracking data, three using webcam data) were fitted to determine the influence of eye gaze (independent, dependent or both independent and dependent on rating difference) on drift rate. These eight fitted models are further described in table 6. Relevant regression coefficient plots are shown in figures 21, 22 and 23.

Results

Rating Task

Stimuli had a mean rating of 19.6 $\pm$ 1.5 on a scale of -100 till 100, indicating that the overall stimuli-set was perceived as appetitive. As ratings are categorized as appetitive or aversive, it is interesting to look at both categories separately. Appetitive stimuli had a mean rating of 47.0 $\pm$ 1.1 while aversive stimuli had a mean rating of -41.6 ± 1.8. The overall frequencies of aversive and appetitive stimuli are shown in table 2, as well as frequencies per stimuli category (fruit or candy). From table 2 it follows that the appetitive/aversive ratio differs between fruit and candy stimuli: fruit stimuli are almost exclusively rated as appetitive, while candy has much more variation. Indeed, figure 3 shows that positively rated fruit stimuli (61.6 $\pm$ 3.0) are rated significantly higher ($F(1,708) = 57.7$, $p <$ .00) than positively rated candy stimuli (42.7 $\pm$ 1.6).

Decision Task

Neutral trials (i.e. trials were both stimuli are rated identically) were excluded because accuracy cannot be determined on these trials ($n = 274$, 6.72%). Table 3 shows the frequencies of trials per condition, trial aversiveness and trial stimuli similarity. Due to its sparse occurrence in the data, aversive dissimilar trials ($n = 16$, 0.4%) are excluded from further analysis.

Response Times On average, participants took 1404 ± 10 ms to respond. As trial difficulty was the main manipulation, its effect on respond times is shown in figure 4: respond times increase with trial difficulty. Figure 5 shows how trial conditions affects RT: participants take significantly ($\chi^2(1) = 8.13, p < .01$) longer to reach a decision on hard trials ($1436 \pm 22ms, b = 2.61, SE = 0.012, t = 2.24$) than easy trials ($1385 \pm 16$).

We then asked if RTs are affected by how long participants have been doing the decision task. We distinguish two possible effects: RTs are affected by time-on-task or RTs are affected by trial novelty. Indeed, RTs significantly decrease ($\chi^2(1) = 15.80, p < .0001, b = -3.1 * 10^{-4}, SE = 7.7 * 10^{-5}, t = -3.99$) when time-on-task increases, but they are not affected by trial novelty ($\chi^2(1) = 1.79, p = 0.18$, *ns*). The effect of time-on-task did not differ between conditions ($\chi^2(1) = 0.08, p = 0.77$, *ns*).

Next we asked if aversiveness influenced RTs (i.e. does the decision process differ for aversive stimuli and appetitive stimuli). Figure 6 shows how how aversiveness affects RTs: participants take significantly longer ($\chi^2(1) = 44.17, p < .0001$) to reach a decision on aversive trials ($1496 \pm 27ms$) than on appetitive trials ($1377 \pm 15ms$, $b = -0.080, SE = 0.014, t = -5.53$).

Finally we investigated if RTs are influenced by trial stimuli similarity (i.e. does the decision process differ when stimuli are from the same category or different categories). Figure 7 shows that there is indeed an influence of trial stimuli similarity: participants are significantly faster ($\chi^2(1) = 7.74, p < .01$) on dissimilar (e.g. fruit-candy) trials ($1333 \pm 27ms$) than on similar trials ($1426 \pm 15ms$, $b = 0.038, SE = 0.014, t = 2.78$). The effect of trial stimuli similarity does not differ between conditions ($\chi^2(1) = 0.08, p = 0.77$, *ns*).

The best fitting LME model on $log(RT)$ is shown in table 4.

Accuracy A trial is labeled as accurate when the highest rated stimulus is chosen. Figure 8 shows that the placement of a stimuli (left/right) does not have any effect on the participants decision ($F(6, 4058.9) = 1.31, p = .25$, *ns*). However, figure 9 shows that decisions are affected by the difference in rating for the presented stimuli: the chance that the

Table 2: Frequencies of fruit- and candy-stimuli rated as aversive or appetitive.

Stimuli	Aversive	Appetitive
Fruit	6	162
Candy	301	581
Total	307	743

Table 3: Frequencies of trials per condition, trial aversiveness and trial stimuli similarity.

	condition			
	easy		hard	
	aversive	appetitive	aversive	appetitive
dissimilar	10 (0.3%)	499 (13.1%)	6 (0.1%)	368 (9.7%)
similar	649 (17.0%)	1236 (32.5%)	251 (6.6%)	786 (20.7%)
	2394 (62.9%)		1411 (37.1%)	

Note. All dissimilar, aversive trials are excluded from analysis due to their sparse occurrence. Percentages shown are in relation to all trials.

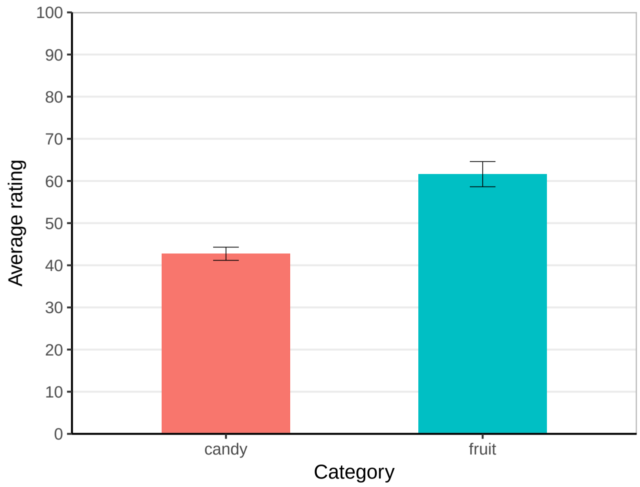


Figure 3: Average appetitiveness rating of positively rated fruit and candy stimuli. Error bars denote standard error of the mean.

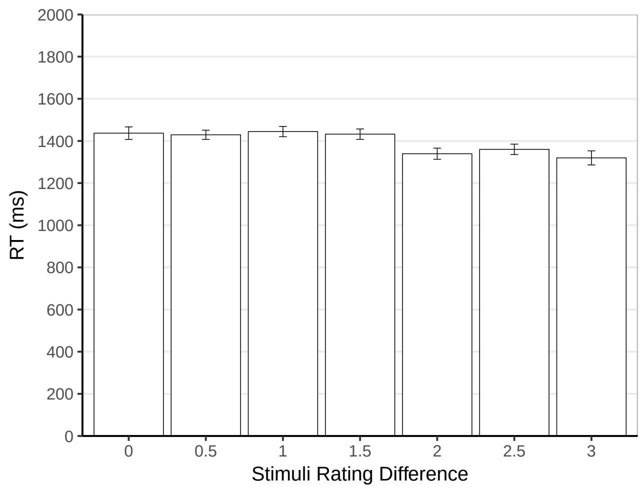


Figure 4: Response Times by trial difficulty, defined as the relative difference in rating between presented stimuli. The higher the difference, the easier the trial. Error bars denote SE of the mean.

left or right item is chosen is affected by the score of that item in reference the other item.

On average, participants made an accurate decision (i.e. the highest rated stimulus of the presented pair is chosen) on 64.6% $\pm$ 0.9% of the trials. Figure 10 shows how accuracy is affected by trial condition. Participants made significantly more accurate decisions ($\chi^2(1) = 43.06, p < .0001$) on easy (68.4% $\pm$ 1.3%) than on hard (58.2% $\pm$ 1.9%, $b = -0.106, SE = 0.016, t = -6.58$) trials.

Additional to condition, trials can be categorized as aversive (i.e. choose between two negatively rated items) or appetitive. Figure 11 shows that accuracy is affected by stimuli aversiveness: accuracy on aversive items (60.1% $\pm$ 2.3%) is significantly lower ($\chi^2(1) = 9.06, p < .001$) than on appetitive items (65.8% $\pm$ 1.2%, $b = 0.074, SE = 0.019, t = 3.83$). This effect does not differ per condition ($\chi^2(1) = 2.71, p = 0.10$, *ns*).

Trials can also be categorized on the basis of stimuli similarity (i.e. are the stimuli presented from the same stimuli category). A fruit-candy trial is marked as dissimilar, while both candy-candy and fruit-fruit trials are marked as similar. Figure 12 shows how trial similarity affects accuracy: participants are significantly ($\chi^2(1) = 4.87, p < .05$) less accurate on dissimilar trials (63.8% $\pm$ 2.3%) than on similar trials (64.7% $\pm$ 1.3%, $b = 0.042, SE = 0.019, t = 2.20$).

