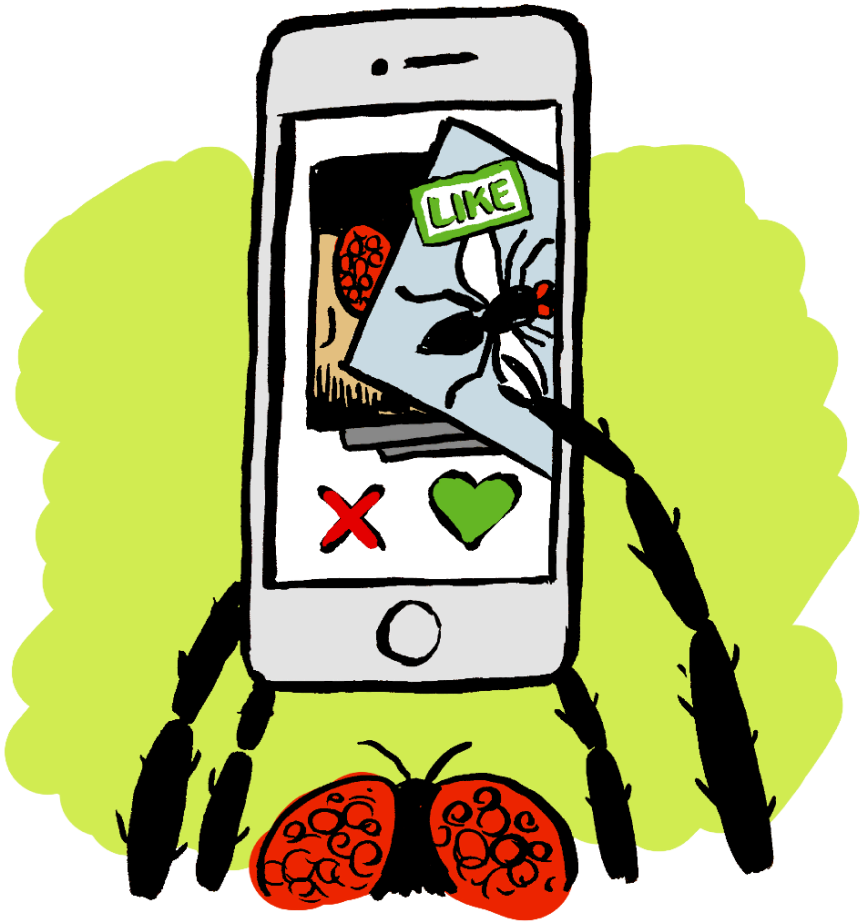
Decisions on the Fly

The applicability of diffusion models to decision making in *Drosophila melanogaster*



Franciska de Beer
S2700409
30-06-2019
Behavioural and Cognitive Neurosciences (BCN) – C-track
Faculty of Science and Engineering, University of Groningen

Supervised by:

T.A. Verschut	GELIFES
M.K. van Vugt	Bernoulli Institute
J.C. Billeter	GELIFES

Illustration by Melle Veltman

Abstract

The Drift Diffusion Model (DDM) describes the process of fast and binary human decision making as evidence accumulation over time until a threshold is reached. Although the DDM can describe a wide variety of human behaviours, there is still a lack of neural evidence for the mechanism that the model describes. The fruit fly *Drosophila melanogaster* might provide a solution as experimental tools available for this species would allow to causally manipulate the neural substrates of decision making. The first step in this approach is to verify whether the DDM can adequately describe decision making in *D. melanogaster*. Therefore, a courtship choice experiment was designed to generate *D. melanogaster* decision making data to which the DDM was applied, in order to assess whether the DDM would account for the data. The experiment consisted of two conditions: in the easy condition, a male fly had to decide whether to court a male or virgin female, whereas in the difficult condition, the male had to choose between a mated and virgin female. Since the DDM expects that evidence accumulates faster in easy than in difficult decisions, the male flies were hypothesized to respond fast and with high accuracy in the easy condition, and to respond slower and with low accuracy in the difficult condition. As hypothesized, males responded with higher accuracy in the easy than the difficult condition. Contrary to the hypothesis, responses were slower in the easy than in the difficult condition. Fitting the DDM to *D. melanogaster* data shows that the DDM might be able to *D. melanogaster* data. Thus, future research might further assess whether *D. melanogaster* is a suitable model organism to study the DDM and eventually investigate the neural mechanisms of decision making in *D. melanogaster* to test DDM predictions at the neural level.

Table of contents

Introduction

Decision making: the big picture and open questions
A decision can be defined as a choice between multiple options. Currently, much research on decision making focuses on which brain areas are involved in this process, which neural circuits play a role, and how neural activity is related to the act of choosing (Glimcher and Fehr 2014). While these particular aspects of decision making have been studied by various disciplines, little is known on how they are connected (Sanfey 2007). In other words, the exact mechanisms underlying decision making have not yet been defined (Doya and Shadlen 2012). Studying these mechanisms would not only improve knowledge of a fundamental human capacity, but might also provide new insights in mental disorders that come along with a dysfunction in decision making (Goschke 2014).

Computational models to study the mechanisms of decision making
Decision making is currently studied by various disciplines, such as economics, psychology, neurosciences, and mathematics. However, the area of cognitive modelling is particularly concerned with the mechanisms of choice. In general, there are three main types of cognitive models: verbal-conceptual, mathematical, and computational models (Sun 2012). With verbal-conceptual models, scientists describe a cognitive process or relations with concepts and words, whereas with mathematical models formulae and equations are used to present a relationship between parts of a cognitive process. With computational models, scientists create algorithmic simulations which simulate the steps of a cognitive process as they theoretically occur in the human brain (Ashby and Helie 2011). In the past few decades, the field of computational modelling has received increasing attention, in part due to two major advantages: firstly, cognitive models generate a clear theory and falsifiable hypotheses on how people actually behave (Glimcher and Fehr 2014), and secondly, these detailed hypotheses can be tested with behavioural experiments (Ashby and Helie 2011). However, one major problem with computational modelling is the difficulty in determining the validity of a model (Sun 2012). That is, are the processes that the computational model assumes really occurring in the human brain? Matching the predictions of a computational model with human behaviour only generates evidence on the outcome of the process, but does not directly test the process itself. Thus, neural evidence is required to verify whether the mechanisms that a computational model describes occur in the human brain. However, detailed neural evidence in humans is difficult to obtain, as most human experimental methods necessarily rely on non-invasive techniques, such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI).

The Drift Diffusion Model
As with many cognitive models, there is ample behavioural evidence but little neural evidence for Ratcliff's influential Drift Diffusion Model (DDM) for fast and binary decision making (Ratcliff and McKoon 2008). The DDM describes human decision making as the accumulation of evidence over time until a threshold is reached (Ratcliff et al. 2016) (Figure 1). In the past few decades, the model has been applied to various kinds of human decision behaviour, such as lexical decisions and categorization tasks (Ratcliff et al. 2016). Specifically, the DDM describes behavioural data from fast and binary decisions which are summarized in the so-called response time (RT) distributions. That is, every decision generates two measures, response time (RT) and accuracy. RT is how fast the subject made a decision, and accuracy is whether the subject choose the correct option. The RT and accuracy can be summarized in a RT distribution, which displays the proportion of decisions made at a specific RT for correct or incorrect choices. How well the DDM fits the data depends on the similarity between the DDM predicted RT distribution and the observed RT distribution based on behavioural data. Note that after Ratcliff first proposed the DDM (1978), various mathematical and computational variants of the DDM have been developed (Ratcliff and Childers 2015).

The DDM includes four main parameters which determine the predicted RT distribution (Figure 1). First of all, the speed of evidence accumulation, called the 'drift rate', depends on the quality of information from sensory stimuli or memories. Secondly, the height of the thresholds indicates the amount of information that is necessary to make a decision. The height of the thresholds may for instance vary due to mental states, such as stress or cautiousness (Ratcliff and McKoon 2008). With cautiousness, the thresholds move upwards and hence more information is required before a decision

is made, whereas stress is associated with lowered thresholds so decisions are made with less information. Thirdly, in case there is a bias for one of the response options, the starting point from which evidence is accumulated will shift towards the corresponding threshold, so less information is necessary to reach that boundary. Fourth, there are processes that do not belong to decision making itself, such as processing information and executing a response, which are all included in the 'non-decision time' parameter.

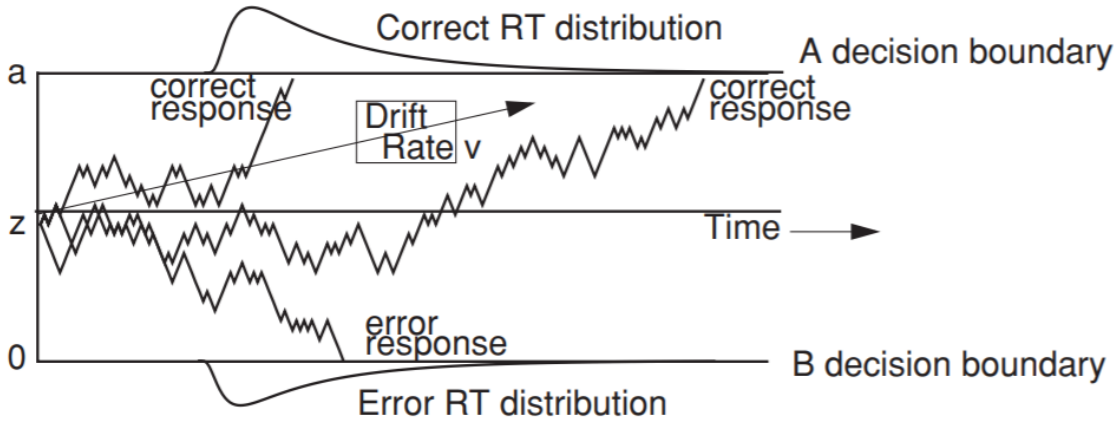


Figure 1. Illustration of the drift diffusion model. During decision making, evidence accumulates over time (drift rate 'v') from the starting point 'z' until one of the two decision boundaries ('a' or '0') is reached and a response is executed (Ratcliff and McKoon 2008).

Although the model successfully describes behavioural data, there is a lack of neural evidence for the DDM in humans (Smith and Ratcliff 2004; vanVugt et al. 2012). This poses the question of whether the mechanisms that the DDM describes actually occurs in the human brain. Besides providing knowledge on a basic human trait, the relevance of finding neural evidence for the DDM is that the model is increasingly used in clinical settings. The model is currently used to determine which aspect of the decision making process is different in clinical populations (White et al. 2010). For instance, the DDM has been used to specify how decision making is impaired in people with schizophrenia (Moustafa et al. 2015) and in children with an attention deficit hyperactivity disorder (ADHD) (Fosco et al. 2017; Huang-Pollock et al. 2017; Pirrone et al. 2017). Thus, while there is no definite answer on whether the model is valid, the DDM is used in clinical settings more often.

***Drosophila melanogaster* as a possible solution**

Due to the difficulty in finding neural evidence for the DDM in humans, a possible solution might be to focus on a different model organism in which the DDM predictions can be tested at a neural level. One potential model organism is the fruit fly *Drosophila melanogaster*. Experimental tools available in this species can be used to map the neural mechanisms of decision making in great detail (St Johnston 2002; Venken et al. 2016). Importantly, techniques such as optogenetics can easily be deployed to causally test the DDM predictions at single neuron level (Simpson and Looger 2018). In addition, the information on which *D. melanogaster* bases its decision is well-known, and the neural pathways by which this information is processed are well-identified (Depetris-Chauvin et al. 2015). Thus, it is possible study *D. melanogaster* decision making by changing sensory information and manipulating the neural processing pathways. Although it may seem inadequate to use *D. melanogaster* to test a computational model designed for humans because humans and flies are not much alike in external appearances, there are numerous similarities between humans and *D. melanogaster*. For instance, nearly 75% of the human disease-causing genes have a homologue in *Drosophila melanogaster* (Pandey and Nichols 2011). In addition, *D. melanogaster* has frequently been used to study human diseases like Alzheimer's disease (Moloney et al. 2010), schizophrenia (Shao et al. 2011), and ADHD (Kim et al. 2017). Since *Drosophila melanogaster* has been used to study various human diseases and their underlying mechanisms, it might be a suitable model organism to study decision making as well.

The first step in this model organism approach is to test whether the DDM can describe decision making data in *D. melanogaster*. By performing a model fit, the similarity between the DDM predicted and observed RT distributions is assessed. In addition, behavioural manipulations should lead to changes in estimates of the corresponding DDM parameter. For instance, when difficulty is manipulated in an experiment, this should only lead to changes in the drift rate estimates and should not affect the threshold and non-decision time parameters. In case the DDM cannot account for *D. melanogaster* data, it may seem unlikely that the model assumptions underlie decision making in *D. melanogaster*. However, if the model would fit *D. melanogaster* data, the second step would be to determine the neural mechanisms of decision making in *D. melanogaster* and test whether these are in line with the DDM predictions.

It must be noted that DasGupta Ferreira, and Miesenböck (2014) already applied a simple DDM to *D. melanogaster* decision making data. They found that RTs and accuracy changed in line with the psychometric function predicted by the DDM. From an odour discrimination task, the RT and accuracy for each fly were used to compute the population average RT and accuracy, and secondly, a data transformation was applied to these averages to generate a single drift rate estimate and a single threshold estimate. Although the article suggests that the RT and accuracy population averages are in line with the DDM, the distribution of observed RTs was not compared to the model-predicted RT distribution to assess whether the DDM could describe the data. Furthermore, although the experiment included a manipulation of difficulty which could be used to assess whether the drift rate estimates would change accordingly, there were no drift rate estimates given for the easy, intermediate, and hard condition separately. Hence, it could not be assessed whether the observed behaviour changed in accordance with the DDM predictions due to the difficulty manipulation. If the behaviour was in line with the DDM, only the drift rate estimates would change due to the difficulty manipulation, and the threshold and non-decision time estimates would have remained constant. In addition, one problem with the odour discrimination task is that the behaviour which indicates a decision (walking) is the same as the preceding and following behaviour. Hence, it is difficult to determine exactly when the fly made a decision. Thus, while this study suggests that *D. melanogaster* data is in line with the DDM, there is no definite answer yet on whether the DDM can account for these data. Therefore, the aim of this thesis is to test whether the DDM can describe decision making in *D. melanogaster.*

A courtship experiment to test *Drosophila melanogaster* decision making

In order to generate *D. melanogaster* data to which the DDM can be applied, a courtship experiment was designed. There were three main reasons to focus on courtship behaviour for this experiment. Firstly, courtship behaviour is typically initiated fast, which is important for testing a model on rapid decision making. Secondly, because courtship behaviour consists of a stable pattern of distinctive actions, it is possible to label one action as the 'decision moment'. Courtship behaviour can be divided into six steps (Figure 2) where the male orients towards the female, taps the female with its forelegs, performs the courtship song by vibrating its wings, licks the female's genitals with its proboscis, bows its abdomen in an attempt for copulation, and copulation results (Billeter et al. 2006). In this experiment, the courtship song in which the male extends and vibrates his wings is taken as the prime indication that the male fly has made the decision to court the female. This is both because wing extension is an easily observable behaviour, and because the male's decision to court a fly is based on volatile and non-volatile cues which were gathered during the two steps preceding the courtship song. The third reason was that courtship behaviour in the male is based on well-identified olfactory and gustatory cues. Thus, the pathways by which these cues are processed are already known, and hence it is possible to precisely manipulate the neural pathways of information processing. Notably, to ensure that the male's decision is not based on social cues but entirely on olfactory and gustatory information, the heads and legs of the presented stimuli flies were cut off.

Manipulating the drift rate

In the courtship experiment, the male fly has to choose which of the two presented flies it wants to court. In this experiment, the drift rate model parameter was manipulated. The drift rate, the speed of evidence accumulation which depends on the quality of information, is altered by creating an easy and difficult condition. It is hypothesised that in the easy condition, there is clear evidence and thus a fast drift rate, which lead to small RTs and high accuracy. On the contrary, information in the difficult condition is hypothesized to be more ambiguous, leading to a slower drift rate, higher RTs and lower

accuracy. In the easy condition, the stimuli consist of a virgin male and a virgin female. This condition is 'easy' as it is relatively simple for the male to distinguish between a male and virgin female due to a clear difference in pheromones. The male fly produces higher levels of (Z)-7-tricosene (7-T), whereas the virgin female generates the pheromones (7Z,11Z)-heptacosadiene (7,11-HD) and (7Z,11Z)-nonacosadiene (7,11-ND) (Billeter et al. 2009). In the difficult condition, a mated female with a mating plug and a virgin female are presented. Namely, during copulation, the mated female receives a mating plug which consists of sperm from the male, but also pheromones that render the female unattractive to other males. This strategy is called chemical mate-guarding, in which the pheromonal profile of the female is manipulated to prevent the female from polyandry, and hence increases the male's changes to produce offspring (Laturney and Billeter 2016). Thus, the virgin and mated female mainly differ due to the anti-aphrodisiac pheromones cis-Vaccenyl Acetate (cVA) and 7-T, which are transferred into the female by the male during copulation (Laturney and Billeter 2016). All other model parameters, such as the starting point, non-decision time and thresholds are hypothesized to remain constant across conditions.