As no feedback is presented on accuracy of the choice, we

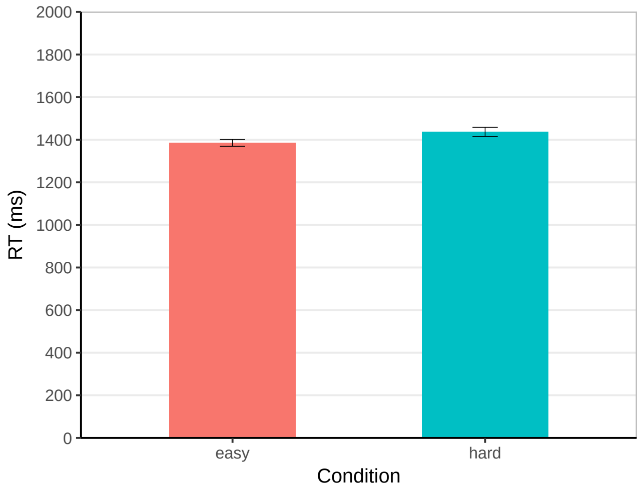


Figure 5: Response Times by condition. Participants need more time to reach a decision on hard trials. Error bars denote SE of the mean.

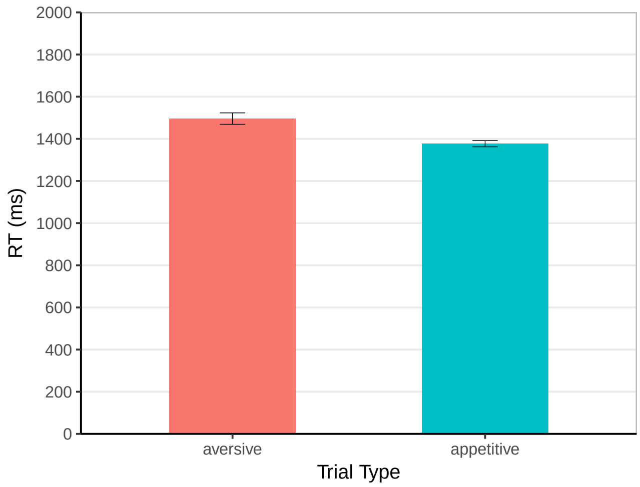


Figure 6: Response Times by trial appetitiveness. Participants take longer to reach a decision on aversive trials. Error bars denote SE of the mean.

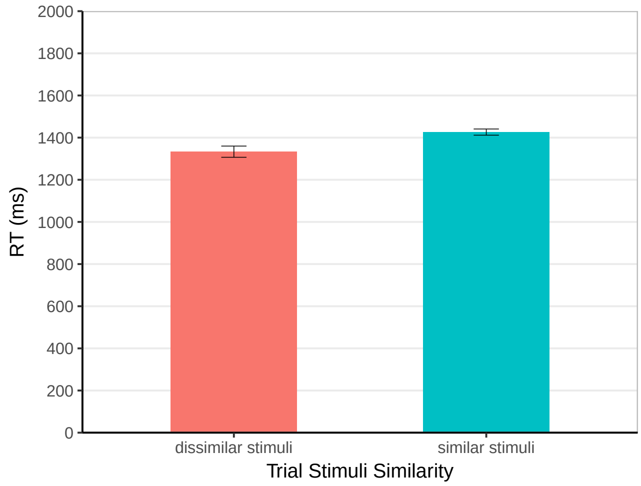


Figure 7: Response Times by trial stimuli similarity. A decision is reached quicker on dissimilar trials. Error bars denote SE of the mean.

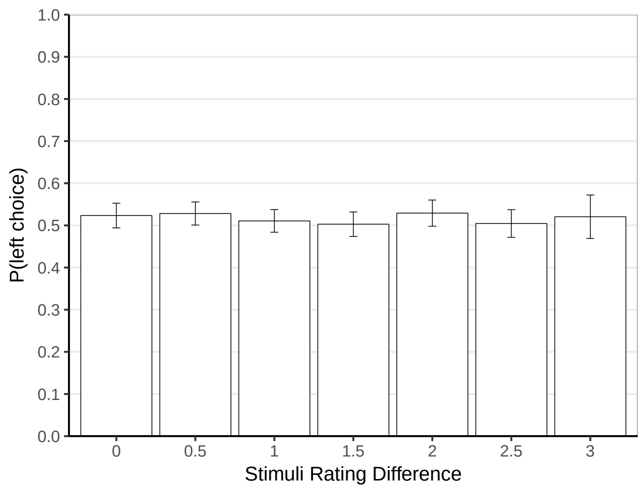


Figure 8: Decisions are not influenced by stimuli position.

did not expect participants accuracy scores to shift during the experiment. Indeed, accuracy was not affected by time on task ($\chi^2(1) = 0.76, p = 0.38$).

The best fitting LME model on trial accuracy is shown in table 5.

Conclusions from LME models

From the best fitting LME models on RT (table 4) and accuracy (table 5) we distinguish three effects on the decision process:

Effect of difficulty Participants are slower ($\sim 51ms$) and less accurate ($\sim 10.2\%$) on hard trials (i.e. two stimuli with similar ratings) than on easy trials. Because both stimuli have similar ratings in the hard condition, relative evidence for both options is much more slowly aggregated. This results in higher RTs and lower accuracy, as random noise inherit to the decision process can bias the slightly lesser rated option.

Table 4: Specification of the best fitting LME model and statistics on $log(RT)$.

Fixed effects	β	t-value	p-value
(Intercept)	7.231	133.62	< .0001
Condition (hard)	$2.608 * 10^{-2}$	2.24	< .05
Positive trial	$-8.006 * 10^{-2}$	-5.53	< .001
Similar trial	$3.828 * 10^{-2}$	2.78	< .05
Trial (time on task)	$-3.086 * 10^{-4}$	-3.99	< .0001

Random effects	Variance
Subject	0.04225

Note. df = 21.66 for all parameters. Df was calculated using Kenward-Roger approximation.

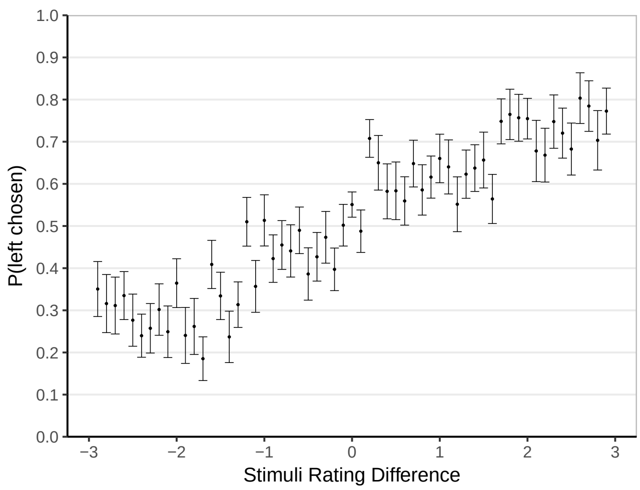


Figure 9: Decisions are influenced by the difference in stimuli rating.

As fixations are the only external source of information available to participants during the decision process it is expected that more fixations are present on hard trials than on easy trials.

Effect of aversion Participants are slower ($\sim 119ms$) on aversive trials, but are also less accurate ($\sim 5.7\%$). These findings are unexpected in terms of the attentional DDM (aDDM), which predicts that gaze duration (i.e. more and/or longer fixations) should results in better avoidance on aver-

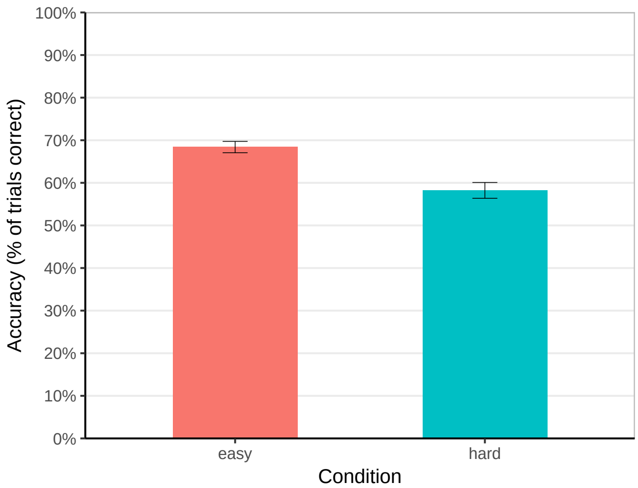


Figure 10: Accuracy scores by trial condition. Accuracy is significantly worse on hard trials ($b = -0.106, SE = 0.016, t = -6.58$). Error bars denote SE of the mean.