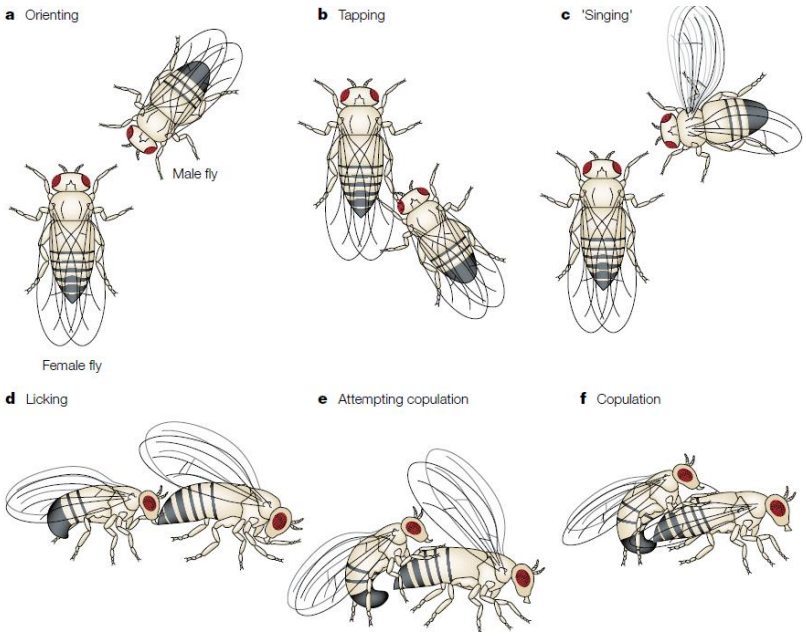


Figure 2. Overview of *D. melanogaster* courtship behavior which consists of six distinguishable steps: a. orienting, b. tapping, c. wing extension and vibration, d. licking, e. attempting copulation, f. copulation (Sokolowski 2001).

Model fitting

The courtship experiment in which the male chooses between two flies generates two measures, response time (RT) and accuracy. RT is the time between the male entering the arena and wing extension. Accuracy depends on whether the male initiated courtship towards the virgin female (correct) or towards the male or mated female (incorrect). In both conditions, the correct choice for the male is to court the virgin female, because the male would have a higher chance to produce offspring with the virgin female if the stimuli flies were still alive. These measures will generate the RT distributions that can be fitted to the DDM in order to test whether the DDM can describe decision making in *D. melanogaster*. Specifically, two diffusion models will be used to fit *D. melanogaster* data: the EZ-diffusion model (Wagenmakers et al. 2007), and the hierarchical drift diffusion model (HDDM; Wiecki et al. 2013). The EZ model is a mathematical model which transforms the data to obtain estimates of the drift rate, threshold height, and non-decision time. The HDDM is a computational model which can produce estimates of the drift rate, threshold, non-decision time, starting point, and variance in drift rate, starting point, and non-decision time. One main advantage of the HDDM is that it has greater flexibility than the EZ model, because the HDDM can be tailored to include a selection of or all these parameters. Thus, various HDDM models (including different parameters) can be applied to the data to assess which particular model best describes the data. However, the HDDM does require larger sample sizes, while the EZ can provide parameter estimates with a small amount of observations. Therefore, both models were applied to *D. melanogaster* data.

Methods

Fly lines and rearing

In the experiment, wild-type *D. melanogaster Canton-S* flies were used. The flies were maintained in bottles containing 50 ml standard diet at 25°C on a 12 h light-dark cycle (light on at 9 a.m.). The fly food contained molasses (30 g/L), agar (10 g/L), sucrose (15 g/L), glucose (30 g/L), cornmeal (15 g/L), wheat germ (10 g/L), soy flour (10 g/L), yeast (35 g/L), propionic acid (5 ml/L), tegosept (10 ml/L). Virgin flies were collected within 2-3 hours after eclosion, housed with 20 flies per vial, and transferred to new vials after 3-4 days. The flies were used for experiments at 5-7 days after eclosion.

Courtship assay

To test whether the DDM can describe *D. melanogaster* decision making, a courtship assay was designed to test the effect of difficulty on response time (RT) and accuracy. For each assay, two decapitated flies were placed in a dish with forceps and a mesh was added on top (supplementary material 5). After the recording was started, the male responder fly was inserted into the dish with an aspirator. Two assays were placed on the diffuser plate and recorded at the same time. The recording was stopped when the males in both assays showed wing vibration, or after 3 minutes, since pilot experiments showed that if a male did not extend his wings within 3 minutes, it rarely performed courtship behaviour later on (only 5 out of 200 flies responded between 3 and 5 minutes). All trials with RTs smaller than 6 seconds were discarded, as it is unsure whether the male responder processed all information within only 5 seconds.

In the experiment, task difficulty was manipulated (Figure 3). For this, the easy condition included a male versus a virgin female stimuli fly, which produce different odours (introduction) and hence are relatively easy to distinguish. On the contrary, in the difficult condition a mated and a virgin female stimulus fly were presented, which are more difficult to distinguish. In both cases, courting the virgin female is seen as correct, and choosing the mated female or male as incorrect, because the male would have the greatest chance to produce offspring with the virgin female if the stimuli flies were still alive.

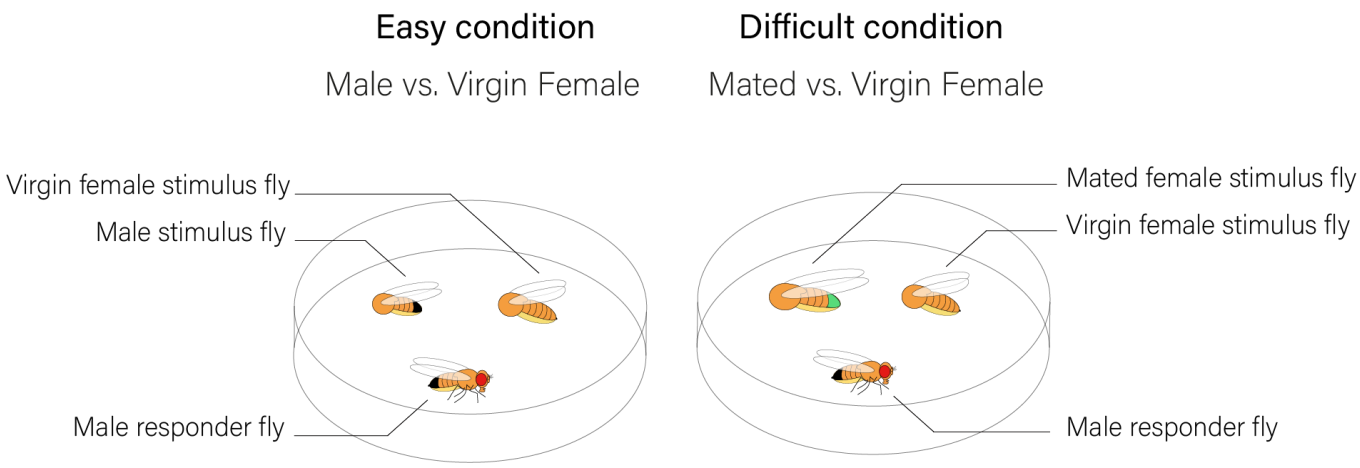


Figure 3. Courtship assays illustrating the difficulty manipulation. In the easy condition, the male and virgin female are relatively easy to distinguish because males produce higher levels of 7-T, whereas the virgin female generates 7,11-ND and 7,11-HD. In the difficult condition, the virgin and mated females are less distinct in chemosensory cues because both produce 7,11-ND and 7,11-HD, and the females only differ due to the mating plug in the mated female which includes cVA and 7-T.

Prior to the experiment, ten virgin males and females were placed in a 50 mm Petri dish for 60 to 90 minutes to generate the mated females used in the assays. Male responder flies were isolated in 2 ml centrifuge tubes (70737002, Sarstedt AG & Co., Germany) 60 to 120 minutes before the onset of the experiment. The heads and legs of the stimuli flies used in the assays were removed with a scalpel under CO_2 anaesthesia. Each assay consisted of the bottom part of a 35 mm Petri dish which walls were coated with a siliconizing reagent (Sigmacote, Sigma-Aldrich, The Netherlands) to prevent the flies from walking upwards. A drop of artificial diet containing sucrose (100g/L), yeast (50 g/L), agar (12 g/L), and distilled water was pipetted in the dish, and a grey polyester mesh (curtain 'Lima',

EHDP, The Netherlands) was used as a cover. A mesh was used because trial experiments showed that the males could not distinguish between the presented male and female in a closed arena, whereas the males could distinguish between them in an open arena with a mesh (supplementary material 5).

Behavioural experiments were performed in a lit fume hood at 21°C between 09.00 and 11.00 a.m. The experimental set-up consisted of a 30 x 30 cm plate with 8 x 240 mm LED strips (1 cm between the strips, 6000-6500 K, 12V, Velleman, Belgium) and a 30 x 30 cm diffuser plate (polytethylene terephthalate glycol, 69% light transmission, PyraSied, The Netherlands) on top (6 cm between the plates). A webcam (Logitech c920, Logitech, Switzerland) was placed above the diffuser plate to record the trials with OBS Studio (https://obsproject.com).

Statistical analysis

RTs and accuracy were generated by one experimenter scoring the video recordings manually. RT is the time between the male entering the arena and wing extension, and accuracy is whether the male showed courtship behaviour towards the virgin female (correct), mated female or male (both incorrect). In the first model, effects of difficulty on RTs were assessed with a generalized linear model (GLM) with a Gaussian error distribution, and in which difficulty was included as explanatory factor. In the second model, effects of difficulty on accuracy were assessed with a GLM including a binominal error distribution and which contained difficulty as explanatory factor. Fligner-Killeen tests were used to check for homogeneity of variances, and overdispersion was assessed by comparing the residual deviance and the degrees of freedom. As RTs showed signs of heteroscedasticity, a logarithmic transformation was applied which rendered the data normal. Statistical analysis was executed in R (v.3.4.3, R Core Team, 2018). The car package (Fox and Weisberg 2019) was applied to perform likelihood ratio tests based on χ^2 values, and the ggplot2 package (Wickham 2016) was employed to visualize the RT distributions and frequency plots. Simulated RT distributions were generated using the rtdists package in R (Singmann et al. 2018).

Diffusion models

To fit the models to the data, the experiment included a manipulation of task difficulty. The hypothesis is that the drift rate is higher in the easy than in the difficult condition. The two models that were applied are the EZ diffusion model (Wagenmakers et al. 2007), which is a mathematical data transformation, and the hierarchical drift diffusion model (HDDM; Wiecki et al. 2013), which is a computational model based on hierarchical Bayesian analysis that aims to find the best parameters to describe a particular set of RT distributions. These particular models were used because the EZ model is able to handle small sample sizes, and the HDDM is flexible in incorporating and excluding parameters to generate the best model fit. In addition, the HDDM can also provide the variance in drift rate (s_v) and the variance in non-decision time (s_t) parameters which indicate the variability in the drift rate and non-decision time parameters. The RTs were transformed (i.e. divided by 94.931) to obtain a mean of 500 milliseconds, because the DDM is designed to fit RT distributions with a mean of about 800 ms (Ratcliff and McKoon 2008), and fitting several transformations indicated that data with a mean of 500 ms provided the best model fit (supplementary material 1). The EZ model was applied to the data using the EZ.R package in R (Wagenmakers et al. 2007; supplementary material 2). Notably, because approximately 100 trials are required per subject to provide reliable parameter estimates, and male flies could not be tested that often, in this study all trials within each condition are considered to come from one subject only. Although this method might be unconventional compared to human studies (Ratcliff and Childers 2015), it enabled the computation of reliable parameter estimates. The EZ model includes three assumptions: the RT distributions should be skewed to the right, RT distributions should be identical for correct and incorrect responses, and there should be no bias in the starting point. Assumption checks are presented in the results section (Figure 6) and none of the tests indicated that an assumption was violated.

The HDDM uses hierarchical Bayesian parameter estimation to generate parameter estimates, the posterior distribution of the estimates, and measures of uncertainty for the estimates (supplementary material 3). A main advantage of the HDDM is its flexibility in incorporating and excluding model parameters. Hence, one way to assess which parameters changed due to the difficulty manipulation is by fitting various HDDM models (including different sets of parameters) and checking which model best describes the data (Table 1 and 3). For instance, if the model in which the drift rate depends on difficulty best describes the data, instead of a model in which the thresholds depend on the

manipulation, this may indicate that the drift rate (and not thresholds) changed due to the manipulation. Thus, multiple HDDM models were applied to the data to assess which model best described the data from the current study. The HDDM applied using the HDDM package (v.0.6.0, http://ski.clps.brown.edu/hddm_docs) and executed in Jupyter notebook (https://jupyter.org) with a script written in Python (supplementary material 2). The assumption of convergence (i.e. whether the Markov chain converged with the posterior distribution) was checked with posterior plots (Figures S2-9) and the Gelman-Rubin statistic presented in the results section, which indicated no convergence problems. Furthermore, as the HDDM package also includes a function to provide EZ parameter estimates, EZ estimates were generated by using both the the EZ.R package and the HDDM software.

Results

Part I. Courtship experiment results

Courtship behaviour by male *D. melanogaster* was tested under easy (male vs. virgin female) and difficult (mated vs. virgin female) conditions. In the easy condition, fast responses with high accuracy were expected, whereas in the difficult condition, slower responses with lower accuracy were hypothesized. Courtship initiation towards the virgin female was considered as correct, while courting the mated female or male was seen as incorrect. Accuracy was higher in the easy condition (66.29%) compared to the difficult condition (50.59%) ($\chi^2 = 0.016$, $df = 1$, $p < 0.001$; Figure 4a), which is in line with the hypothesis. However, RTs were on average slower in the easy condition (55.090±5.056 seconds (SEM)) than in the difficult condition (39.482±4.249 seconds (SEM)) ($\chi^2 = 5.400$, $df = 1$, $p = 0.020$; Figure 4b) which is contrary to the hypothesis. The RT distributions for both conditions (Figure 5) show that a large proportion of the males responded fast when two females were presented, whereas the RT distribution for the easy condition is more spread across time. Furthermore, when dividing the males that responded correctly from the males that responded incorrectly, the correctly responding males did not court faster than the incorrectly responding males within each condition ($\chi^2 = 0.950$, $df = 1$, $p = 0.330$; Figure 4b). The mean RT for correct trials in the easy condition is 58.831±6.100 seconds (SEM), and 47.733±8.997 seconds (SEM) for the incorrect trials. In the difficult condition, the average RT is 40.302±5.924 seconds (SEM) for correct trials, and 38.643±6.166 seconds (SEM) for incorrect trials.

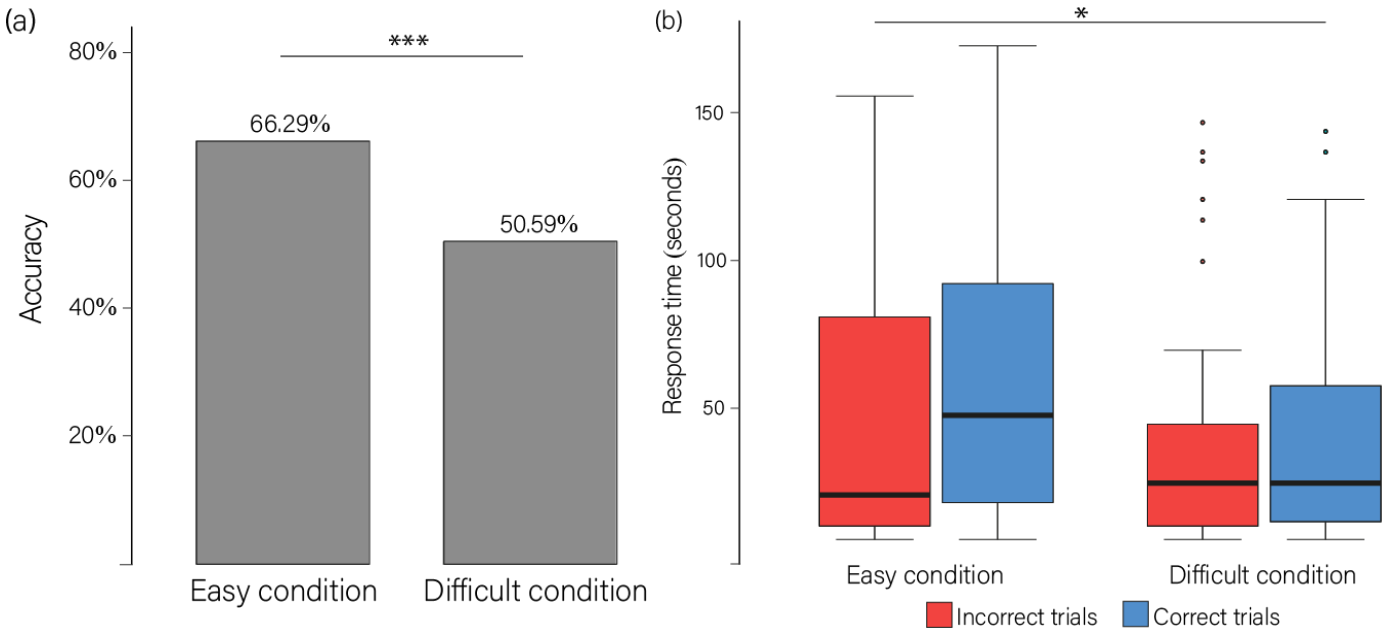


Figure 4. (a) Bar graph showing accuracy for the easy (male vs. virgin female, n = 89) and difficult condition (mated vs. virgin female, n = 85). (b) Boxplot illustrating the RT distribution per condition for correct and incorrect trials.