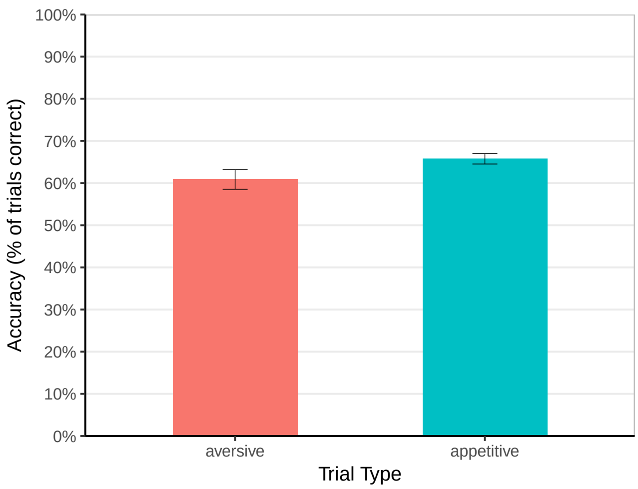


Figure 11: Accuracy scores by trial appetitiveness. Participants are significantly more accurate on appetitive than aversive trials ($b = 0.074, SE = 0.019, t = 3.83$). Error bars denote SE of the mean.

sive items (Armel et al., 2008), as gaze dwell time amplifies the proportional value of the fixated relative to the non-fixated stimulus and thus influences evidence accumulation for a decision, quantified with the *driftrate* parameter (Krajbich et al., 2010; Krajbich & Rangel, 2011; Krajbich et al., 2012). However, more recent research found that gaze duration influences stimulus selection regardless of proportional stimulus value, even on aversive items (Cavanagh et al., 2014).

Effect of similarity Participants are faster ($\sim 93ms$) on dissimilar trials, while being less accurate ($\sim -1\%$). We hypothesize that this effect is the result of a change in task perception by participants. Compared to *similar* trials (i.e. where both stimuli are from the same category), participants do not evaluate the individual stimuli values but the values of the stimuli-category (e.g. fruit or candy). As fruit stimuli are higher rated than candy stimuli on average (see fig. 3) less information is needed to reach a decision as only the category of each stimulus has to be determined, instead of the intrinsic value. Therefore we expect less fixations to be found during dissimilar trials.

But why are participants less accurate on dissimilar trials? It is possible that the presented choice in a trial breaks with the heuristic of choosing the highest rated stimulus-category: a candy stimulus can be rated higher than the alternative fruit stimulus. If the categorical heuristic is applied, less accurate decision are to be expected on dissimilar trials.

Fixations One of the main goals of this study was to develop a method to enable decision making research using commonly available hardware such as an integrated webcam instead of a standalone eyetracker. To enable comparison between both the webcam-based eyetracker and the native eyetracker, both devices were active during the decision making experiment.

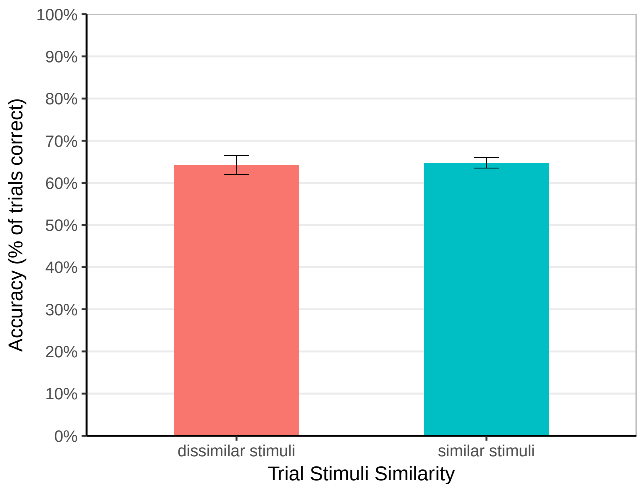


Figure 12: Accuracy scores by trial stimuli similarity. Participants are significantly less accurate on dissimilar trials (fruit-candy) ($b = 0.042, SE = 0.019, t = 2.20$). Error bars denote SE of the mean.

Figure 13 shows the similarity in proportions of fixations per trial for both the eyetracker and webcam source. Furthermore, in figure 14 the overall heatmaps for both the eyetracker and webcam are shown. While the webcam-heatmap shows a more smoothed image, it is clear from both images (figs. 13, 14) that the webcam-based eyetracker data shares a large similarity with the native eyetracker data (Pearson's correlation $\rho_(3654) = 0.63$, $p < 0.001$). However, differences are also found in eyetracker and webcam gaze durations (figure 19). This poses a potential problem, as gaze duration is a paramount parameter in the attentional DDM.

The attentional DDM predicts that longer fixations to an item should result in a higher probability of that item being chosen (Krajbich et al., 2010; Krajbich & Rangel, 2011; Krajbich et al., 2012). To investigate this, we look at the first fixation in a trial as that marks the start of the decision process. Figure 15 shows the proportion of trials where the first fixated item was chosen as a function of the duration of that fixation.

There is no significant effect of first fixation duration on first fixated item selection ($F(1,3130) = 0.487, p = 0.49, ns$). This observation breaks with the expectations from the aDDM, which states that the first fixation duration should predict selection (Krajbich et al., 2010; Krajbich & Rangel, 2011; Krajbich et al., 2012).

Both total fixations and total fixation duration per trial increase with trial difficulty (figure 16 and 17 respectively), which is in line with earlier research (Krajbich et al., 2010; Krajbich & Rangel, 2011; Krajbich et al., 2012; Cavanagh et al., 2014).

The importance of (relative) gaze duration per option for decision making is also evident from figure 18: as relative gaze time for an option increases, the choice for that option increases as well.

Another interesting insight into the decision process comes from the duration of consecutive fixations within a trial. Figure 19 shows how fixation duration differs between first, sec-

Table 5: Specification of the best fitting LME model and statistics on trial accuracy.

Fixed effects	β	t-value	p-value
(Intercept)	0.595	23.242	$< .0001$
Condition (hard)	-0.106	-6.577	$< .0001$
Positive trial	0.074	3.834	$< .001$
Similar trial	0.042	2.198	$< .05$

Random effects	Variance
Subject	$9.4 * 10^{-5}$

Note. df = 556.72 for all parameters. Df was calculated using Kenward-Roger approximation.

ond, last and other fixations. The first and second fixations are relatively long, followed by variable short fixations and finally the last fixation is of a shorter duration than the first fixation due to the reaching of the decision threshold during the last fixation. Participants are more likely to choose the last fixated item when relative rating difference increases (see figure 20).

Drift Diffusion Model

First we fitted a reference model with only reward value as a predictor for drift-rate ($DIC = 9890$). We then expanded this reference model to three alternative models in a similar way to Cavanagh et al. (2014): an independent model in which both gaze and value independently influence drift-rate, an attention DDM (aDDM) in which gaze and value interact and influence drift-rate and hybrid aDDM in which both the interaction of gaze and value and an independent effect of gaze