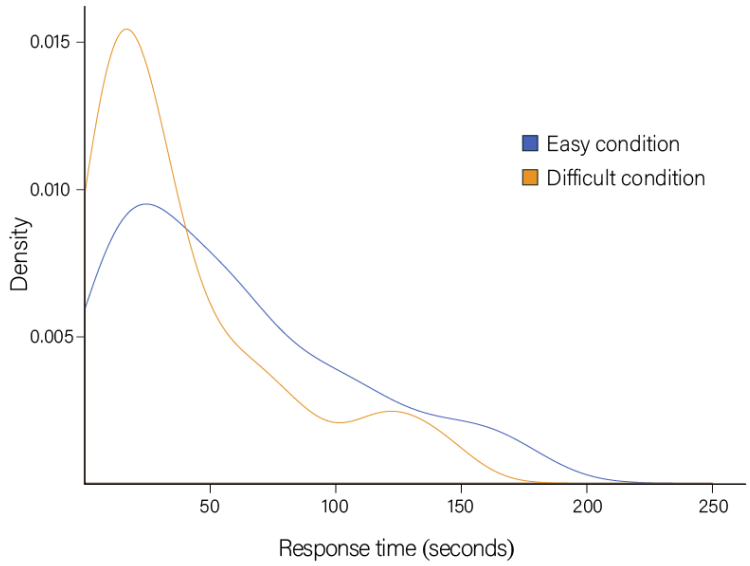


Figure 5. Density graph depicting the RT distributions for the easy (n = 89) and difficult (n = 85) condition including both correct and incorrect responses. Note the characteristic right-skew in the distribution which is specific to response time distributions.

Part II. Model fitting on easy condition data

Because the males in the difficult condition initiated courtship equally often and equally fast towards virgin as mated females, the males did not seem to distinguish between the two females. Hence, it is uncertain whether a decision process occurred as the DDM describes. Therefore, the EZ model and HDDM are fitted first to the easy condition data only and secondly to data from both conditions. In case there was no decision making process in the difficult condition, the models might still fit the easy condition data. Fitting several models to the data (supplementary material 2) indicated that a data transformation to obtain a mean of 500 ms (i.e. RTs were divided by 94.931) provided the best model fit and hence this transformation will be used throughout the model fitting procedure.

EZ model fit on easy condition data

The EZ model (Wagenmakers et al. 2007) provides estimates of the drift rate, threshold, and non-decision time and is based on three assumptions: RT distributions should be skewed to the right, RT distributions should be identical for correct and incorrect responses, and the starting point should be unbiased. Firstly, a distribution plot (Figure 6a) indicates that the RT distributions are right-skewed, which was affirmed by the D'Agostino test (skew = 0.864, $z = 2.685$, $p = 0.007$). Secondly, the RT distributions for correct and incorrect responses were identical, as required (Figure 6a and b; $\chi^2 = 1.141$, $df = 1$, $p = 0.285$). Thirdly, a bias in starting point is identical to a bias in response options, and hence the unbiased starting point assumption was be checked by assessing the RT distribution for each response option (left or right side of the dish). Namely, if the starting point is shifted towards one of the boundaries, the drift rate will reach this boundary faster and hence there is a bias towards that response option. As an example, if there were a bias for the right side, RTs for the right side would be faster in the correct than the incorrect trials (Wagenmakers et al. 2007). This would result in an interaction effect with fast correct responses for one side, and slow correct responses for the other side. As the RT distributions were equal for both options ($\chi^2 = 2.265$, $df = 1$, $p = 0.132$; Figure 6b) there appears to be no bias in starting point. Thus, there were no indications that any of the assumptions were violated.

Applying the EZ model on the easy condition data with the EZ.R package (supplementary material 1) generated the following parameter estimates: 0.045 for the drift rate 'v', 0.151 for the threshold 'a', and 0.106 (i.e. 11.396 seconds) for the non-decision time 'Ter'. Furthermore, the EZ parameter estimates from the HDDM package were 0.4491 for the drift rate 'v', 1.506 for the threshold 'a', and -0.304 (i.e. non-transformed -32.700 seconds). Note that the estimates of the EZ.R and HDDM packages mainly differ due to their noise coefficient. Since the EZ model assumes that the

noise coefficient is 0.1, while the HDDM sets it to 1, a factor 10 scaling difference occurs. Thus, the drift rate from the EZ.R package is 0.151, while the drift rate of the HDDM package is 1.506.

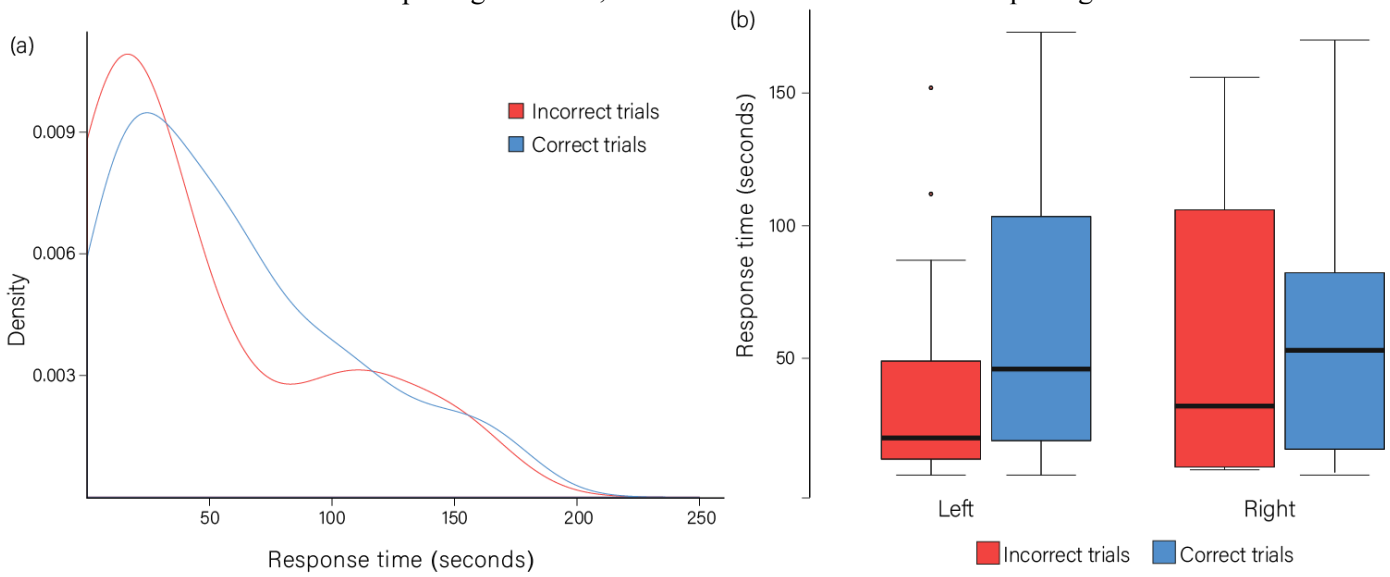


Figure 6. (a) Density plot showing RT distributions for correct and incorrect responses in the easy condition. (b) Boxplot illustrating the RT distribution for correct and incorrect trials for right or left side responses.

In order to visualize the EZ model fit, RT distributions were simulated based on the EZ parameter estimates using the rtdists package in R (Singmann et al. 2018). If the EZ parameter estimates describe the data well, the simulated RT distribution based on these estimates will be similar to the empirical RT distribution. Figure 7a shows that the simulated RT based on the EZ estimates from the EZ.R package has a longer tail, but overall, the RT distributions show a major overlap, indicating that the EZ estimates from the EZ.R package seem to fit the data well. The RT distribution based on EZ estimates derived from the HDDM package predicts fewer fast responses (within 500 ms) than was observed in the empirical data (Figure 7b). In addition, the simulated RT distribution is more right-skewed. On the whole, the distributions largely overlap, indicating a good model fit considering the small amount of observed data (n = 89).

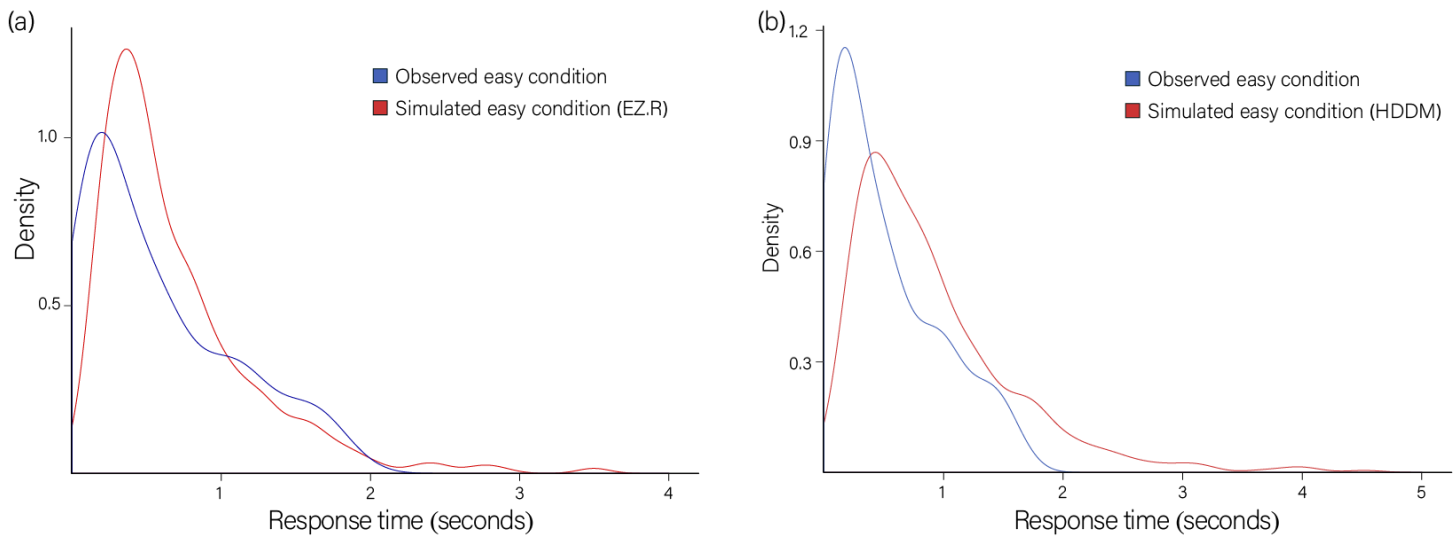


Figure 7. (a) Density plot showing easy condition RT distributions for observed and simulated data, the latter based on EZ parameter estimates which were generated with the EZ.R package. (b) Density plot showing easy condition RT distributions for observed and simulated data, the latter based on EZ parameter estimates derived from the HDDM package. Note that the observed RTs were transformed (i.e. divided by 94.931) to obtain a mean of 500 ms.

HDDM model fit on easy condition data

The HDDM uses hierarchical Bayesian parameter estimation to provide parameter estimates per subject which are informed by the group averages (Wiecki et al. 2013). To find the best parameter estimates, Markov Chain Monte Carlo (MCMC) is used, a class of algorithms to generate random samples from a distribution. However, when using MCMC, the assumption of convergence must be assessed (Sinharay 2003). That is, has the Markov chain converged to the posterior distribution from the Bayesian parameter estimation, so the algorithm can be stopped and its output used for further inferences. Although there is no decisive test to check convergence, a good indication can be provided by posterior plots and the Gelman-Rubin statistic. The Gelman-Rubin statistic indicates no problems with convergence (1.000 for the threshold 'a', 1.000 for the drift rate 'v', and 0.999 for non-decision time 'Ter'. Values close to 1 indicate good convergence, values close to 1.02 indicate convergence problems). In addition, the posterior plots (Figures S2-4) indicate good convergence as well. The posterior trace plots, which display the variability in estimates by the chain, show fairly constant means and variances. Furthermore, the autocorrelation plot, indicating the relationship among the drawn samples (ideally about 0 and at least below 0.5), shows fluctuations around zero. Finally, the marginal posterior histogram displays the distribution of estimates, and the normal distribution in Figures S2-4 indicate good convergence. Thus, there are no indications for convergence problems and hence it is assumed that the MCMC converged with the posterior distribution.

Multiple HDDM models were applied to the easy condition data to determine which model best described the data. Table 1 shows that the HDDM model without outlier correction, variance in drift rate and variance in non-decision time best fitted the data, as it resulted in the smallest deviance information criterion (DIC). DIC indicates the amount of information that is not explained by the model, while it corrects for model complexity, since complex models typically fit the data better. Hence, the model with the lowest DIC describes the data best. Applying the selected HDDM model to the easy condition data generates the following parameter estimates: 0.430 for drift rate 'v', 1.320 for the threshold 'a' and 0.010 (i.e. 1.079 seconds) for non-decision time 'Ter'. The model fit displayed in Figure 8 indicates a good model fit given the relatively small number of trials (n = 89). In addition, posterior predictive checks can be used to assess the uncertainty of parameter estimates by the HDDM. For this, new data are simulated which also provide parameter estimates. That is, every simulation produces its own estimates and hence the collection of simulations provides a distribution of the model estimates. The parameter estimates of the current model can then be compared to this distribution. The posterior predictive statistics (Table S1) indicate that 11 of the 15 model estimates are likely to occur in the posterior distribution of the HDDM.

Table 1. Comparison of HDDM models on easy condition data

	DIC
1. Model with no added parameters	**182.442**
2. Model incl. outlier-correction	186.583
3. Model incl. outlier-correction & St	187.973
4. Model incl. outlier-correction & Sv	188.177
5. Model incl. outlier-correction, Sv, & St	189.837
6. Model incl. St	184.546
7. Model incl. Sv	184.273
8. Model incl. Sv, St	185.701

The deviance information criterion (DIC) indicates the amount of information not accounted for by the HDDM. The lowest DIC indicates the best model fit. Outlier correction entails that the outliers were assigned to an uniform distribution in order to effect the model fit less strong. St is the variance in non-decision time. Sv is the variance in drift rate. The selected model is indicated with a bold font.

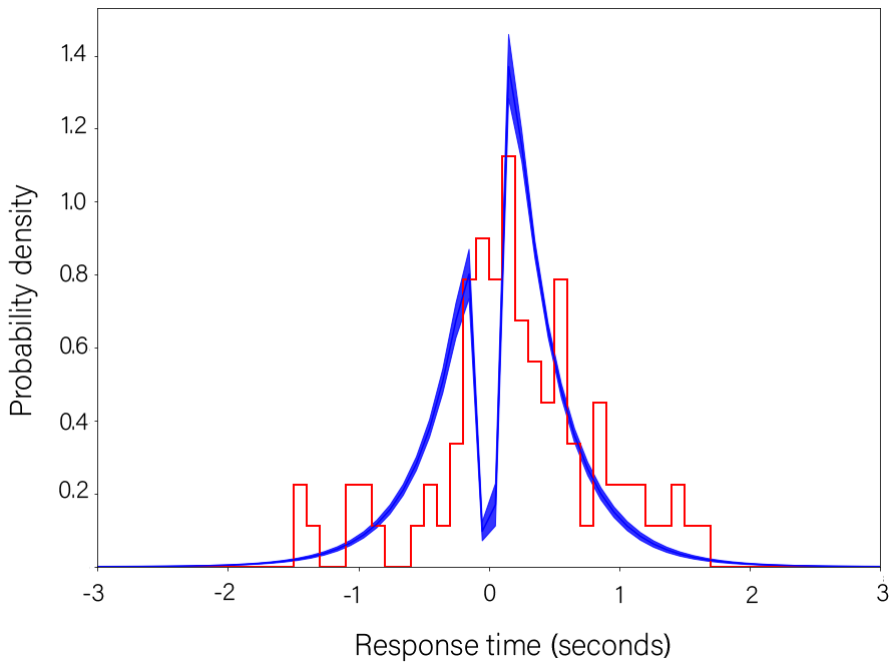


Figure 8. Posterior predictive plot showing the model fit. RT distribution of the *D. melanogaster* easy condition data (n = 89) is given in red, the predictive likelihood is given in blue. RTs for incorrect trials are plotted negatively, so the RT distribution for incorrect trials is given on the left side of zero, while the distribution for correct trials is shown on the right side of zero.