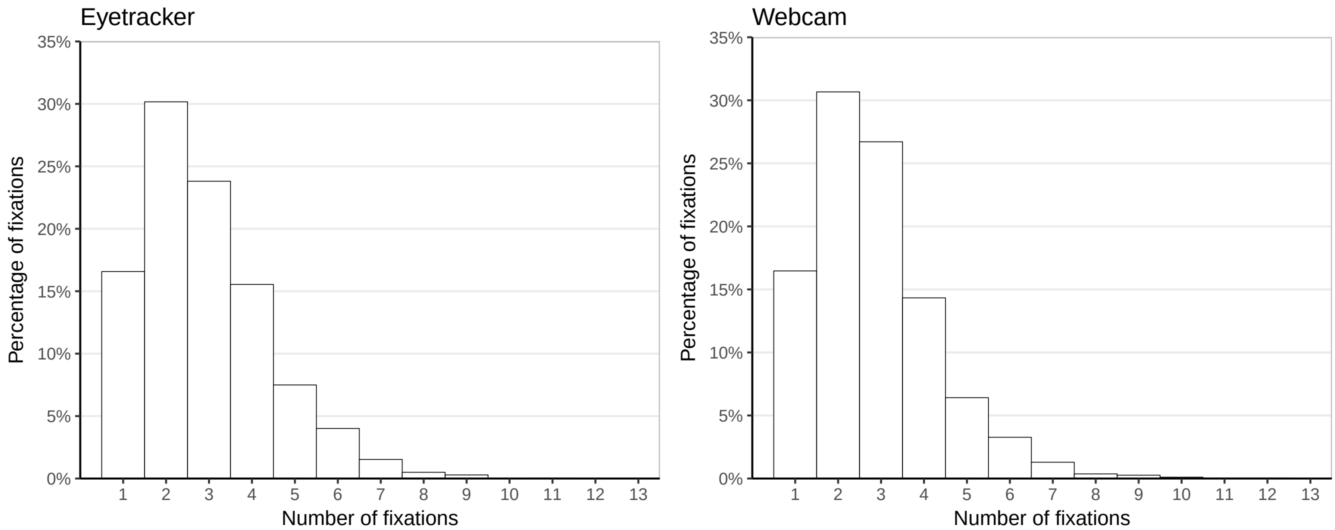
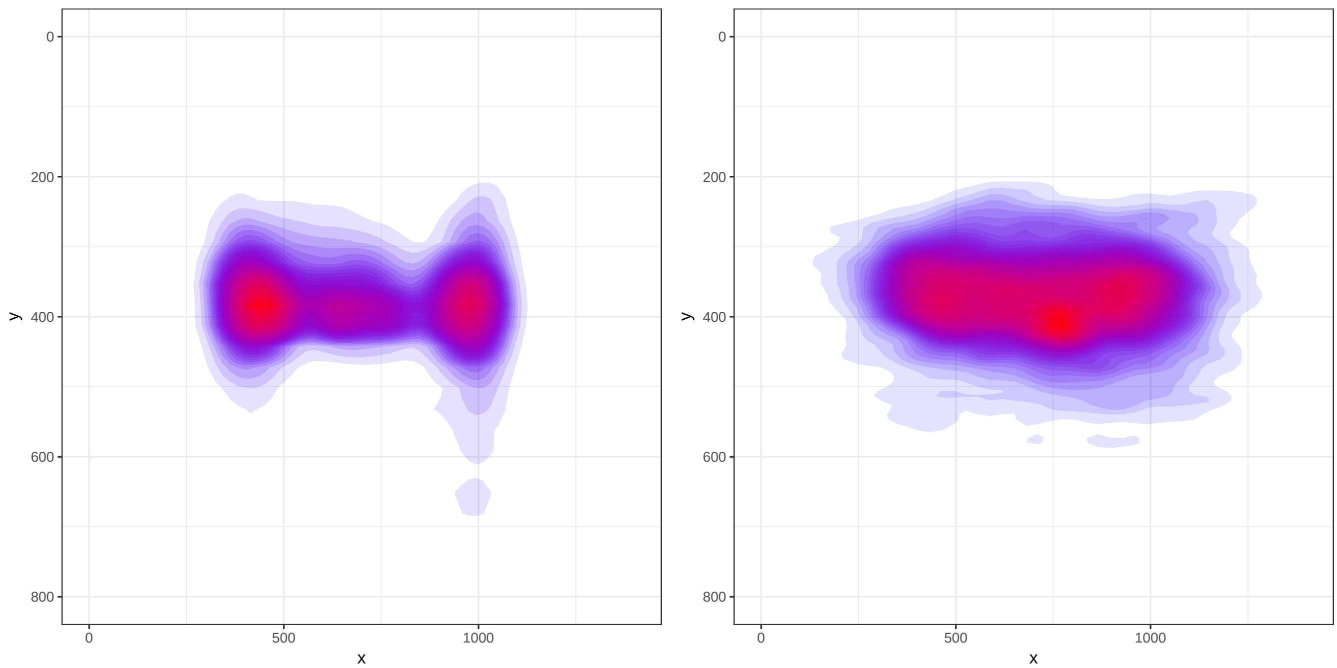


Figure 13: Proportions of trials with number of fixations per trial. The webcam and eyetracker show a similar distribution.



(a) Eyetracker

(b) Webcam

Figure 14: Heatmaps of eyetracker and webcam fixations. Colors denote density of fixations (blue = lowest, red = highest).

influences drift-rate.

Additionally we tested whether trial stimuli similarity influences the decision threshold, as a clear effect on RTs was visible in figure 7 and this effect was hypothesized as the applying of a heuristic by participants on dissimilar trials.

Table 6 shows the fitted HDDMs and figures 21, 22 and 23 show the relevant regression coefficient plots for the various models.

Both table 6 and figure 21b clearly show that the decision threshold is lower on dissimilar trials. This explains the similarity effect visible in figure 7 and 12: due to the lowering of decision threshold, a decision is reached more quickly (thus lowering RT) but the quality of the decision is not improved (as drift rate is not influenced) thus accuracy is lower on dissimilar trials.

Adding relative gaze duration as a parameter to drift-rate estimation improves the model fit considerably ($DIC_{withoutgaze} = 9890$, $DIC_{withgaze} = 9857$). This is to be expected as fixations to stimuli are the only (external) way of collecting information about the stimulus and has been extensively reported in previous research (Cavanagh et al., 2014; Krajbich & Rangel, 2011; Krajbich et al., 2012).

To determine how both intrinsic value of stimuli and relative gaze duration influences the decision process we need to examine differences in the three alternating models (independent, aDDM, extended aDDM). Focusing on

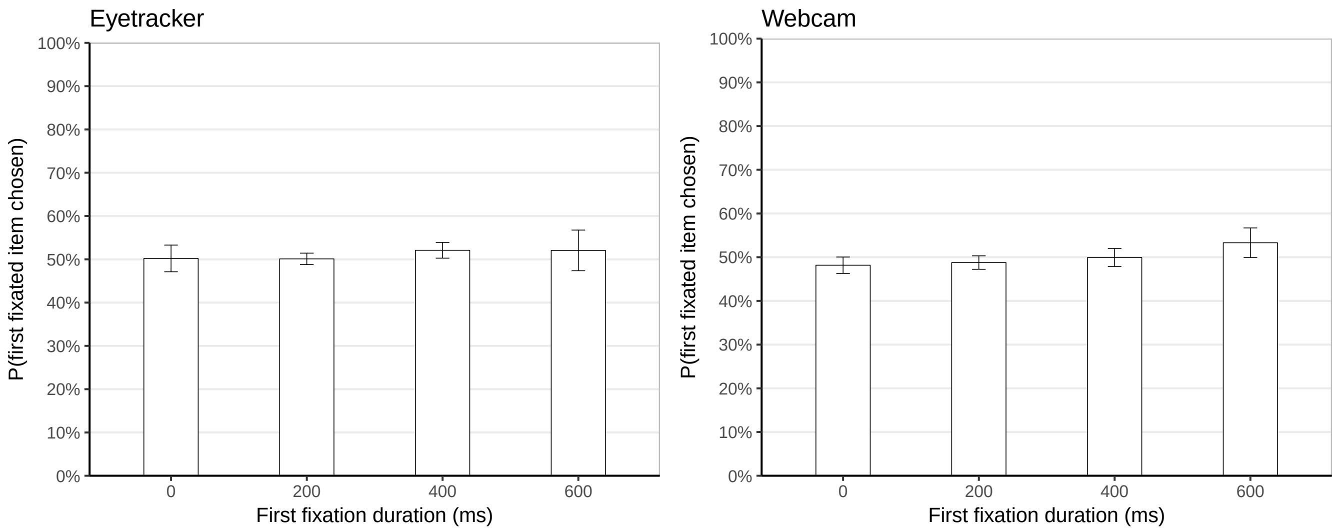


Figure 15: Duration of first fixation does not affect choice. Error bars denote SE of the mean.

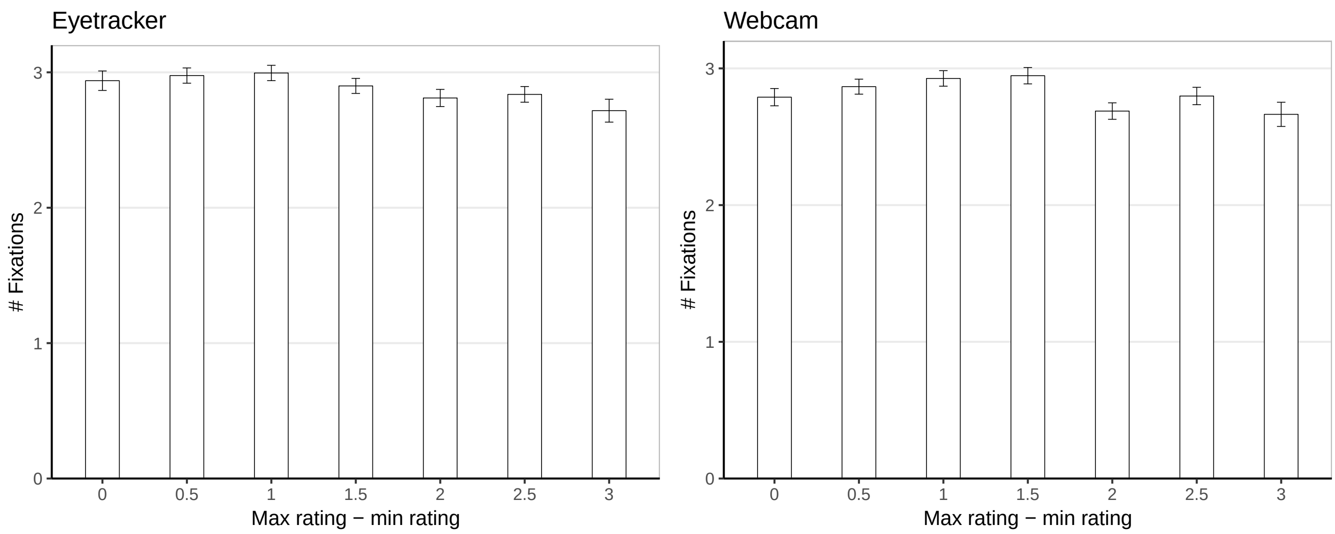


Figure 16: Average number of fixations per trial per relative rating difference. When trial difficulty decreases (i.e. larger relative rating difference) the number of fixations decreases as well. Error bars denote SE of the mean.