Comparison of EZ and HDDM model fits on easy condition data

Note that the absolute values of the estimates (Table 2) might in themselves not be indicative of how well the model fits the data, the values are informative when they are compared to estimates of treatments where model parameters have been manipulated. However, an exception is the non-decision time estimate, which is the part of the RT not dedicated to the decision process, such as the time it takes to process information and execute a response. In humans, non-decision time typically ranges from 0.3 to 0.7 seconds (Wagenmakers et al. 2007). The non-decision time for *D. melanogaster* would ideally be below 2 seconds, since RTs of 2 seconds have been observed in this project, and DasGupta et al. (2014) reported a *D. melanogaster* non-decision time of 1.452 seconds. Therefore, the non-decision time of 11 seconds which the EZ model from the EZ.R package indicates (Table 2), seems unlikely given that faster RTs were observed. Hence, this estimate might indicate that the model is unable to describe the data set. Similarly, the non-decision time of -32 seconds predicted by the EZ model from the HDDM package indicates that the EZ model might not fit the data as well. On the contrary, the HDDM does provide a plausible non-decision time of 1 seconds, which is close to the non-decision time estimates in humans. The model fits for the EZ models (derived from the EZ.R and HDDM package), and the HDDM (Figure 7-8) show that the model-predicted and observed RT distributions largely overlap, which indicates that the models might be able to describe *D. melanogaster* decision making data.

Table 2. Parameter estimates by the EZ model and the Hierarchical Drift Diffusion Model (HDDM) on easy condition data

	EZ model from EZ.R package	EZ model from HDDM package	HDDM
Drift rate (v)	0.049	0.449	0.423
Threshold (a)	0.151	1.506	1.320
Non-decision time (Ter)	0.106	-0.304	0.010
Non-transformed non-decision time (Ter)	11.340 seconds	-32.700 seconds	1.079 seconds

Part III. Model fitting on both conditions

Note, that because the male responder flies in the difficult condition courted the virgin and mated females equally fast and often, it cannot be assured whether a decision process occurred as the DDM describes. Therefore, cautiousness is required when assessing the model fits on both conditions data.

EZ model fit on both conditions

First of all, none of the EZ assumptions seem to be violated in the data set including both conditions. Figure 10a and the D'Agostino test (skew = 1.048, $z = 3.923$, $p < 0.001$) confirmed that the data were right-skewed. In addition, the RT distributions for correct and incorrect responses showed no difference (Figure 10a; $\chi^2 = 1.457$, $df = 1$, $p = 0.228$). Finally, there was no bias towards one of the two response options ($\chi^2 = 0.383$, $df = 1$, $p = 0.536$; Figure 10d). The EZ parameters for the easy condition are 0.042 for the drift rate 'v', 0.160 for the threshold 'a', and -0.049 (i.e. -3.776 seconds) for non-decision time 'Ter'. For the difficult condition, the EZ parameter estimated from the EZ.R package are 0.002 for drift rate 'v', 0.142 for the threshold 'a' and -0.90 (i.e. -8.508 seconds) for non-decision time 'Ter'. The negative non-decision times (-3,776 seconds for the easy condition and -8,508 seconds for the difficult condition) indicate that the EZ estimates from the EZ.R package might not fit the data well. On the contrary, Figures 9a and b show the EZ model fit from the EZ.R package and illustrate that the shape of the observed and simulated RT distributions are similar and hence indicate a good model fit. The EZ parameters for both condition data using the HDDM package are 0.229 for drift rate 'v', 1.521 for the threshold 'a' and -0.073 (i.e. -6.924 seconds) for non-decision time 'Ter'. The negative non-decision time indicates a bad model fit. In addition, no RT distribution could be provided with the rtdists function, since the combination of parameters could not produce a RT distribution.

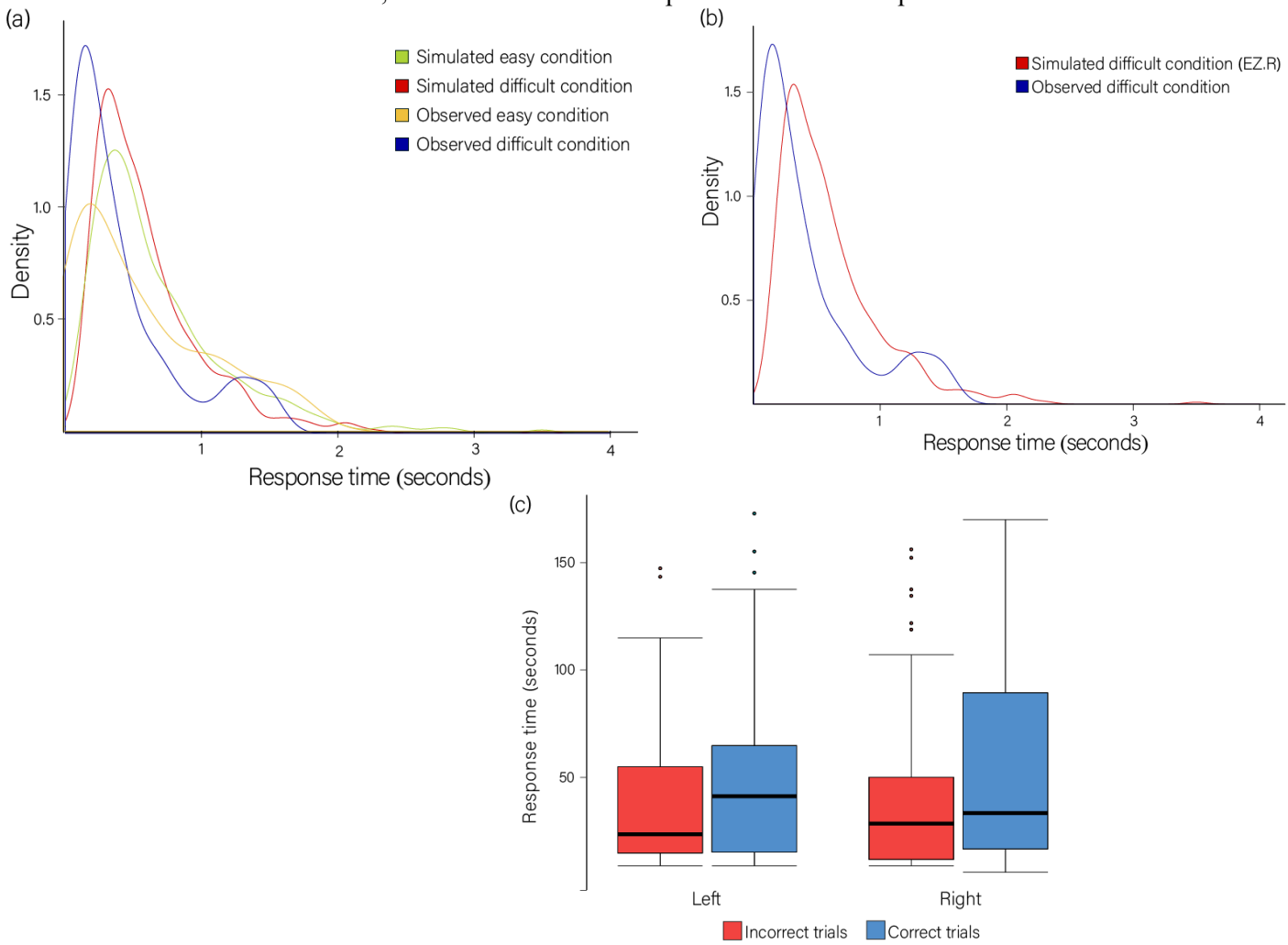


Figure 9. (a) Density plot showing RT distributions from observed and simulated data based on EZ estimates with EZ.R package. (b) Density plot showing the difficult condition RT distributions for the observed and simulated data based on EZ.R package estimates. (c) RT distribution for correct and incorrect trials for right or left side responses based on both condition data.

HDDM model fit on both conditions

There are no indications the Markov chain did not converge with the posterior distribution based on the the Gelman-Rubin statistic (1.000 for both the easy and difficult condition thresholds 'a', 1.001 for the easy condition drift rate 'v', 1.000 for the difficult condition drift rate, and 1.000 for non-decision time 'Ter', and 1.002 for the variance in non-decision time 'St') or the posterior plots (Figures S5-9).

Fitting multiple HDDM models to the data set with both conditions indicated that the two HDDM models that best describe the data are the models in which only the thresholds, and both the drift rate and thresholds depend on the difficulty manipulation (Table 3). In order to assess whether the experimental manipulation significantly altered the drift rate and thresholds estimates, the model in which both the drift rate and thresholds depend on the manipulation was used. This model estimates that the drift rate estimates are higher for the easy condition (0.409) than the difficult condition (0.034), which is in line with the hypothesis that the drift rate is slower in the difficult condition. Furthermore, the threshold estimate is higher for the easy condition (1.417) than the difficult condition (1.205), which is contrary to the hypothesis that the thresholds would remain constant. To assess whether the drift rate and thresholds estimates differed significantly between the conditions, a Bayesian posterior probability test was performed using the HDDM package. This test assessed the probability that the drift rate and thresholds were greater in one condition than the other (supplementary material 1). The test indicated that the probability that the drift rate of the easy condition is greater than the drift rate of the difficult condition is 0.964 (Figure 10a). This means that there is 96.4% probability that the drift rate is higher in the easy than in the difficult condition. In addition, the probability that the threshold heights are greater in the easy than in the difficult condition is 0.998 (Figure 10b). Hence, there is 99.8% chance that the boundary heights are greater in the easy than in the difficult condition. These results indicate that the HDDM estimates that both the drift rate and threshold heights decreased due to the difficulty manipulation.

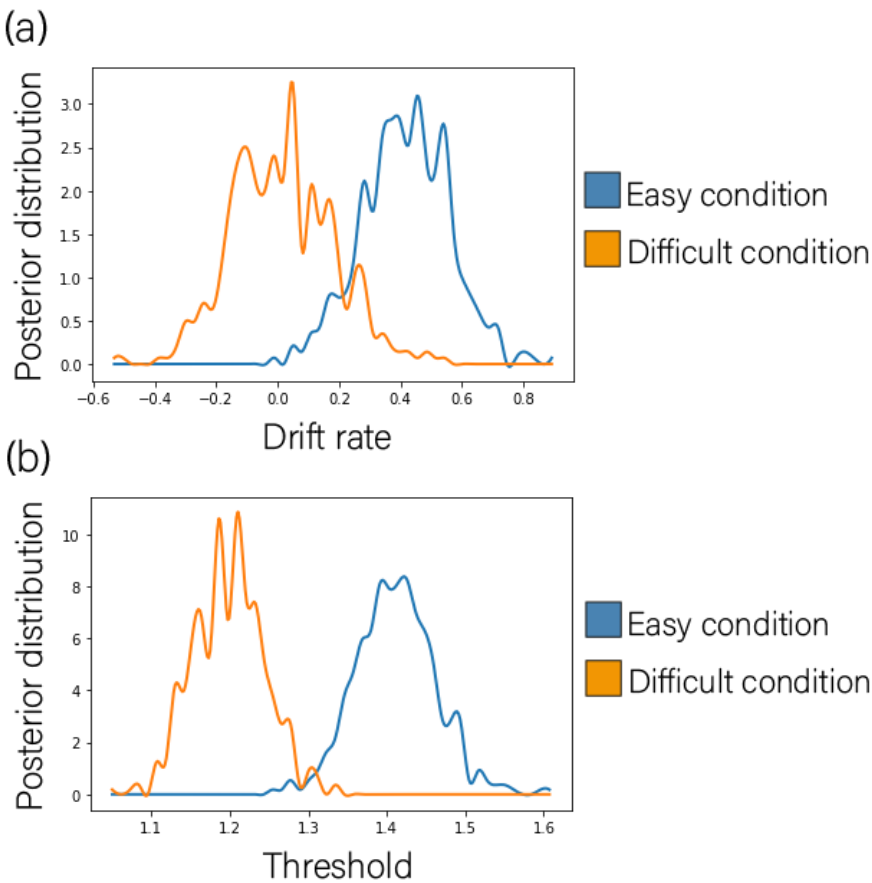


Figure 10. HDDM posterior probability plots illustrating (a) differences in drift rate for the easy and difficult condition, and (b) differences in the threshold heights between the easy and difficult condition.

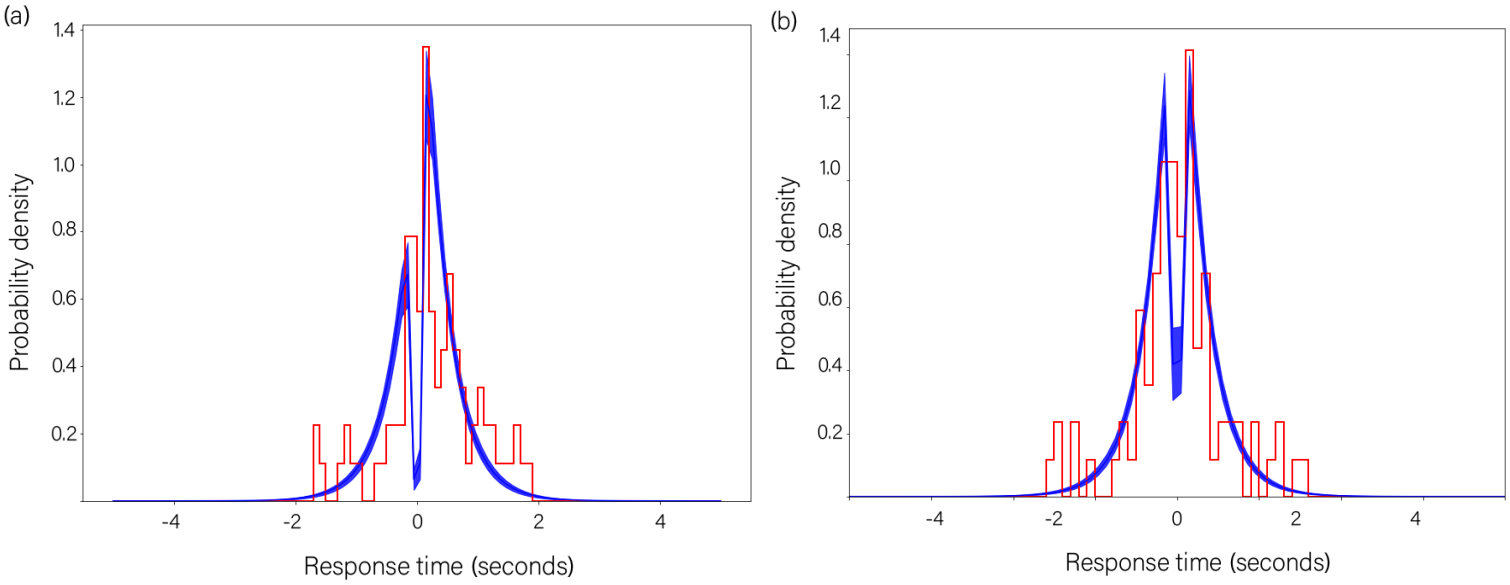


Figure 11. (a) Posterior predictive plot showing the HDDM model fit on easy condition data (n = 89). (b) Posterior predictive plot showing the HDDM model fit on difficult condition data (n = 85). RT distribution of the *D. melanogaster* data is presented in red, the predictive likelihood is given in blue RTs for incorrect trials are plotted negatively, so the RT distribution for incorrect trials is given on the left side of zero, while the distribution for correct trials is shown on the right side of zero.

The HDDM model fit is displayed in Figure 11 and shows that the model describes the data relatively well given the small number of trials. The small amount of observations is visible by the red line, which indicates the observed RTs, dropping to zero at some points because there were no trials with that RT observed. In addition, the posterior predictive statistics (Table S2) indicate that 13 out of the 15 estimates were likely to occur in the posterior distribution of the HDDM. Since the non-decision time of 830 ms (Table 4) is similar to the non-decision times often found in humans (Wagenmakers et al. 2007), this may indicate that the HDDM might be able to fit the data from both conditions.

Comparison of EZ and HDDM model fits on both condition data

Firstly, the EZ models (both using the EZ.R and HDDM package) provide unlikely non-decision time estimates that range from -3.323 to -8.508 seconds. However, the HDDM estimates the non-decision time to be 0.830 seconds, which is close to the non-decision time range of 0.3 to 0.7 seconds typically found in humans (Wagenmakers et al. 2007). In addition the EZ model fit from the EZ.R package displayed in Figure 9 shows that the simulated RT distributions largely overlap with the observed RT distributions, which indicates a good model fit. Besides, the model fit of the HDDM to both condition data shows overlapping RT distributions as well (Figure 10). Importantly, both the EZ model and HDDM indicate that the drift rate is higher in the easy condition than in the difficult condition (Figure 11a), which is in line with the hypothesis that in the difficult condition a smaller drift rate would be observed. Furthermore, both the EZ and HDDM models describe a small decrease in threshold height in the difficult condition (Figure 11b). This is unlike the hypothesis that the threshold would remain constant across trials. Thus, the EZ model and HDDM both indicate a decrease in drift rate and threshold heights in the difficult condition.