DIC-values, both the independent ($DIC = 9559$) and extended aDDM ($DIC = 9559$) perform better than the aDDM ($DIC = 9565$). However, inspecting the posterior plots in figure 22 shows that the extended aDDM gaze parameter encloses zero, indicating that this parameter does not adequately influence drift rate. Therefore, the independent model provides the best approximation of the collected data.

While the webcam-based models (model 5, 6 & 7) perform worse than the eyetracker-based models, inspecting the regression coefficients in both table 6 and figure 23 shows that the overall pattern is consistent with the eyetracker-based models. This shows that it is possible to use webcam-collected gaze information to adequately model the decision process using HDDM.

Discussion

In this thesis we developed and tested a new methodology of studying decision making processes on a value-based decision task using an integrated webcam for collecting gaze information.

A preference for healthy food?

We noticed that almost all participants reported abnormal concerns for dieting on the Eating Restraint Scale (ERS). Could this have affected the results of this study?

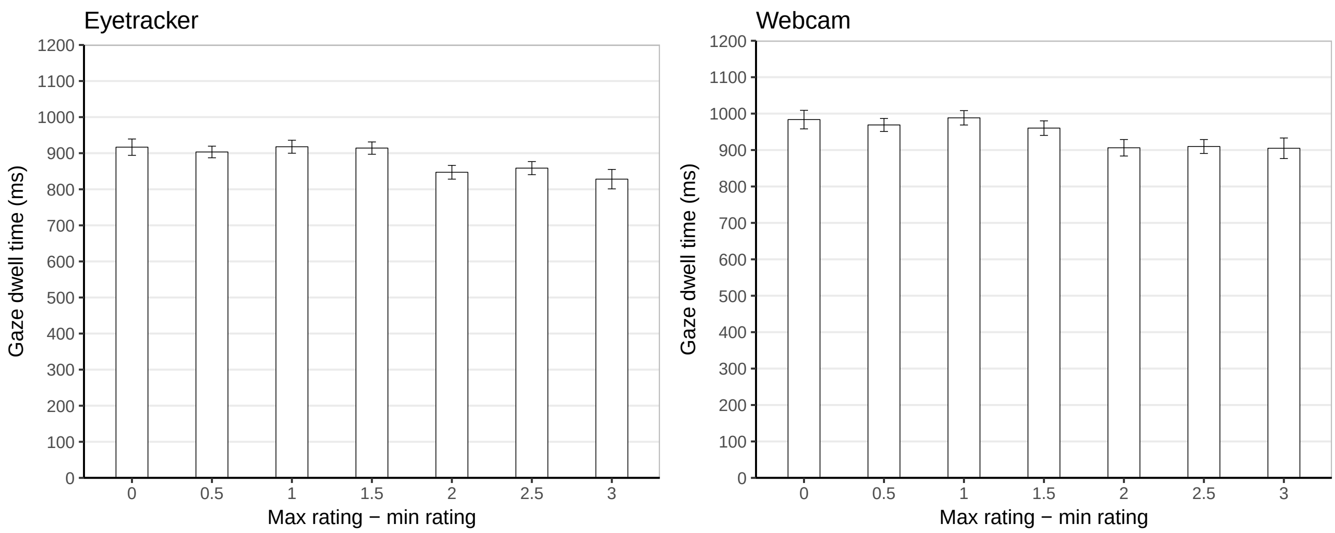


Figure 17: Total gaze duration per trial per relative rating difference. Note the larger total duration values of the webcam. Total gaze duration decreases as trial difficulty decreases. Error bars denote SE of the mean.

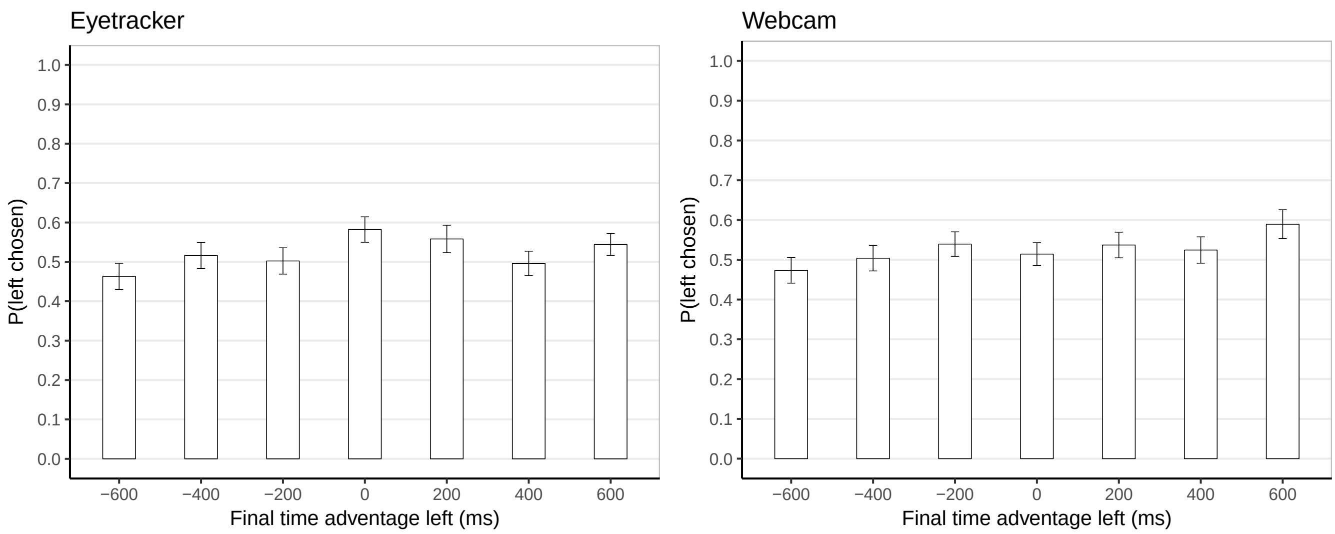


Figure 18: Proportion of stimuli chosen by difference in gaze time for that stimulus vs the other stimulus. Relative gaze time seems to influence choice, especially at -600 and 600 ms gaze difference. Error bars denote SE of the mean.

If participants are more concerned with dieting, it could very well be that their valuation of healthy fruit (and alternatively, unhealthy candy) stimuli is influenced by this increased concern. Indeed, we found that fruit stimuli were, on average, rated much higher than candy stimuli (see figure 3). However, it should be noted that fruit stimuli were sparsely represented (7 out of 50, 14%) in the stimulus set, which could give rise to some form of reversed mere-exposure effect, in which the choice for fruit stimuli is influenced by the higher occurrence of candy trials. Future research should try to control for this by increasing the overall stimulus set, while aiming for equal stimuli groups and sampling a random subset of the rated stimuli for the decision task, which should decrease any stimuli related biasing effects.