Table 3. Comparison of HDDM models on both condition data

	DIC
Models in which the drift rate 'v' depends on the difficulty manipulation	
1. No added parameters	346.712
2. Incl. outlier-correction	351.692
3. Incl. outlier-correction, St	338.710
4. Incl. outlier-correction, Sv	353.671
5. Incl. outlier-correction, Sv, St	353.671
6. Incl. St	345.868
7. Incl. Sv	347.908
8. Incl. Sv, St	346.586
Models in which the thresholds 'a' depends on the difficulty manipulation	
9. No added parameters	340.332
10. Incl. outlier-correction	346.206
11. Incl. outlier-correction, St	347.128
12. Incl. outlier-correction, Sv	343.541
13. Incl. outlier-correction, Sv, St	347.128
14. Incl. St	333.088
15. Incl. Sv	341.530
16. Incl. Sv, St	341.590
Models in which the drift rate 'v' and thresholds 'a' depend on the difficulty manipulation	
17. No added parameters	338.950
18. Incl. outlier-correction	345.433
19. Incl. outlier-correction, St	345.433
20. Incl. outlier-correction, Sv	340.761
21. Incl. outlier-correction, Sv, St	339.709
22. Incl. St	**334.536**
23. Incl. Sv	340.631
24. Incl. Sv, St	343.465

The deviance information criterion indicates (DIC) the amount of information not accounted for by the HDDM. The lowest DIC indicates the best model fit. Outlier correction entails that the outliers were assigned to an uniform distribution in order to effect the model fit less strong. St is the variance in non-decision time. Sv is the variance in drift rate. The selected model is indicated with a bold font.

Table 4. Parameter estimates by the EZ model and the Hierarchical Drift Diffusion Model (HDDM) based on data from both conditions

	EZ model from EZ.R package	EZ model from HDDM package	HDDM
Drift rate (v)		0.229	
Easy condition	0.042		0.409
Difficult condition	0.002		0.034
Threshold (a)		1.521	
Easy condition	0.160		1.416
Difficult condition	0.142		1.207
Non-decision time (Ter)		-0.073	0.009
Easy condition	-0.040		
Difficult condition	-0.090		
Non-transformed non-decision time (Ter)		-6.924 seconds	0.830 seconds
Easy condition	-3.776 seconds		
Difficult condition	-8.508 seconds		

Discussion

This study tested whether the DDM can describe decision making in *Drosophila melanogaster*. Applying the HDDM and EZ model to the data indicates that both the EZ model and HDDM seem able to describe decision making in *D. melanogaster*. Besides, both models indicate that the drift rate is lower in the difficult than in the easy condition, which is in line with Ratcliff's theory that the drift rate is faster in the easy condition (Ratcliff and McKoon 2008). Unlike the hypothesis, the models estimate that the thresholds also lowered. This indicates that either the models do not accurately describe the data, or that the difficulty manipulation influenced other factors besides difficulty. In addition, it must be noted that the data set contained fewer trials than commonly used in other studies (Cavanagh et al. 2011; Huang-Pollock et al. 2017), making it likely that the data set was too small to determine whether the EZ and HDDM models can describe *D. melanogaster* decision making. Nonetheless, the current study suggests that the EZ model and HDDM might be able to describe decision making in *Drosophila melanogaster*.

Three main conclusions can be derived from this research. Firstly, the HDDM and EZ models seem able to describe *D. melanogaster* decision making (Table 2 and 4, Figure 7-9 and 11). Even though the RT distributions from the EZ models and HDDM are all similar to the empirical RT distribution, the HDDM non-decision time estimates are closer to those previously found in humans (Wagenmakers et al. 2007) than the negative non-decision time estimates by the EZ models. These negative estimates from the EZ models might be due to some relatively fast and slow responses in the current data, since Ratcliff and Childers (2015) concluded that the EZ model is unable to fit data with very fast or slow observations while the HDDM is largely resistant to outliers. Although HDDM seems well able to describe the current data set, the low number of observations does hamper its model fit. This can be derived from the box-shaped red lines which drop to zero at various points (Figures 9 and 11) when there are no observations for those RTs. Thus, while the current study indicates that the HDDM might fit the data, larger data sets are required to determine with certainty whether the HDDM can describe decision making in *Drosophila melanogaster*.

The second main conclusion is that the male flies responded faster in the difficult than easy condition, contrary to the hypotheses. The DDM predicts that in the difficult condition, information is more ambiguous, leading to a slower drift rate, slower responses, and lower accuracy. The third conclusion is that accuracy is higher in the easy than difficult condition, which is in line with the hypothesis. However, the males in the difficult condition were not expected to respond to both females equally often and equally fast. A possible explanation on why the males did not seem to distinguish between the two females is the last male sperm precedence phenomenon. That is, most of the offspring will be sired by the male that last mated the female (Laturney et al. 2018). Thus, the mated female might not be an 'incorrect' option for the responder male, because the responder male is the only male present in the assay, and hence it would be the last male to mate with the mated female. Although Laturney and Billeter (2016) found that males spend more time courting females without a mating plug than females with mating plug, the males still courted the females with a mating plug. Thus, both Laturney and Billeter's study and the current project indicate that males still court mated females with a mating plug when they can also choose a female without mating plug. This might indicate that males do not consider females with a mating plug as an 'incorrect' response option themselves. Thus, instead of presenting a mated and a virgin female in the difficult condition, a better option might be to present a *D. melanogaster* virgin female and a virgin female of a different species, such as *Drosophila suzukii* or *Drosophila elegans*. *D. melanogaster* males typically prefer females from their own species, but may court females form other species as well (Cobb and Jallon 1990). An alternative explanation on why the males courted both females equally often and fast is that the courtship song does not reflect a decision but rather a sampling method. That is, at the moment the male performs the courtship song, the male might not have decided to mate with that fly yet, but rather tests the receptivity of the encountered fly. The latter option might also explain why 30% of the males in the easy condition performed courtship initiation towards a male fly instead of a virgin female.

Two main difficulties were encountered during this project in combining computational models and *D. melanogaster* research. First of all, it appeared difficult to design an experiment for flies that is resembles human experiments on fast and two-choice decision making. Although several experiments exist on rapid decision making in *D. melanogaster*—such as the odour discrimination task used by

DasGupta and his colleagues (2014), or the Flywalk (Steck et al. 2012) in which flies' responses to specific odours are assessed—several of them rely on walking behaviour (towards or away from the presented cue). The main difficulty with these experiments is that the behaviour indicating a decision is the same as the preceding and following behaviours (i.e. walking), and hence it is difficult to determine when exactly the fly made a decision. This difficulty might be circumvented by focusing on courtship or mating behaviour, in which every step of the process is distinct from the others (Figure 2) and hence the timing of these behaviours might be easily determined. The second problem encountered in this study was the difficulty in determining whether a specific behaviour genuinely reflected the process of interest. For instance, in this project the courtship song was used to indicate that the male had decided which fly it wanted to court. However, the results from the difficult condition question whether the courtship song reflects a decision, and suggest that it might in fact function as a sampling method. This example illustrates the difficulty in choosing which behaviour reflects a decision.

Since the current study shows that the DDM might be able to describe decision making in *D. melanogaster*, future research could further study whether *Drosophila melanogaster* is a suitable model organism for testing the DDM. This might be assessed by manipulating model parameters, such as the drift rate or threshold, and testing whether behaviour changes according the DDM predictions. The current study aimed to alter the drift rate by manipulating task difficulty, but this manipulation might not have succeeded, because the males responded equally often and fast to both females in the difficult condition. Future studies might for instance use different stimuli flies to change task difficulty, such as females from different species as discussed earlier in the discussion. In addition, the threshold heights might be manipulated using *D. melanogaster* lines with altered dopaminergic (DA) activity. Zhang et al. (2018) indicated that DA alters the motivation of the males to court and thereby affects the likelihood of courtship initiation. Hypothetically, males which are more motivated to court require less information to perform the courtship song, which would result in lower thresholds estimates. Thus, future experiments might opt to change the threshold parameter by testing transgenic flies with altered DA levels (Friggi-Grelin et al. 2003).

In case future research shows that the DDM is able to explain decision making in *D. melanogaster*, causal and precise (genetic) tools available in this species might be used to find neural evidence for the DDM. For instance, Groschner et al. (2018) demonstrated that the αβ core Kenyon cells in the *D. melanogaster*'s mushroom body integrate olfactory information from subthreshold depolarizations until an action potential is evoked. Thus, future research could for instance assess whether these αβ Kenyon cells might function as a threshold. If activation of a specific proportion of αβ Kenyon cells leads to a response execution, then the threshold might hypothetically consist of a specific amount of firing cells, where a higher threshold would mean that more αβ Kenyon cells need to fire before a response is executed. This idea might be investigated by testing the amount of firing cells shortly before a decision is made, or by activating different amounts of αβ Kenyon cells with optogenetics and test which amount leads to a response execution.

Furthermore, future research might investigate whether the flies' trajectory can be used as a proximate for the drift rate. Previously, evidence accumulation has been linked to populations of frontal eye field (FEF) neurons (Purcell et al. 2010), EEG components (Ratcliff et al. 2009), and computer mouse trajectories (Lepora and Pezzulo 2015). Given that the distance between flies determines which sensory information the male can sense, trajectories might be used to estimate the sensory input at specific moments in time. For instance, volatile chemosensory cues as cVA can be sensed over a longer distance than non-volatile cues, such as 7-T or 7,11-ND, for which the male needs to touch the other fly (Griffith and Ejima 2009; Depetris-Chauvin et al. 2015). Thus, it can be estimated where in the trajectory the male was able to detect cVA, and at what point it could sense 7-T. In addition, this information might be used to estimate the probability of decision outcomes at specific distances. For instance, the percentage of males performing courtship initiation after tapping a virgin female could be used to indicate the probability of male courtship initiation after sensing a virgin female's non-volatile cues.

Finally, the experiment presented in this study might also be used to assess how genes influence *D. melanogaster* decision making. In particular, genes that have a homologue in humans might be of interest (Pandey and Nichols 2011). For example, a human gene causing Parkinson's disease (PD) called *Parkin* (Fortini et al. 2000) could be investigated, because PD patients are notorious for their impulsive decision making (Averbeck et al. 2014) and *D. melanogaster* research might provide

detailed knowledge on how this gene influences choice behaviour. In addition, behavioural tests might indicate whether humans and flies with PD symptoms show similar changes in decision behaviour, and specifically whether the DDM parameter estimates change similarly for both humans and flies. In case DDM parameters change similarly in humans and flies, this could indicate that *Drosophila melanogaster* might be a suitable model organism to study the DDM.

Bibliography

Ashby GF, Helie S. 2011. A tutorial on computational cognitive neuroscience: Modeling the neurodynamics of cognition. J Math Psychol. 55(4):273–289. doi:10.1016/j.jmp.2011.04.003.

Averbeck BB, Sullivan SSO, Djamshidian A. 2014. Impulsive and Compulsive Behaviors in Parkinson ' s Disease. doi:10.1146/annurev-clinpsy-032813-153705.

Billeter JC, Atallah J, Krupp JJ, Millar JG, Levine JD. 2009. Specialized cells tag sexual and species identity in Drosophila melanogaster. Nature. 461(7266):987–991. doi:10.1038/nature08495.

Billeter JC, Rideout EJ, Dornan AJ, Goodwin SF. 2006. Control of Male Sexual Behavior in Drosophila by the Sex Determination Pathway. Curr Biol. 16(17):766–776. doi:10.1016/j.cub.2006.08.025.

Cavanagh JF, Wiecki T V., Cohen MX, Figueroa CM, Samanta J, Sherman SJ, Frank MJ. 2011. Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nat Neurosci. 14(11):1462–1467. doi:10.1038/nn.2925.

Cobb M, Jallon JM. 1990. Pheromones, mate recognition and courtship stimulation in the Drosophila melanogaster species sub-group. Anim Behav. 39(6):1058–1067. doi:10.1016/S0003-3472(05)80778-X.

DasGupta S, Ferreira CH, Miesenboč̌k G. 2014. FoxP influences the speed and accuracy of a perceptual decision in Drosophila. Science (80-). 344(6186):901–904. doi:10.1126/science.1252114.

Depetris-Chauvin A, Galagovsky D, Grosjean Y. 2015. Chemicals and chemoreceptors: ecologically relevant signals driving behavior in Drosophila. Front Ecol Evol. 3(April). doi:10.3389/fevo.2015.00041.

Doya K, Shadlen MN. 2012. Decision Making. Curr Opin Neurobiol. 22:911–013. doi:10.1080/00207599208246888.

Fortini ME, Skupski MP, Boguski MS, Hariharan IK. 2000. A survey of human disease gene counterparts in the Drosophila genome. J Cell Biol. 150(2):23–29.

Fosco WD, White CN, Hawk LW. 2017. Acute Stimulant Treatment and Reinforcement Increase the Speed of Information Accumulation in Children with ADHD. J Abnorm Child Psychol. 45(5):911–920. doi:10.1007/s10802-016-0222-0.

Fox J, Weisberg S. 2019. An {R} Companion to Applied Regression. Third. Thousand Oaks {CA}: Sage.

Friggi-Grelin F, Coulom H, Meller M, Gomez D, Hirsh J, Birman S. 2003. Targeted gene expression in Drosophila dopaminergic cells using regulatory sequences from tyrosine hydroxylase. J Neurobiol. 54(4):618–627. doi:10.1002/neu.10185.

Glimcher PW, Fehr E. 2014. Introduction: A Brief History of Neuroeconomics: Neuroeconomics. Neuroeconomics.:xvii–xxviii. doi:10.1016/B978-0-12-416008-8.00035-8.

Goschke T. 2014. Improving study design for antidepressant effectiveness assessment. Int J Methods Psychiatr Res. 23(S1):41–57. doi:10.1002/mpr.

Griffith LC, Ejima A. 2009. NIH Public Access. :394–398. doi:10.1111/j.1749-6632.2009.04367.x.Multimodal.

Groschner LN, Chan Wah Hak L, Bogacz R, DasGupta S, Miesenböck G. 2018. Dendritic Integration of Sensory Evidence in Perceptual Decision-Making. Cell. 173(4):894-905.e13. doi:10.1016/j.cell.2018.03.075.

Huang-Pollock C, Ratcliff R, McKoon G, Shapiro Z, Weigard A, Galloway-Long H. 2017. Using the Diffusion Model to Explain Cognitive Deficits in Attention Deficit Hyperactivity Disorder. J Abnorm Child Psychol. 45(1):57–68. doi:10.1007/s10802-016-0151-y.

Kim S, Lee HS, Park Y. 2017. Perinatal exposure to low-dose imidacloprid causes ADHD-like symptoms: Evidences from an invertebrate model study. Food Chem Toxicol. 110(August):402–407. doi:10.1016/j.fct.2017.10.007.

Laturney M, Billeter JC. 2016. Drosophila melanogaster females restore their attractiveness after mating by removing male anti-aphrodisiac pheromones. Nat Commun. 7:1–11. doi:10.1038/ncomms12322.

Laturney M, van Eijk R, Billeter J-C. 2018. Last male sperm precedence is modulated by female remating rate in Drosophila melanogaster . Evol Lett. 2(3):180–189. doi:10.1002/evl3.50.

Lepora NF, Pezzulo G. 2015. Embodied Choice: How Action Influences Perceptual Decision Making. PLoS Comput Biol. 11(4):1–22. doi:10.1371/journal.pcbi.1004110.
Moloney A, Sattelle DB, Lomas DA, Crowther DC. 2010. Alzheimer's disease: Insights from Drosophila melanogaster models. Trends Biochem Sci. 35(4):228–235. doi:10.1016/j.tibs.2009.11.004.
Moustafa AA, Kéri S, Somlai Z, Balsdon T, Frydecka D, Misiak B, White C. 2015. Drift diffusion model of reward and punishment learning in schizophrenia: Modeling and experimental data. Behav Brain Res. 291:147–154. doi:10.1016/j.bbr.2015.05.024.
Pandey UB, Nichols CD. 2011. Human Disease Models in. Pharmacol Rev. 63(2):411–436. doi:10.1124/pr.110.003293.411.
Pirrone A, Dickinson A, Gomez R, Stafford T, Milne E. 2017. Understanding perceptual judgment in autism spectrum disorder using the drift diffusion model. Neuropsychology. 31(2):173–180. doi:10.1037/neu0000320.
Purcell BA, Heitz RP, Cohen JY, Schall JD, Logan GD, Palmeri TJ, Purcell BA, Heitz RP, Cohen JY, Schall JD, et al. 2010. Psychological Review Neurally Constrained Modeling of Perceptual Decision Making Neurally Constrained Modeling of Perceptual Decision Making. doi:10.1037/a0020311.
Ratcliff R. 1978. A theory of memory retreival. Psychol Rev. 85(2):59–108. doi:10.1037/h0021465.
Ratcliff R, Childers R. 2015. Individual differences and fitting methods for the two-choice diffusion model of decision making. Decision. 2(4):237–279. doi:10.1037/dec0000030.
Ratcliff R, McKoon G. 2008. The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. Neural Comput. 4(164):1–44. doi:10.1126/scisignal.2001449.Engineering.
Ratcliff R, Philiastides MG, Sajda P. 2009. Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. 106(16):6539–6544.
Ratcliff R, Smith PL, Brown SD, McKoon G. 2016. Diffusion Decision Model: Current Issues and History. Trends Cogn Sci. 20(4):260–281. doi:10.1016/j.tics.2016.01.007.
Sanfey AG. 2007. Decision Neuroscience: New directions in studies of judgement and decision making. Curr Dir Psychol Sci. 16(3):151–155. doi:10.1111/j.1467-8721.2007.00494.x.
Shao L, Shuai Y, Wang J, Feng S, Lu B, Li Z, Zhao Y, Wang L, Zhong Y. 2011. Schizophrenia susceptibility gene dysbindin regulates glutamatergic and dopaminergic functions via distinctive mechanisms in Drosophila. Proc Natl Acad Sci. 108(46):18831–18836. doi:10.1073/pnas.1114569108.
Simpson JH, Looger LL. 2018. A diffusion equation approach to spin diffusion in biomolecules. Genetics. 208:1291–1309. doi:https://doi.org/10.1534/genetics.117.300228.
Singmann H, Brown S, Gretton M, Heathcote A. 2018. rtdists: Response Time Distributions.
Sinharay S. 2003. Assessing Convergence of the Markov Chain Monte Carlo Algorithms: A Review. Res Dev Div Princet.
Smith PL, Ratcliff R. 2004. Psychology and neurobiology of simple decisions. Trends Neurosci. 27(3):161–168. doi:10.1016/j.tins.2004.01.006.
Sokolowski MB. 2001. Genetics meets behaviour. Genetics. 2(November). doi:10.1038/35098592.
St Johnston D. 2002. The art and design of genetic screens: Drosophila melanogaster. Nat Rev Genet. 3:176–88. doi:10.1038/nrg751.
Steck K, Veit D, Grandy R, Badia SBI, Mathews Z, Verschure P, Hansson BS, Knaden M. 2012. A high-throughput behavioral paradigm for Drosophila olfaction - The Flywalk. Sci Rep. 2:1–10. doi:10.1038/srep00361.
Sun R. 2012. Introduction to Computational Cognitive Modeling. In: The Cambridge Handbook of Computational Psychology. p. 3–20.
vanVugt MK, Simen P, Nystrom LE, Holmes P, Cohen JD. 2012. EEG oscillations reveal neural correlates of evidence accumulation. Front Neurosci. 6(JULY):1–13. doi:10.3389/fnins.2012.00106.
Venken KJT, Sarrion-Perdigones A, Vandeventer PJ, Abel NS, Christiansen AE, Hoffman KL. 2016. Genome engineering: Drosophila melanogaster and beyond. Wiley Interdiscip Rev Dev Biol. 5(2):233–267. doi:10.1002/wdev.214.
Wagenmakers E-J, Van der Maas HLJ, Grasman RPPP. 2007. An EZ-diffusion model for response time and accuracy. Psychon Bull Rev. 14(1):3–22.
White CN, Ratcliff R, Vasey MW, McKoon G. 2010. Using diffusion models to understand clinical

disorders. J Math Psychol. 54(1):39–52. doi:10.1016/j.jmp.2010.01.004.
Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
Wiecki T V., Sofer I, Frank MJ. 2013. HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. Front Neuroinform. 7(August):1–10. doi:10.3389/fninf.2013.00014.
Zhang SX, Miner LE, Boutros CL, Rogulja D, Crickmore MA. 2018. Motivation, Perception, and Chance Converge to Make a Binary Decision. Neuron. 99(2):376-388.e6. doi:10.1016/j.neuron.2018.06.014.

Supplementary material 1. Scripts for model fitting

```
# R script for EZ model fit on easy condition data

# Prepare & Check data
setwd("~/set/to/working_directory")
courtdat <- read.csv("Data_file.csv")
names(courtdat) <-
c("trial","condition","accuracy","rt","date","side","time","rtold")
print(courtdat)
courtdat$condition <- as.factor(courtdat$condition)
levels(courtdat$condition)
courtdat$rt <- as.numeric(courtdat$rt)
courtdat$side <- as.factor(courtdat$side)
levels(courtdat$side)
str(courtdat)
summary(courtdat)


# Conditions & variables with correct or incorrect trials only
Easycond <- courtdat[courtdat$condition=="Male_Virgin",]
CorrEasyCond <- Easycond[easycond$accuracy==1,]
IncorrEasyCond <- Easycond[easycond$accuracy==0,]


# DESCRIPTIVE DATA

# Mean accuracy, rt and variance
# Variables with acc, rt and varrt are created to use in the model fit

# Accuracy
meanAccEasy <- mean(Easycond[,3])
print(meanAccEasy)

# RT
meanRteasy <- mean(Easycond[,4])
meanRtCorreasy <- mean(CorrEasyCond[,4])
meanRtIncorreasy <- mean(IncorrEasyCond[,4])
print(meanRteasy)
print(meanRtCorreasy)
print(meanRtIncorreasy)

# Variance
varRteasy <- var(Easycond[,4])
varRtCorreasy <- var(CorrEasyCond[,4])
varRtIncorreasy <- var(IncorrEasyCond[,4])
print(varRteasy)
print(varRtCorreasy)
print(varRtIncorreasy)


# FIGURES

# Plot rt histograms for correct or incorrect trials
plot(density(CorrEasyCond[,4]),xlab = "rt(sec)", main ="rt histogram Male
vs. Virgin condition (correct trials)")
plot(density(IncorrEasyCond[,4]),xlab = "rt(sec)", main ="rt histogram Male
vs. Virgin condition (incorrect trials)")
```

```
# Plot accuracy histogram
plot(density(Easycond[,3]),xlab = "Accuracy (percentage of males that made
the correct decision)", main = "Accuracy histogram Male vs. Virgin
condition")

# Plot rt distribution for correct & incorrect trials
courtdat$accuracy <- as.factor(courtdat$accuracy)
levels(courtdat$accuracy)
library(ggplot2)
theme_set(theme_bw())
theme_bw <- theme_update(panel.grid.minor=element_blank(),
                         panel.grid.major=element_blank(),
                         panel.background=element_blank(),
                         panel.border=element_blank(),
                         legend.position="bottom",
                         legend.title = element_blank(),
                         axis.line.x=element_line(size=0.75),
                         axis.line.y=element_line(size=0.75),
                         axis.text.x=element_text(colour="black", size=11),
                         axis.text.y=element_text(colour="black", size=11),
                         axis.title.x=element_text(size=13),
                         axis.title.y=element_text(size=13),
                         axis.ticks.x=element_line(colour="black",size=0.5,
                         axis.ticks.length=unit(0.3,"cm") )

ggplot(courtdat, aes(rt,colour=accuracy)) +geom_density() +xlim(0,3.5)
ggplot(Easycond, aes(rt,colour=accuracy)) +geom_density() +xlim(0,3.5)

# Bar graph on accuracy
df <- data.frame(cond=c("Male vs. Virgin Female","Mated vs.Virgin Female"),
per=c(66.29, 100))
head(df)
library(scales)
ggplot(data=df, aes(x=cond, y=per)) + geom_bar(stat = "identity", width =
0.5, colour="darkgreen", fill="dodgerblue1") +
  ggtitle("Accuracy per condition") +
  ylab("Percentage of trials correct") + ylim(0, 100)

# Boxplot on RT per condition & accuracy
ggplot(courtdat, aes(y=rtold, x=condition, colour=accuracy)) +
stat_boxplot(geom = 'errorbar') + geom_boxplot(fatten=5)


---------------------------------------------------------------------

# FIT DATA TO EZ DIFFUSION MODEL

# EZ Check 1. RT distribution should be right-skewed
# First, Visual inspection
ggplot(courtdat, aes(rtold,colour=accuracy)) +geom_density() +xlim(0,250)
# Second, D'Agostino skewness test
library(moments)
agostino.test(courtdat[courtdat[,3]==1,4])
# Thus, all data have a skewness


# Check2. Correct & error RTs should be equally fast
# Look at table with RTs
mean(CorrEasyCond[,4])
mean(IncorrEasyCond[,4])
```

```
# GLM
Model_IdenticalRT_both <- glm(courtdat$rt~courtdat$accuracy, data =
courtdat)
summary(Model_IdenticalRT_both)
Anova(Model_IdenticalRT_both)
par(mfrow=c(2,2))
plot(Model_IdenticalRT_both)
par(mfrow=c(1,1))

# Check3. There should be no bias in RT for one of the two choice
alternatives

# GLM
Model_Side <- glm(easycond$rt~easycond$side, data = courtdat)
summary(Model_Side)
Anova(Model_Side)
par(mfrow=c(2,2))
plot(Model_Side)
par(mfrow=c(1,1))
# Plot

ggplot(courtdat, aes(y=rtold, x=side, colour=accuracy)) + stat_boxplot(geom
= 'errorbar') + geom_boxplot(fatten=5)


# Perform EZ model fit
# Output: drift rate, threshold, nondecision time
source("EZ.R")
parsEasyCond <- get.vaTer(meanAccEasy,varRteasy,meanRTeasy)
parsDiffCond <- get.vaTer(meanAccDiff,varRtdiff,meanRTdiff)
print(parsEasyCond)
print(parsDiffCond)


---------------------------------------------------------------------------

# RT DISTRIBUTIONS

# Simulate RT distributions based on EZ estimates generated by EZ.R
# Load package
library(rtdists)

# Data frame with the parameter estimates of the EZ model (input is output
of previous get.vaTer function)
SimEasyDat <- rdiffusion(500, a= 0.16044020, v= 0.04215527, t= 0.03977583,
s= 0.1)
SimDiffDat <- rdiffusion(500, a= 0.142195074, v= 0.001654804, t=
0.089627361, s = 0.1)

# Create data sets including observed & simulated data
EasyDat <- data.frame(easycond$rt)
EasyDat$type <- "Obs_easy"
names(EasyDat) <- c("rt", "type")

DiffDat <- data.frame(diffcond$rt)
DiffDat$type <- "Obs_diff"
names(DiffDat) <- c("rt", "type")

SimEasyDat <- data.frame(SimEasyDat$rt)
SimDiffDat <- data.frame(SimDiffDat$rt)
SimEasyDat$type <- "easy"
```

```
SimDiffDat$type <- "difficult"
names(SimEasyDat) <- c("rt", "type")
names(SimDiffDat) <- c("rt", "type")

simTot <- rbind(SimEasyDat, SimDiffDat, EasyDat, DiffDat)
simTot$type <-as.factor(simTot$type)
levels(simTot$type)

DatEasy <- rbind(SimEasyDat, EasyDat)
DatEasy$type <-as.factor(DatEasy$type)
levels(DatEasy$type)

DatDiff <- rbind(SimDiffDat, DiffDat)
DatDiff$type <-as.factor(DatDiff$type)
levels(DatDiff$type)


# PLOTS

# Plot: all data sets + easy condition + difficult condition
ggplot(simTot, aes(rt, colour = type))+geom_density() +xlim(0 ,4)
ggplot(DatEasy, aes(rt, colour = type))+geom_density() +xlim(0 ,4)
ggplot(DatDiff, aes(rt, colour = type))+geom_density() +xlim(0 ,4)

# Simulated easy condition density plot: with upper & lower boundary
par(mfrow=c(1,2))
plot(density(SimEasyDat[SimEasyDat[,2]=='upper',1]),main="RT simulation
easy cond")
lines(density(SimEasyDat[SimEasyDat[,2]=='lower',1]),lty=3)

# Simulated difficult condition density plot: with upper & lower boundary
par(mfrow=c(1,2))
plot(density(SimDiffDat[SimEasyDat[,2]=='upper',1]),main="RT simulation
easy cond")
lines(density(SimDiffDat[SimEasyDat[,2]=='lower',1]),lty=3)

-------------------------------------------------------------------------------

# RT DISTRIBUTION USING THE HDDM PACKAGE

# Simulate RT dist based on EZ estimates from the HDDM package
# Create a data frame with the parameter estimates of the EZ model (input
is output of previous get.vaTer function)
SimHDDMest <- rdiffusion(500, a= 1.884, v= 0.359, t=0.0548)

# Create plot of this simulated data
par(mfrow=c(1,2))
plot(density(SimHDDMest[SimHDDMest[,2]=='upper',1]),main="RT simulation
HDDM estimates of EZ")
lines(density(SimHDDMest[SimHDDMest[,2]=='lower',1]),lty=3)

# Create data frame with simulated & observed data
dfHDDM <- data.frame(SimHDDMest$rt)
dfHDDM$type <- "HDDMsim"
obsimhddm <- rbind(dfHDDM, EasyDat)
obsimhddm$type <-as.factor(obsimhddm$type)
levels(DatEasy$type)
obsimhddm$rt <- as.numeric(obsimhddm$rt)

# Plot RT distributions from simulated HDDM + observed data
ggplot(obsimhddm, aes(rt, colour=type)) + geom_density() +xlim(0, 5)
```

```
# EZ model fit on both conditions

# Prepare & Check data
setwd("~/set/wd")
courtdat <- read.csv("Data_file.csv")
names(courtdat) <- c("trial","condition","accuracy","rt","date","rtold")
print(courtdat)
courtdat$condition <- as.factor(courtdat$condition)
levels(courtdat$condition)
courtdat$rt <- as.numeric(courtdat$rt)
str(courtdat)
summary(courtdat)


# Conditions & variables with correct or incorrect trials only
easycond <- courtdat[courtdat$condition=="Male_Virgin",]
diffcond <- courtdat[courtdat$condition=="Mated_Virgin",]
CorrEasyCond <- easycond[easycond$accuracy==1,]
CorrDiffCond <- diffcond[diffcond$accuracy==1,]
IncorrEasyCond <- easycond[easycond$accuracy==0,]
IncorrDiffCond <- diffcond[diffcond$accuracy==0,]
correctdat <- courtdat[courtdat$accuracy==1,]
errordat <- courtdat[courtdat$accuracy==0,]


# DESCRIPTIVE DATA

# Mean accuracy, rt and variance
# Variables with acc, rt and varrt are created to use in the model fit

# Accuracy
meanAccEasy <- mean(easycond[,3])
meanAccDiff <- mean(diffcond[,3])

# RT
meanRTeasy <- mean(easycond[,4])
meanRTeasycorrect <- mean(CorrEasyCond[,4])
meanRTeasyerror <- mean(IncorrEasyCond[,4])

meanRTdiff <- mean(diffcond[,4])
meanRTdiffcorrect <- mean(CorrDiffCond[,4])
meanRTdifferror <- mean(IncorrDiffCond[,4])

# Variance
varRteasy <- var(easycond[,4])
varRtCorreasy <- var(CorrEasyCond[,4])
varRtIncorreasy <- var(IncorrEasyCond[,4])

varRtdiff <- var(diffcond[,4])
varRtCorrDiff <- var(CorrDiffCond[,4])
varRtIncorrDiff <- var(IncorrDiffCond[,4])


-----------------------------------------------------------------------------

# FIT DATA TO EZ DIFFUSION MODEL

# EZ Check 1. RT distribution should be right-skewed
# First, Visual inspection
ggplot(courtdat, aes(rtold,colour=accuracy)) +geom_density() +xlim(0,250)
# Second, D'Agostino skewness test
```

```
library(moments)
agostino.test(courtdat[courtdat[,3]==1,4])
# Thus, all data have a skewness


# Check2. Correct & error RTs should be equally fast
# Look at table with RTs
mean(CorrEasyCond[,4])
mean(IncorrEasyCond[,4])

# t-test
# H0: the RTs in the accurate trials are equal to the RTs in the incorrect
trials (or: RT(correct)-RT(incorrect)=0)
t.test(courtdat$rt~courtdat$accuracy)


# Check3. There should be no bias in RT for one of the two choice
alternatives
#t.test(courtdat$rt~courtdat$side)
#ggplot(courtdat, aes(y=rtold, x=side, colour=accuracy)) +
stat_boxplot(geom = 'errorbar') + geom_boxplot(fatten=5)


# Perform EZ model fit
# Output: drift rate, threshold, nondecision time
source("EZ.R")
parsEasyCond <- get.vaTer(meanAccEasy,varRteasy,meanRTeasy)
parsDiffCond <- get.vaTer(meanAccDiff,varRtdiff,meanRTdiff)
print(parsEasyCond)
print(parsDiffCond)
```

```
# Python script for HDDM model fit

# Load all packages
import pandas as pd
import matplotlib.pyplot as plt
import hddm
from kabuki.analyze import gelman_rubin

# Set wd & retrieve data
%cd "C:/set/to/wd"

# Load data
data = hddm.load_csv('Data_file.csv')

# Check data file
data.head(10)

# Check convergence with the Gelman-Rubin statistic
models = []
for i in range(5):
    Court = hddm.HDDM(data, depends_on={'v': 'stim', 'a'= 'stim'},
     include=('st') )
    Court.find_starting_values()
    Court.sample(5000, burn=20, thin=10)

    models.append(Court)

# Check for convergence with Gelman-Rubin statistic
hddm.analyze.gelman_rubin(models)

# Convergence: visual check
Court.plot_posteriors(['a', 't', 'v'])

# Plot predictive posteriors
Court.plot_posterior_predictive(figsize=(14,10))
plt.title('Posterior predictive')
plt.xlabel('RT')
plt.xlim(-4,4)
plt.ylabel('Probability density')

# Statistics
stats = Court.gen_stats()
Court.print_stats()

# Posterior predictive statistics
ppc_data = hddm.utils.post_pred_gen(Court)

# Look at first lines of PPC
ppc_data.head(10)

# Compute summary statistics & compare to actual/own data
ppc_compare = hddm.utils.post_pred_stats(data, ppc_data)
print ppc_compare

#Fit the EZ model on the dataset
#Insert: pc, vrt, mrt, scaling parameter
#Output: v, a, Ter
hddm.utils.EZ(0.6629213, 0.1960009, 0.5113133, s=1)
```

```
# HDDM model fit on both conditions

# Load all packages
import pandas as pd
import matplotlib.pyplot as plt
import hddm
from kabuki.analyze import gelman_rubin

# Set wd & retrieve data
%cd "C:/set/to/wd"

# Load data
data = hddm.load_csv('Data_file.csv')

# Check data file
data.head(10)

# Check convergence with the Gelman-Rubin statistic
models = []
for i in range(5):
Court =hddm.HDDM(data, depends_on={'v': 'stim', 'a'= 'stim'},
    include =('st') )
    Court.find_starting_values()
    Court.sample(10000, burn=1000, thin=10)

   models.append(Court)

# Check for convergence with Gelman-Rubin statistic
hddm.analyze.gelman_rubin(models)

# Convergence: visual check
Court.plot_posteriors(['a', 't', 'v'])

# Plot predictive posteriors
Court.plot_posterior_predictive(figsize=(14,10))
plt.title('Posterior predictive')
plt.xlabel('RT')
plt.xlim(-4,4)
plt.ylabel('Probability density')

# Statistics
stats = Court.gen_stats()
Court.print_stats()

# Posterior predictive statistics
ppc_data = hddm.utils.post_pred_gen(Court)

# Look at first lines of PPC
ppc_data.head(10)

# Compute summary statistics & compare to actual/own data
ppc_compare = hddm.utils.post_pred_stats(data, ppc_data)
print ppc_compare

#Fit the EZ model on the dataset
#Insert: pc, vrt, mrt, scaling parameter
#Output: v, a, Ter
hddm.utils.EZ(0.5862, 0.217926, 0.499954, s=1)
```

```
# Test whether the drift rate estimates are different across conditions

# Plot the drift rate for both conditions
v_Male, v_Mated, = Court2.nodes_db.node[['v(Male_Virgin)',
    'v(Mated_Virgin)']]
hddm.analyze.plot_posterior_nodes([v_Male, v_Mated])
plt.xlabel('drift-rate')
plt.ylabel('Posterior probability')
plt.title('Posterior of drift-rate group means')

# Test probability that the drift rate is larger in one of the conditions
print "P(v_Easy > v_Difficult) = ",(v_Male.trace()> v_Mated.trace()).mean()
print "P(V_Difficult > V_Easy) = ",(v_Mated.trace()> v_Male.trace()).mean()


# Test whether the thresholds estimates are different across conditions

# Plot the drift rate for both conditions
a_Male, a_Mated, = Court2.nodes_db.node[['a(Male_Virgin)',
    'a(Mated_Virgin)']]
hddm.analyze.plot_posterior_nodes([a_Male, a_Mated])
plt.xlabel(‘thresholds’)
plt.ylabel('Posterior probability')
plt.title('Posterior of thresholds group means')

# Test probability that the drift rate is larger in one of the conditions
print "P(a_Easy > a_Difficult) = ",(a_Male.trace()> a_Mated.trace()).mean()
print "P(a_Difficult > a_Easy) = ",(a_Mated.trace()> a_Male.trace()).mean()
```