Another explanation for the abnormal concerns for dieting could lie in longitudinal changes in dieting concerns for the general population since the research by van Strien et al. (2007): since 2007 many social networking sites where people frequently share photos of themselves have gained a lot of popularity (e.g. Facebook, Instagram, Snapchat), which could influence people's physical self-appreciation. However, more recent research by van Strien, Herman, and Verheijden

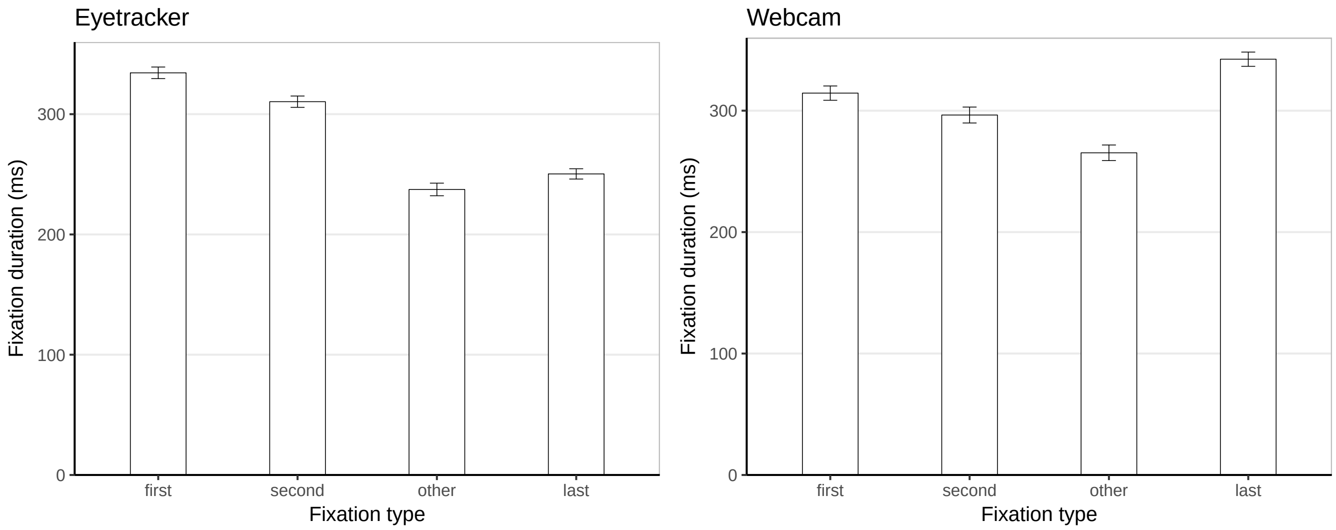


Figure 19: Fixation durations for first, second, other and last fixation per trial. The last fixation is cut short, as hypothesized by Krajbich and Rangel (2011) due to the decision threshold being reached. Note that the webcam shows a different pattern where the last fixation is the longest, indicating incorrect gaze duration estimation. Error bars denote SE of the mean.

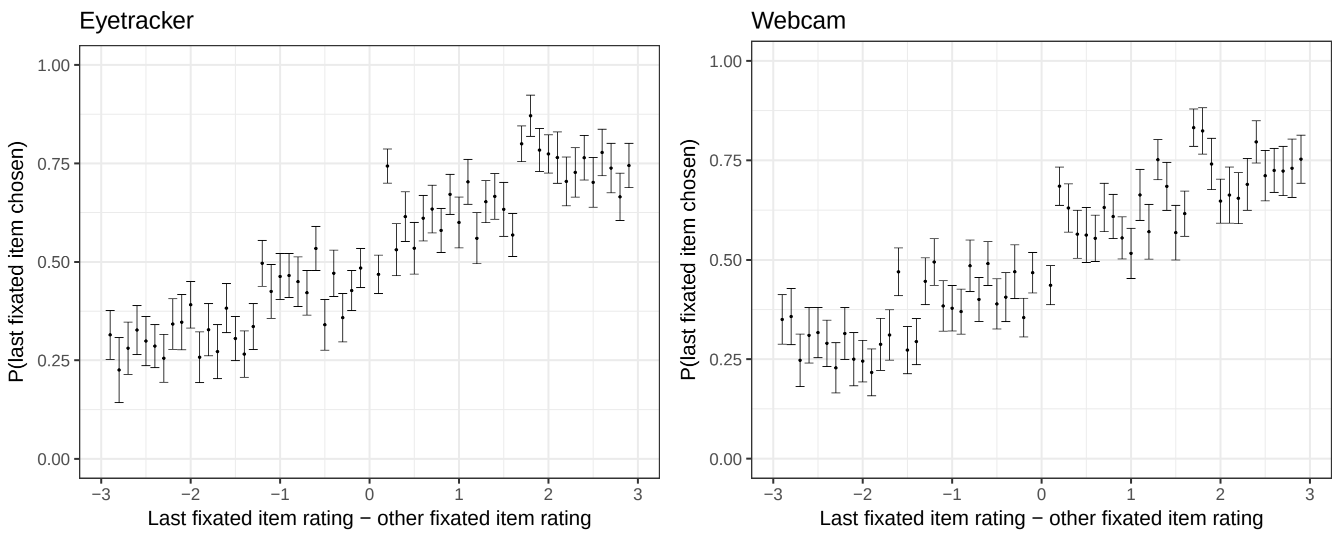


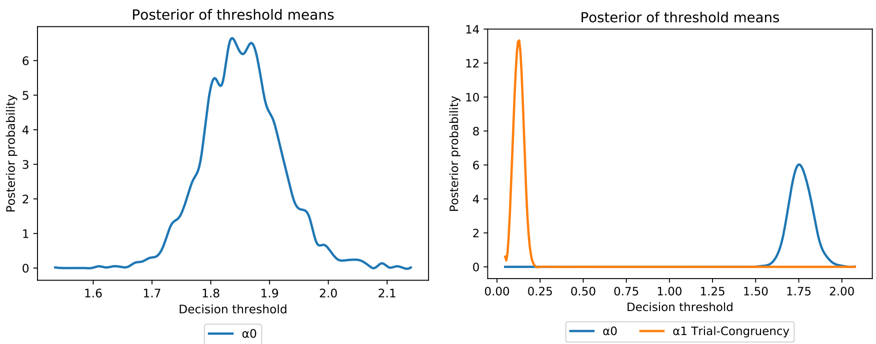
Figure 20: Proportion of stimuli chosen per relative rating difference of last fixated stimuli vs other stimuli. Participants are more likely to cast a final look at the desired stimulus when the relative rating increases. Error bars denote SE of the mean.

Table 6: Deviance Information Criterion Fits and Parameter Values for Each Model of the Influence of Gaze Time and Value on Drift Rate and Trial Congruency on Decision Threshold. Adapted from Cavanagh et al. (2014).

		drift-rate (v)						decision threshold (a)	
			β_1						
Variable	DIC	β_0	Value	Gaze * Value	Gaze	(1 - Gaze) * Value	β_3 Gaze	α_0	α_1 Trial Congruency
Model 0: Value	9890	.08 (.02)	.57 (.00)					1.87 (.00)	
Model 1: Independent	9857	.08 (.02)	.57 (.00)		.10 (.00)			1.85 (.00)	
Model 2: Independent (similarity)	9559	.09 (.01)	.56 (.00)		.10 (.00)			1.75 (.00)	.14 (.00)
Model 3: aDDM	9565	.09 (.01)		.69 (.00)		.55 (.00)		1.75 (.00)	.14 (.00)
Model 4: aDDM + Gaze	9559	.09 (.01)		.46 (.00)		.56 (.00)	.15 (.00)	1.75 (.00)	.14 (.00)
Model 5: Independent (webcam)	9864	.08 (.01)	.57 (.00)		.15 (.00)			1.77 (.00)	.14 (.00)
Model 6: aDDM (webcam)	9864	.08 (.01)		.67 (.00)		.57 (.00)		1.77 (.00)	.14 (.00)
Model 7: aDDM + Gaze (webcam)	9865	.08 (.02)		.68 (.00)		.57 (.00)	-.01 (.58)	1.77 (.00)	.14 (.00)

Note. DIC = deviance information criterion; aDDM = attention drift diffusion model.

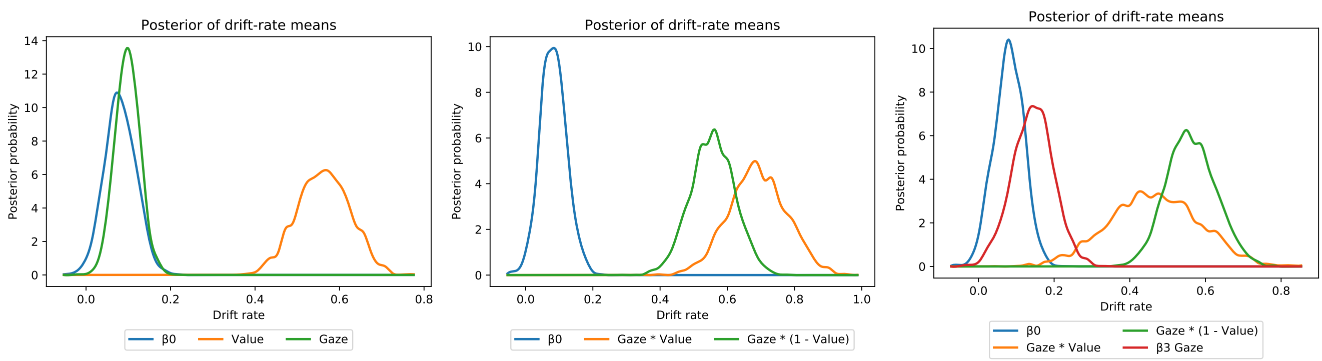
(2014) does not show significant changes in CD subscales over the general population in comparison with van Strien et al. (2007), and while these social networking sites have since than gained even more popularity, some effect should have



(a) Model 1

(b) Model 2

Figure 21: Posteriors of decision threshold parameters in model 1 and 2. Note the smoother posterior distribution shape of the intercept parameter in model 2, indicating less noise in the posterior distribution.



(a) Model 2

(b) Model 3

(c) Model 4

Figure 22: Posteriors of drift parameters in model 2, 3 and 4. Note that gaze values correspond to eyetracker gaze values. In model 2, the drift rate is driven by the value of the stimulus and gaze direction and this is added to the baseline drift β_0. In model 3, the drift rate is driven by the interaction of gaze direction and stimulus value (i.e. looking at a highly valued stimulus), and the interaction of gaze direction and inverse stimulus value (i.e. looking at a lowly valued stimulus) and this is added to the baseline drift β_0. Model 4 is similar to model 3, but drift rate is additionally driven by the gaze direction, independent from stimulus value.

been visible.

Decision Task

Participants took significantly longer to reach a decision on hard trials and were significantly less accurate (figures 5 & 10), indicating that our main manipulation functioned as expected.

Additionally, it was found that participants take longer while making a worse decision on aversive trials than on appetitive trials (figures 6 & 11) . These findings are not in line with an earlier study (Armel et al., 2008), in which longer respond times on aversive trials were associated with improved avoidance of the aversive item. However, in that study stimuli were presented sequentially instead of simultaneously, which results in a different avoidance strategy: when stimuli are simultaneously presented, avoiding to look at a specific stimuli is achieved by looking at the other stimuli, while sequentially presented stimuli are avoided by not looking at the stimuli at all. In the sequential case, no new external information is collected for the decision process, while in the simultaneous case information about the alternative option is collected.

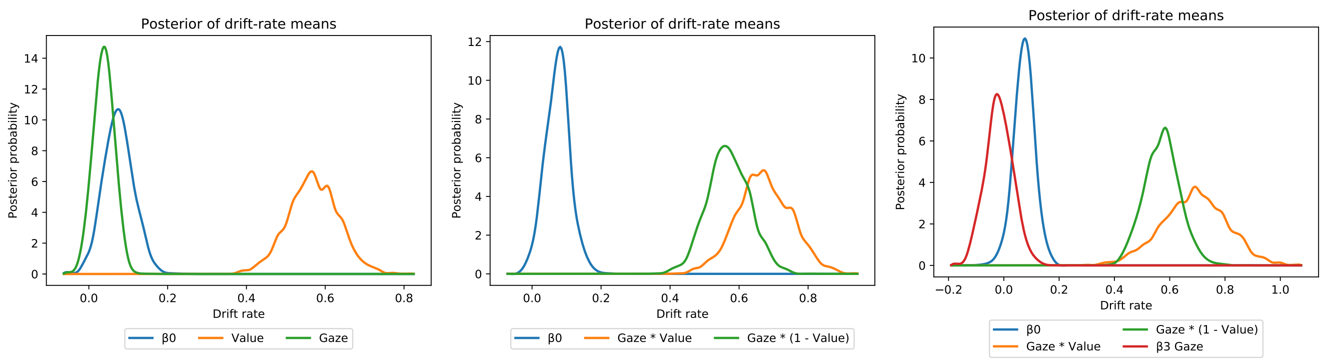
Indeed, a more recent study by Cavanagh et al. (2014) found a similar effect as found here (i.e. longer RT's and worse avoidance on aversive items) with simultaneously presented stimuli, indicating that the decrease in accuracy on aversive trials could indeed be the result of stimuli presentation (simultaneously or sequentially).

Finally, we found that participants respond slightly faster (fig. 7), but make more mistakes (fig. 12), on trials where the stimuli come from different categories (e.g. fruit and candy). We hypothesized that participants view the decision of these type of trials differently than trials were the presented choices are from the same stimuli category: they don't look at the actual stimulus anymore but only try to choose between the two presented stimuli categories (i.e. they choose between fruit or candy, not between an apple and a Mars candybar). Drift-diffusion model estimates showed that the decision threshold parameter is influenced by trial stimulus similarity and inclusion of this parameter improves the model fit (table 6 & figure 21), indicating that participants task perception is changed, and not the decision process, on these trials.

Webcam gaze-tracking

Comparison of collected webcam data and eyetracker data (figures 13 & 14) showed that while being less accurate, similar information regarding gaze can be collected through the webcam and the eyetracker.

However, the current methodology has several issues. Firstly, the current method of extracting gaze coordinates from the webcam relies solely on data collected from the eyetracker. This dependency can be removed by including a webcam calibration (similar to the classical eyetracker calibration) while recording through the webcam. This calibration should be in the form of a simple game (e.g. "follow the cat with your eyes") to prevent dropout during this stage. Secondly, due to an implementation error regarding timestamp collection, the webcam data is linearly fitted to the eyetracker data as a workaround. Future implementations should include proper timestamping on webcam collected data. Lastly, the eyetracker used performed some preprocessing on the raw data, preventing a completely fair comparison between the webcam and eyetracker data. Still, we are confident that these



(a) Model 5 (b) Model 6 (c) Model 7

Figure 23: Posteriors of drift parameters in model 5, 6, and 7. Note that gaze values correspond to webcam gaze values. In model 5, the drift rate is driven by the value of the stimulus and gaze direction and this is added to the baseline drift β_0. In model 6, the drift rate is driven by the interaction of gaze direction and stimulus value (i.e. looking at a highly valued stimulus), and the interaction of gaze direction and inverses stimulus value (i.e. looking at a lowly valued stimulus) and this is added to the baseline drift β_0. Model 7 is similar to model 6, but drift rate is additionally driven by the gaze direction, independent from stimulus value. Note the similarities between model 5 and 6 with model 2 and 3, indicating that the webcam eyetracker can extract similar information concerning gaze as a commercial eyetracker. However, this does not apply to model 7 and 4, as the interaction parameters of gaze and stimulus value switch and the posterior distribution of the independent gaze direction encompasses zero in model 7, indicating that this effect is not significant.

problems are easily fixed and that this study's results show the usefulness of webcam-collected gaze information.

Fixations

The similarities between webcam- and eyetracker-collected fixations are also evident from figure 13 which shows the distribution of fixations per trials: both distributions look alike, indicating a common source of information. However, differences between the webcam and eyetracker are present in computed fixation durations (figure 19): the webcam-based fixations show a longer last fixation per trial than the eyetracker-based fixations. We propose several reasons for this difference: First, the timing alignment (see method) of the webcam could be slightly off which would result in an incorrect start or ending of the trial. Second, the eyetracker preprocessed the data, which could influence fixation durations in an unknown way. Third, the reduced accuracy of the webcam-gaze collection (see figure 14) makes it more difficult to recognize saccades, thus not recognizing different fixations, resulting in larger fixation durations and less fixations overall.

The effect of trial difficulty (the main manipulation of the decision task) on fixations is shown in figures 16 & 17 and is in line with earlier research (Krajbich et al., 2010; Krajbich & Rangel, 2011; Krajbich et al., 2012; Cavanagh et al., 2014): Participants need more information to reach a decision as the presented options have a similar intrinsic value on difficult trials. Additional information to reach a decision can only be collected externally by fixating on the options, thus to reach a decision, more and longer fixations are needed.

First fixation duration per trial did not predict choice in this study (figure 15), which is not in line with the attentional DDM. Cavanagh et al. (2014) found that first fixation duration only predicted choice on easy (win-lose, a choice between an appetitive and aversive option) trials, which we're not presented in this study. It could be that the easy trials in this study are not easy enough compared to these win-lose trials (i.e. the difference in choices should be greater than the 10 rating points used in this study), or that the effect of first fixation duration on choice is too subtle to establish in 150 decision trials.

Future research should perform a longer decision task (i.e. many more trials) and try to create a clearer distinction between easy and hard trials: for example, hard trials have options within 10 rating points of each other, while easy trials have options which are at least 30 rating points apart. Additionally, win-lose trials (i.e. appetitive vs aversive stimulus) could be included to investigate if first fixation duration indeed predicts choice on these trials.