Supplementary material 2. HDDM posterior plots for multiple data transformations

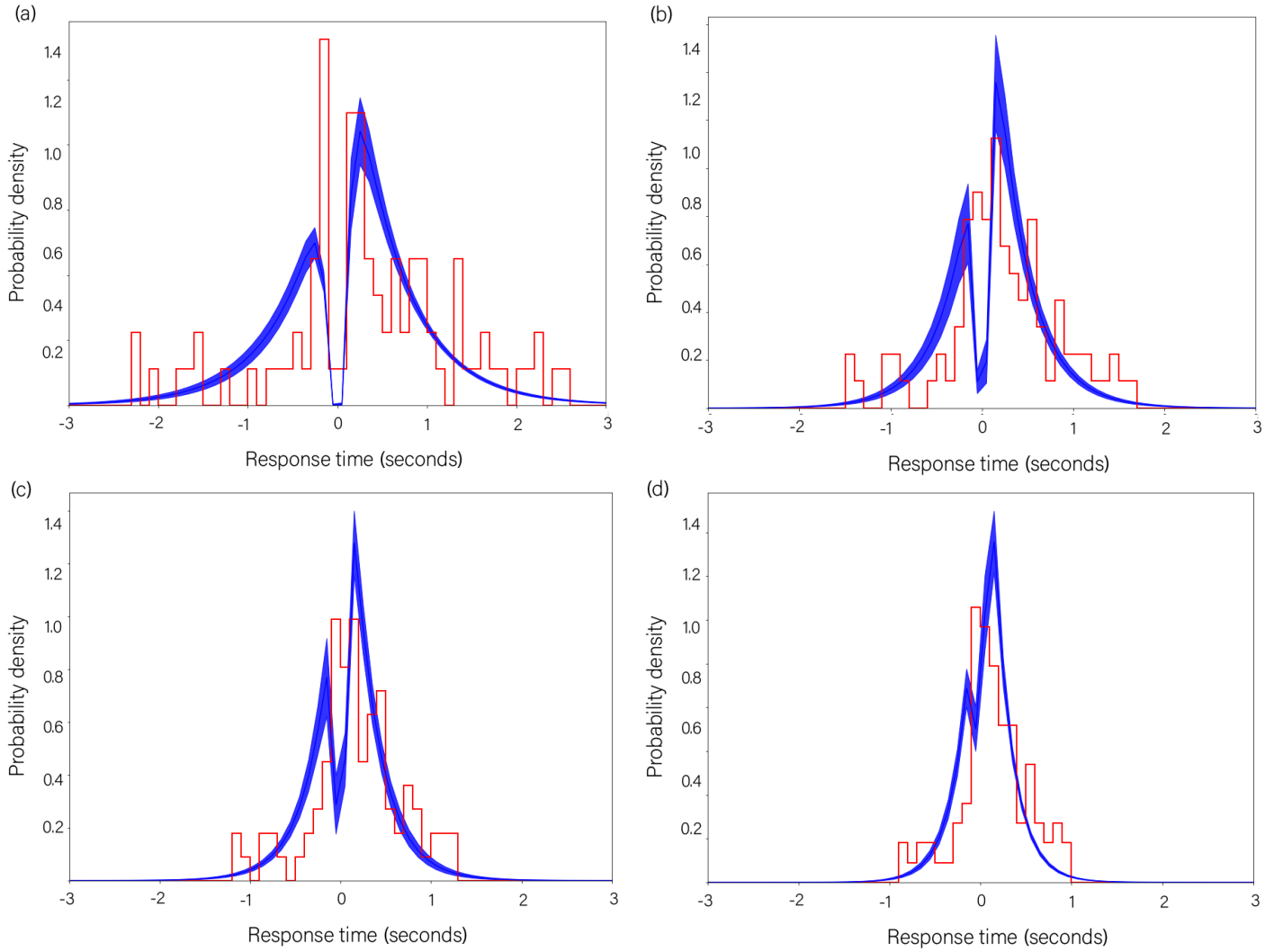


Figure S1. HDDM posterior plots show the HDDM model fit on easy condition data which was transformed to obtain a mean of (a) 800 ms, (b) 500 ms, (c) 400 ms, (d) 300 ms. Note that posterior plots are used to assess which data transformation generates the best model fit instead of the deviance information criterion (DIC). This is because the DIC computes the amount of data not accounted by the model, and the transformation that generates the smaller RTs will provide the lowest DIC.

Supplementary material 3. HDDM posterior plots and posterior predictive statistics

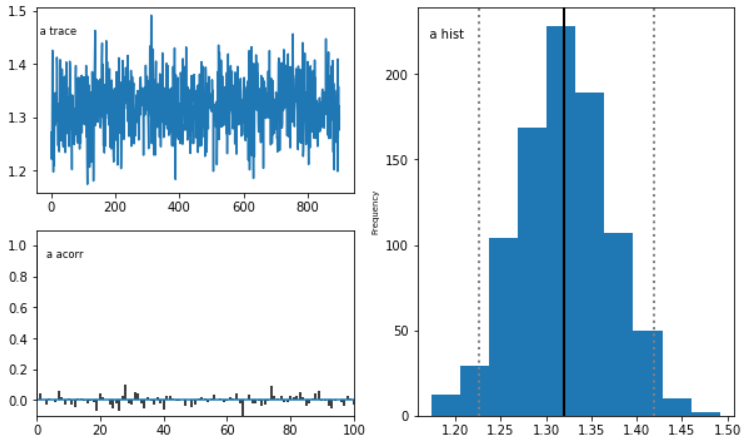


Figure S2. Posterior plots for the subject mean of the decision threshold 'a' based on the easy condition data. (upper left) posterior trace, (lower left) autocorrelation, (right) marginal posterior histogram, with the solid line indicating the mean of the posterior distribution and the dotted lines the 2.5% and 97,5% percentiles.

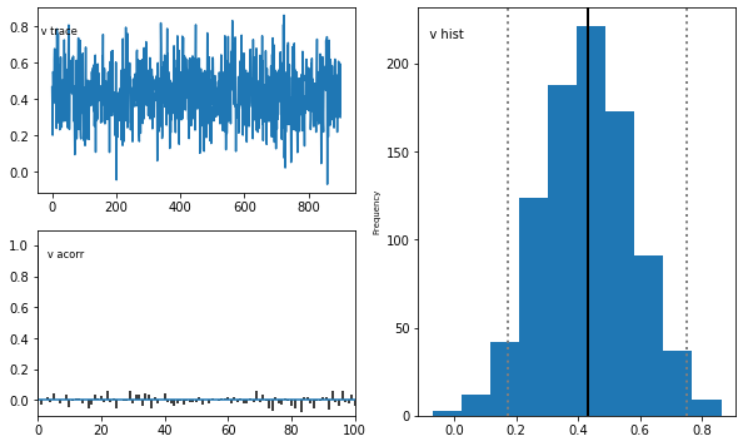


Figure S3. Posterior plots for the subject mean of the drift rate 'v' based on the easy condition data. (upper left) posterior trace, (lower left) autocorrelation, (right) marginal posterior histogram.

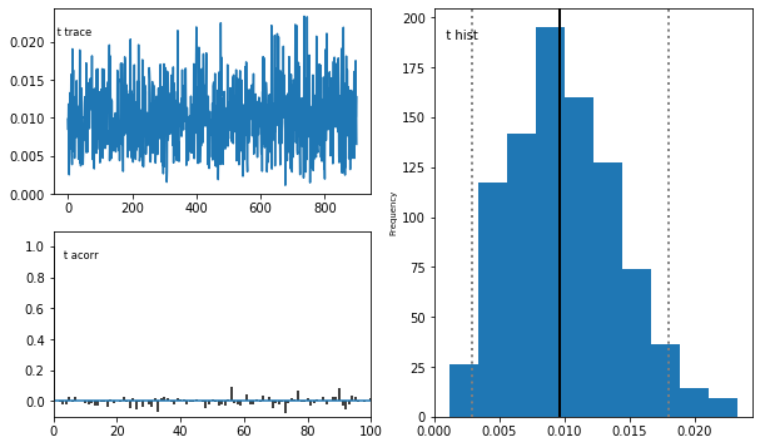


Figure S4. Posterior plots for the subject mean of non-decision time 'Ter' based on the easy condition data. (upper left) posterior trace, (lower left) autocorrelation, (right) marginal posterior histogram.

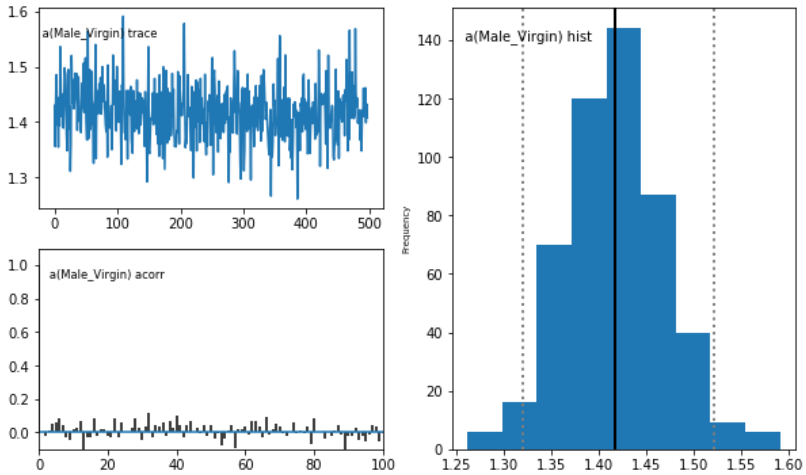


Figure S5. Posterior plots for the subject mean of the decision threshold 'a' in the easy condition based on data from both conditions. (upper left) posterior trace, (lower left) autocorrelation, (right) marginal posterior histogram, with the solid line indicating the mean of the posterior distribution and the dotted lines the 2.5% and 97,5% percentiles.

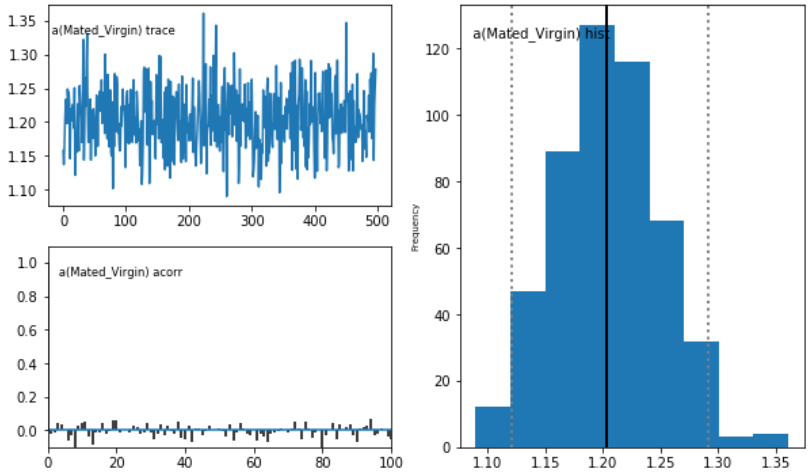


Figure S6. Posterior plots for the subject mean of the decision threshold 'a' in the difficult condition based on data from both conditions. (upper left) posterior trace, (lower left) autocorrelation, (right) marginal posterior histogram.

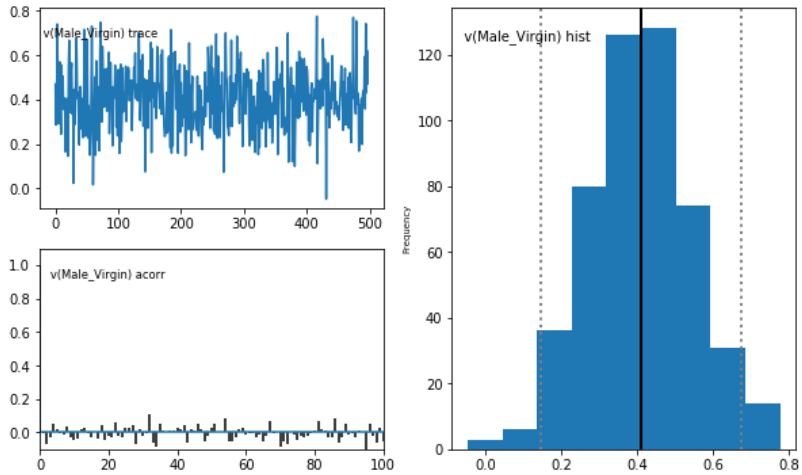


Figure S7. Posterior plots for the subject mean of the drift rate 'v' in the easy condition based on data from both conditions. (upper left) posterior trace, (lower left) autocorrelation, (right) marginal posterior histogram.

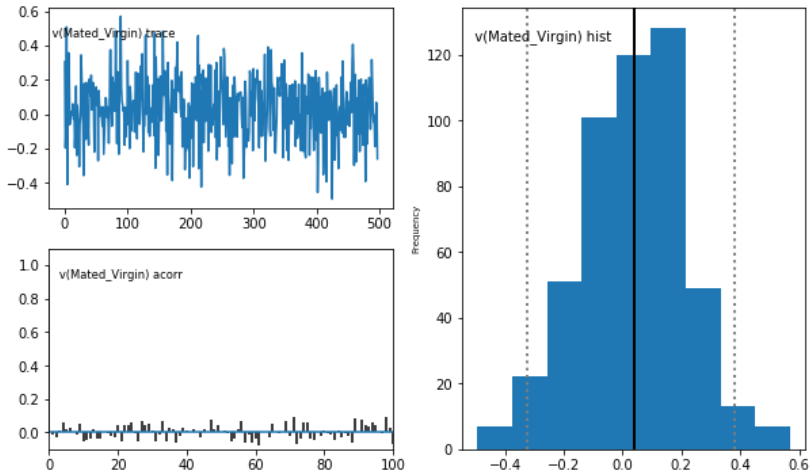


Figure S8. Posterior plots for the subject mean of the drift rate 'v' in the difficult condition based on data from both conditions. (upper left) posterior trace, (lower left) autocorrelation, (right) marginal posterior histogram.