The Decision Making Process

While the empirical data discussed earlier provides an interesting insight in the various effects on the decision process, such as the effect of trial stimuli similarity, it cannot be used to infer how these effects influence the decision process. For that algorithmic decision making models such as the Hierarchical Drift Diffusion Model (HDDM) are needed. These DDM's are used to test different theories of how the decision process works by investigating which model best approximates the empirical data.

We used HDDM's to determine how the decision process is influenced by trial stimulus similarity (by lowering decision threshold), how gaze and intrinsic stimulus value influence the decision process and if webcam-collected gaze information can be used to reach similar conclusions as eyetracker data.

Formal comparison of both eyetracker- and webcam-based HDDM's (table 6) provides evidence that the Independent Model (in which gaze and value independently influence choice) is the best approximation of the decision process in comparison with both the attentional DDM and hybrid attentional DDM, which is in line with earlier research by Cavanagh et al. (2014) which attributes the earlier preference for the attentional DDM (Krajbich et al., 2010; Krajbich & Rangel, 2011; Krajbich et al., 2012) to a lack of aversive items in the decision task. Aversive items are important to include, as the attention DDM makes a very clear prediction on the effect of gaze duration on aversive stimuli: avoidance should improve with gaze duration, which is the opposite of what we and Cavanagh et al. (2014) found.

Further research should focus on the difference between aversive and appetitive items and if the decision process differs between those items.

This study did not collect information regarding pupil dilation, while Cavanagh et al. (2014) found that changes in pupil dilation provide an additional predictor on choice in value-based decision tasks. As many eyetrackers report information about pupil dilation, this can easily be integrated in HDDM's to investigate the influence of pupil dilation on the decision process.

However, pupil dilation cannot be easily extracted from webcam-data using the OpenFace framework, so a different approach should be used in future research to enable integration of pupil dilation in Decision Making models on the basis of webcam-collected gaze information.

Conclusions

We conclude that it is possible to perform decision making research through the webbrowser and using a webcam to collect gaze information without losing valuable sources of information needed to distinguish between alternative models of decision making. This could have large implications for future decision making research, as this methodology can vastly decrease costs, reduce time needed for participant recruitment and enable researchers to study a more representative sample of the population instead of almost exclusively WEIRD participants.

Additionally, in line with Cavanagh et al. (2014), we conclude that gaze and stimulus value independently influence choice in a value-based decision task. This implicates that it may be possible to influence choice by presenting visual information and directing gaze, regardless of intrinsic value preference, a method used by marketeers for a very long time.

References

Armel, K. C., Beaumel, A., & Rangel, A. (2008). Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making*, *3*(5), 396.

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. In *Proceedings of the ieee international conference on computer vision workshops* (pp. 354–361).

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). Openface: an open source facial behavior analysis toolkit. In *Applications of computer vision (wacv), 2016 ieee winter conference on* (pp. 1–10).

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

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology: General*, *143*(4), 1476.

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(35), 12629–12633.

EyetechDS. (2014). *Eyetech digital systems, vt-2 mini.* Retrieved 2014-05-22, from `http://web.archive.org/web/20140522185745/http://www.eyetechds.com:80/vt2-mini.shtml`

EyetechDS. (2016). *Eyetech digital systems, quicklink sdk.* Retrieved 2016-12-02, from `http://web.archive.org/web/20161202194955/http://www.eyetechds.com/quick-link-sdk.html`

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* Retrieved from `http://www.deeplearningbook.org` (Book in preparation for MIT Press)

Halekoh, U., & Højsgaard, S. (2014). A kenward-roger approximation and parametric bootstrap methods for tests in linear mixed models – the R package pbkrtest. *Journal of Statistical Software*, *59*(9), 1–30. Retrieved from `http://www.jstatsoft.org/v59/i09/`

Henrich, J., Heine, S. J., & Norenzayan, A. (2010a). Most people are not weird. *Nature*, *466*(7302), 29–29.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010b). The weirdest people in the world? *Behavioral and brain sciences*, *33*(2-3), 61–83.

Kenward, M. G., & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis*, *53*(7), 2583–2595.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature neuroscience*, *13*(10), 1292–1298.

Krajbich, I., Lu, D., Camerer, C., & Rangel, A. (2012). The attentional drift-diffusion model extends to simple purchasing decisions. *Frontiers in psychology*, *3*.

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences*, *108*(33), 13852–13857.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). lmertest: Tests in linear mixed effects models [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=lmerTest` (R package version 2.0-33)

Nisbett, R. (2010). *The geography of thought: How asians and westerners think differently... and.* Simon and Schuster.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American psychologist*, *54*(9), 741.

Polivy, J., Herman, C. P., & Warsh, S. (1978). Internal and external components of emotionality in restrained and unrestrained eaters. *Journal of Abnormal Psychology*, *87*(5), 497.

Rangel, A., Camerer, C., & Montague, P. R. (2008). A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, *9*(7), 545–556.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current opinion in neurobiology*, *20*(2), 262–270.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.

RStudio. (2012). *version 0.98.507.* Boston, MA: Author.

San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D. W., & Hansen, J. P. (2010). Evaluation of a low-cost open-source gaze tracker. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 77–80).

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception.* Bobbs-Merrill Indianapolis.

Sewell, W., & Komogortsev, O. (2010). Real-time eye gaze tracking with an unmodified commodity webcam employing a neural network. In *Chi'10 extended abstracts on human factors in computing systems* (pp. 3739–3744).

Skovsgaard, H., Agustin, J. S., Johansen, S. A., Hansen, J. P., & Tall, M. (2011). Evaluation of a remote webcam-based eye tracker. In *Proceedings of the 1st conference on novel*

gaze-controlled applications (p. 7).
van Strien, T., Breteler, M. H., & Ouwens, M. A. (2002). Restraint scale, its sub-scales concern for dieting and weight fluctuation. *Personality and Individual Differences*, *33*(5), 791–802.
van Strien, T., Herman, C. P., Engels, R. C., Larsen, J. K., & van Leeuwe, J. F. (2007). Construct validation of the restraint scale in normal-weight and overweight females. *Appetite*, *49*(1), 109–121.
van Strien, T., Herman, C. P., & Verheijden, M. W. (2014). Dietary restraint and body mass change. a 3-year follow up study in a representative dutch sample. *Appetite*, *76*, 44–49.
von der Malsburg, T. (2015). saccades: Detection of fixations in eye-tracking data [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=saccades` (R package version 0.1-1)
Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). Hddm: hierarchical bayesian estimation of the drift-diffusion model in python. *Frontiers in neuroinformatics*, *7*, 14.
Wood, E., Baltrušaitis, T., Zhang, X., Sugano, Y., Robinson, P., & Bulling, A. (2015). Rendering of eyes for eye-shape registration and gaze estimation. In *Proceedings of the ieee international conference on computer vision* (pp. 3756–3764).

Appendix A Stimuli

Figure 24: maltesers

Figure 25: milka-melocakes

Figure 26: MnMs

Figure 27: twix

Figure 28: snickers

Figure 29: lays-naturel

Figure 30: ah-stroopwafels

Figure 31: pepsels

Figure 32: tony-chocola

Figure 33: roze-koeken

Figure 34: dropfruit-duos

Figure 35: katja-apekoppen

Figure 36: haribo-kikkers

Figure 37: karamel-fudge

Figure 38: haribo-bananen

Figure 39: zure-matjes

Figure 40: frisia-kabelspek

Figure 41: katja-yoghurtgum

Figure 42: red-band-pretsleutels

Figure 43: smarties-v2

Figure 44: chokotoff

Figure 45: kitkat

Figure 46: lion

Figure 47: nuts

Figure 48: tomblerone

Figure 49: bros

Figure 50: milkyway

Figure 51: haribo-aardbeijen

Figure 52: kleine-muntdrop

Figure 53: klene-ovaaltjes

Figure 54: haribo-trekdrop

Figure 55: lookolook-dropveters

Figure 56: lookolook-droptijden

Figure 57: lays-paprika

Figure 58: pringles-v2

Figure 59: lays-bolognese

Figure 60: doritos-nacho-cheese

Figure 61: tortilla-chio-rolls

Figure 62: chio-heartbreakers

Figure 63: lays-thai-sweet-chili

Figure 64: chio-popcorn-sweet

Figure 65: chio-popcorn-salt

Figure 66: appels

Figure 67: mandarijn

Figure 68: snoeptomaatjes

Figure 69: wortels

Figure 70: druiven

Figure 71: bananen

Figure 72: aardbeien

Figure 73: blauwebessen