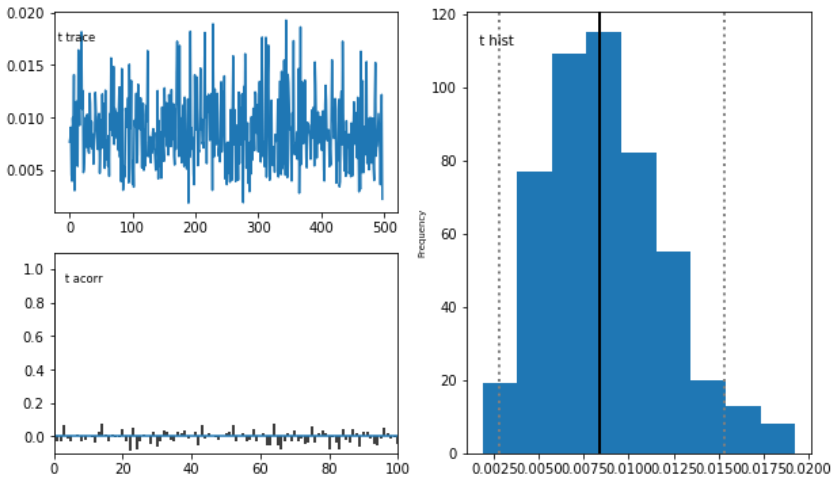


Figure S9. Posterior plots for the subject mean of non-decision time 'Ter' based on data from both conditions. (upper left) posterior trace, (lower left) autocorrelation, (right) marginal posterior histogram.

Table S1. Summary of posterior predictive statistics and observed parameter estimates on easy condition data

Statistic	Observed	Mean	Std.	SEM	MSE	Credible	Quantile
Accuracy	0.663	0.635	0.070	0.001	0.006	True	63.600
Mean ub	0.546	0.432	0.054	0.013	0.016	True	96.600
Std. ub	0.431	0.331	0.059	0.010	0.012	True	94.600
10q ub	0.093	0.128	0.021	0.001	0.003	True	2.600
30q ub	0.215	0.221	0.034	0.000	0.001	True	47.600
50q ub	0.446	0.335	0.052	0.012	0.036	True	96.000
70q ub	0.678	0.505	0.079	0.030	0.015	True	97.000
90q ub	1.194	0.854	0.140	0.116	0.135	False	98.200
Mean lb	-0.443	-0.433	0.067	0.000	0.004	True	41.200
Std. lb	0.450	0.324	0.076	0.016	0.021	True	92.800
10q lb	0.082	0.133	0.030	0.002	0.003	False	2.200
30q lb	0.111	0.225	0.044	0.013	0.015	False	0.200
50q lb	0.195	0.338	0.066	0.021	0.024	False	0.200
70q lb	0.497	0.505	0.095	0.000	0.009	True	49.000
90q lb	1.118	0.845	0.174	0.075	0.105	True	94.600

Parameter estimations by the model based on the data (observed) and estimations based on the simulations from the posterior distribution. Mean is the average estimate of the simulated data. Observed and simulated values are similar if the model fit was good. The standard deviation shows the amount of variation in the parameter estimate in the simulated data. The standard error of the mean (SEM), mean-squared error (MSE) and 95% confidence interval (credible, TRUE means the estimate is within the 95% interval, FALSE means the estimate falls outside the interval) estimate the divergence between the observed and posterior distribution.

Table S2. Summary of posterior predictive statistics and observed parameter estimates on data from both conditions

Statistic	Observed	Mean	Std.	SEM	MSE	Credible	Quantile
Accuracy	0.586	0.576	0.099	0.000	0.010	True	50.900
Mean ub	0.537	0.433	0.085	0.011	0.019	True	87.900
Std. ub	0.466	0.333	0.081	0.018	0.024	True	93.700
10q ub	0.084	0.127	0.028	0.002	0.003	True	4.100
30q ub	0.190	0.219	0.047	0.001	0.003	True	30.900
50q ub	0.374	0.336	0.073	0.002	0.007	True	71.900
70q ub	0.671	0.507	0.111	0.028	0.040	True	92.500
90q ub	1.273	0.863	0.108	0.171	0.212	True	97.300
Mean lb	-0.446	-0.430	0.089	0.000	0.008	True	38.600
Std. lb	0.460	0.329	0.086	0.018	0.025	True	92.500
10q lb	0.064	0.129	0.034	0.003	0.004	False	1.400
30q lb	0.126	0.220	0.054	0.009	0.012	False	0.600
50q lb	0.248	0.332	0.079	0.007	0.014	True	12.500
70q lb	0.464	0.500	0.116	0.001	0.015	True	41.900
90q lb	1.129	0.846	0.210	0.154	0.198	True	94.600

Parameter estimations by the model based on the data (observed) and estimations based on the simulations from the posterior distribution. Mean is the average estimate of the simulated data. Observed and simulated values are similar if the model fit was good. The standard deviation shows the amount of variation in the parameter estimate in the simulated data. The standard error of the mean (SEM), mean-squared error (MSE) and 95% confidence interval (credible, TRUE means the estimate is within the 95% interval, FALSE means the estimate falls outside the interval) estimate the divergence between the observed and posterior distribution.

Supplementary material 4. Manipulate the threshold parameter with DA activity

As this project aims to verify whether the DDM can describe decision making in *D. melanogaster*, experiments and treatments were designed to alter the model parameters. One treatment to differ the decision threshold, which is the amount of information necessary to make a decision, was to include flies with altered dopaminergic (DA) activity. As the threshold depends on the subject's internal state (e.g. stress and cautiousness), the aim was to increase arousal in the flies, so less information would be necessary to make a decision and hence the thresholds would lower. The idea was to arouse the flies with altered DA levels, as lower DA levels are linked to hyperactivity (van der Kooij and Glennon 2007). However, an experiment showed that flies from the lines used to alter DA levels did not display courtship initiation within the 3-7 minutes (only 3 out of 48 males showed wing extension), and hence the data cannot be used to fit the DDM on. More precisely, one fly from the cross w^{1118}*;TH-Gal4* and w^{1118}*;UAS-kir2.1* performed the courtship song towards the virgin female with a RT of 97 seconds; one fly from the cross ^{1118};*TyrosineHydroxylase(TH)-Gal4* and w^{1118}*;UAS-NaChBac* initiated courtship towards the virgin female with a RT of 24 seconds; one fly with w^{1118}*;TH-Gal4* background responded incorrectly to the stimulus male with a RT of 252 seconds. As the goal of this thesis is designing manipulations that alter model parameters, it is important to also describe manipulations that did not work out. Therefore, the treatment including altered DA levels is described in this section.

Dopamine

Dopamine (DA) has frequently been linked with both ADHD and learning and memory. Even though the exact DA mechanisms are not identified yet, a common view is that DA release is crucial to form memories that are motivationally important, which can be either rewarding, novel or punishing events (Genro et al. 2010). Studies on humans, rats, and fruit flies have shown that a loss of DA release can lead to memory deficits (van der Kooij and Glennon 2007; Lisman et al. 2011). Besides its role in memory and learning, DA has been linked to ADHD in the Dopaminergic theory for ADHD. According to the theory by Levy, DA deficits in particular brain areas cause ADHD symptoms (Levy and Swanson 2001). In a more recent DA theory called the DA transfer deficit (DTD) theory, a decrease of DA neural firing is present in children with ADHD. In various animal models, a loss of DA led to the typical ADHD symptoms: hyperactivity, a lack of motivation, and memory deficits (van der Kooij and Glennon 2007). Even though, the amount of DA will not cause memory deficits and ADHD by its own, it is possible to say that a loss of DA is correlated with memory deficits and ADHD. Thus, sensitizing DA neurons can have two effects. Either it solely affects hyperactivity, which is a well-known effect (Friggi-Grelin et al. 2003), or it also affects learning and memory.

Hypotheses

There are two hypotheses on the effects that (de)sensitizing DA neurons will have on male courtship behaviour. The first hypothesis is that flies with sensitized DA neurons will be hyperactive, and hence will respond faster, but also make more errors (both RT and accuracy decrease). In the DDM, this is hypothetically reflected by lowered thresholds, as thresholds partially depend on the subject's state and DA modulates the fly's arousal (Andretic et al. 2005; Van Swinderen and Andretic 2011). On the contrary, flies with desensitized DA neurons will be less hyperactive, and thus respond slower and make less errors (both RT and accuracy increase). In the model, this would then lead to increased thresholds.

The second hypothesis is that sensitized DA neurons also affect, next to its arousal level, learning and memory in *D. melanogaster*. Thus, a fly with sensitized DA neurons is not only hyperactive, but also has a better memory, which will lead to faster responses and better accuracy (RT decreases and accuracy increases). In the DDM, this would hypothetically lead to both a change in threshold height and drift rate, as the threshold is state-dependent, and there is clearer evidence due to better memory. Namely, if the drift rate depends on the quality of information, and if the male remembers better that it was rejected by conspecifics when it tried to court them in the vial when it grew up, it is cleared to that fly that the stimulus female is a better choice than the stimulus male. On the other hand, flies with desensitized DA neurons will be less hyperactive and have a worse memory, leading to slower

responses and more mistakes (RT increases and accuracy decreases). In the diffusion model, this probably is reflected by heightened thresholds and lowered drift rate.

Flies and housing

To alter the activity of dopaminergic neurons, the strains w^{1118}, w^{1118};*TyrosineHydroxylase(TH)-Gal4* (Friggi-Grelin et al. 2003), w^{1118};*UAS-NaChBac* (BDSC#9496), and w^{1118};*UAS-kir2.1* (BDSC#6595) were assessed. The flies with increased (NaChBac) or decreased (kir2.1) DAs activity were collected from the offspring of the crosses w^{1118};*TyrosineHydroxylase(TH)-Gal4* and w^{1118};*UAS-NaChBac*, and w^{1118};*TH-Gal4* and w^{1118};*UAS-kir2.1*. As controls, each of the lines was crossed with w^{1118}, generating w^{1118};*TH-Gal4* and w^{1118}, w^{1118};*UAS-NaChBac* and w^{1118}, and w^{1118} *and* w^{1118};*UAS-kir2.1*. The flies were maintained in bottles with 50 ml standard diet at 25°C on a 12 h light-dark cycle (light on at 12 a.m.). Virgin flies for experiments were collected within 2-3 hours after eclosion, housed with 20 flies per vial, and transferred to new vials after 3-4 days. The flies were used for experiments at 5-7 days after eclosion.

Experiment

The courtship experiment including flies with altered DA neuron activity were performed. Of the 48 male responders tested that day, only 3 showed courtship initiation. Two made an accurate response, and the RTs were 24, 97 and 252 seconds. Even though some trials were recorded for 7 minutes to see whether the 3 minutes time span was too short, this did not lead to more responding males. Possibly, these males required more than 7 minutes to perform courtship initiation and the experiment duration might have been prolonged. However, the aim of the project is to design an experiment on fast decision making, and it is impossible without additional experiments to determine whether choices after 7 minutes still rely on a fast decision process as the DDM describes. Therefore, flies from these crosses will not be used in further experiments to assess the DDM.

Conclusion

The experiment on flies with altered DA neuron activity showed that the included strains were unsuitable to assess fast decision making in this particular courtship experiment. Only few flies performed courtship initiation within 3-7 minutes. The strains with in- and decreased DA neuron activity differ greatly with Canton-S flies in speed and amount of courtship initiation with (see Results for Canton-S). This difference might due to the w1118 background of the lines, since w1118 has been linked to reduced courtship behaviour and a deficit in copulation success (Xiao et al. 2017).

Supplementary material 5. Open versus closed assays

In a trial experiments, assays with closed plastic covers were used which generated 59.26% of the male responders showing courtship behaviour towards males. This was unanticipated, because the male is expected to court the virgin female more often, and hence taken as an indication that perhaps the pheromones might saturate in the dish. In a similar fashion, Laturney and Billeter (2016) found an effect of closed assays, where the pheromones could saturate in the dish, versus open assays, where pheromones did not saturate. Therefore, the following trial experiments were performed using their idea of an 'open assay', i.e. a 35 mm Petri dish covered with a polyester mesh placed in a fume hood, where odours could ascent out of the dish. Using open versus closed assays resulted in a significant improvement of accuracy ($\chi^2 = 5.553$, $df = 1$, $p = 0.018$), as male responders only showed courtship initiation towards males in 43.83% of the trials in experiments with open assays (Figure S4). Males in the closed assays also responded significantly faster (mean RT is 29 sec.) than in the open condition (mean RT is 55.090 sec., $\chi^2 = 6.980$, $df = 1$, $p = 0.008$) (Figure S5). Mean RT was 29±6.445 seconds (SEM) in the closed assay and 55.090±5.056 seconds (SEM) in the open assay. As the mean RT in the closed condition is 29 seconds, the saturation of pheromones may occur within half a minute. Possibly, the males could not distinguish between the male and female stimuli fly as soon as the pheromones saturated in the dish, and hence they might have simply guessed whether the fly they encountered was a male or female, which resulted in faster responses. Performing the experiments in a fume hood and using the polyester mesh as a cover might possibly have prevented some saturation, as the accuracy in the male versus virgin female (easy) condition is higher than in the mated versus virgin female (difficult) condition (results). In the easy condition, the male showed courtship initiation towards the female significantly above chance threshold, and hence it must have been able to distinguish between the presented male and female.

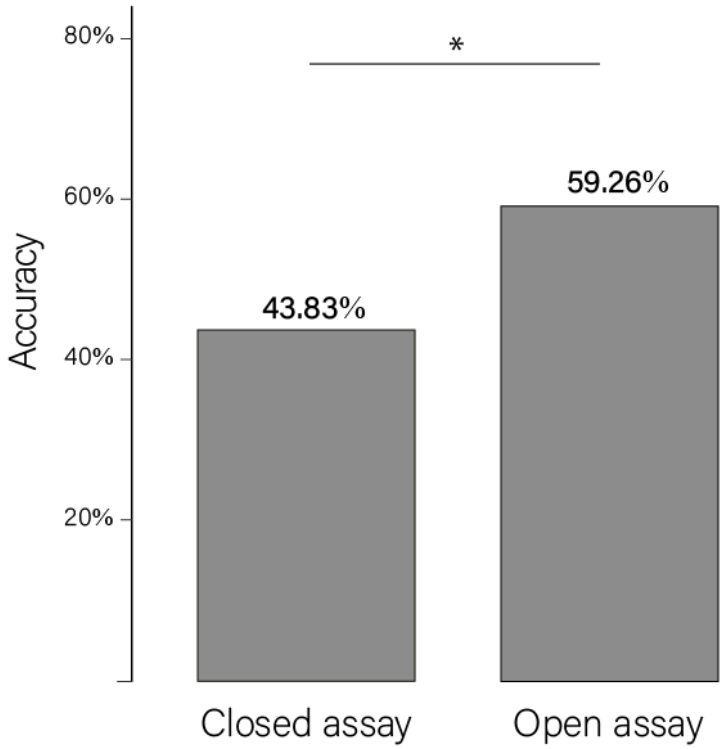


Figure S10. Display of percentages of males that showed courtship initiation towards males in closed and open assays. In the closed assays, 27 male responders were tested and 59.26% initiated courtship towards a male fly. In the open assays 89 responders were assessed and 43.83% showed courtship initiation towards a male fly.

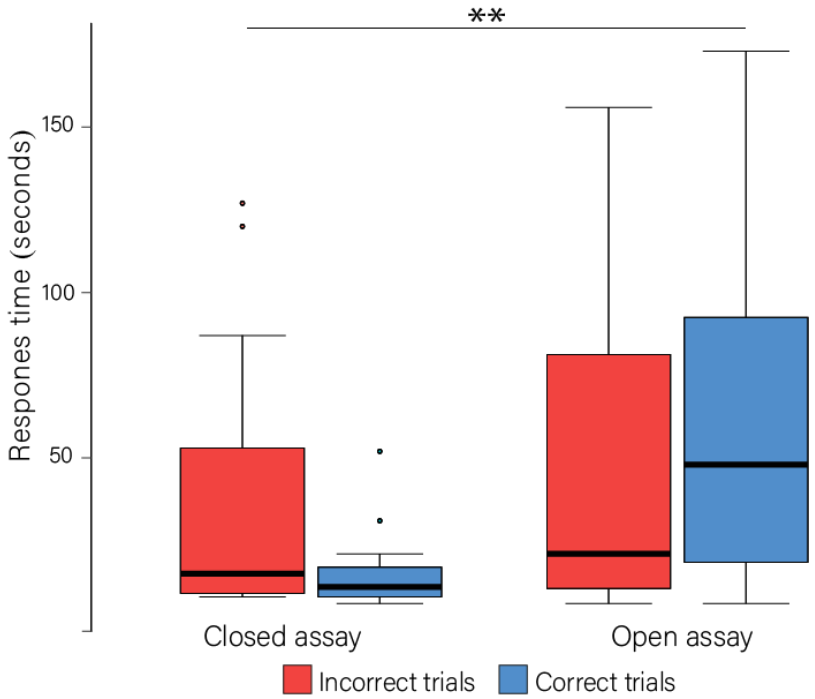


Figure S11. Box plot showing the response time (RT) distribution for correct and incorrect trials in open and closed assays. In the correct trials, the male responder showed courtship initiation towards a female, whereas in the incorrect trials, the male performed courtship initiation towards a male fly. In the closed assays, 27 male responders were tested (11 correct, 16 incorrect), and in the open assays, 89 responders were assessed (59 correct, 30 incorrect).