Visual Feature Dependent Subjective Difficulty and its Ability to Improve Efficiency

Peter Møller Nielsen
Aalborg University
Aalborg, Denmark
pmni15@student.aau.dk

ABSTRACT

In this paper the relation between subjective difficulty, features of stimuli and improvement of efficiency is tested in a visual search experiment. The relation is tested, in hope of getting a better understanding of the relation between perception and subjective difficulty. 11 subjects are tested in a web-based hybrid foraging search experiment, where each subject conducts a visual search for three different targets per trial. There is a total of 20 trials where targets and distractors in each trial are randomised in size, difficulty condition and color. At the end of each trial the subjects rates the subjective difficulty of the search. A manipulation check is conducted with a linear mixed model for the response time and setsize typical for visual search. The manipulation check shows the linear relation and that the search efficiency is dependent on visual features, which is in accordance with visual search models. The dependence between subjective difficulty and the perception of visual features is tested with a linked linear mixed model, using the predictions of the model for response time from the manipulation check. The model shows a relation between the subjective difficulty, the visual features, the predictions and residuals from the manipulation check model. Lastly the interaction between trial and subjective difficulty is used in a linear mixed model to predict efficiency improvements for the visual search. The model shows a significant predictability by the interaction between subjective difficulty and trial for response time in the visual search paradigm, with results indicating a growing subjective difficulty will hinder improvement. The results of the findings are discussed for implications on user centered product design, further research and the relation between subjective difficulty and visual search.

KEYWORDS

Visual Search, Difficulty, Efficiency, Improvement, Linked Linear Mixed Model, User Experience, Features, Experimental Design, Working Memory, Mental Workload„ Linear Mixed Model,

ACM Reference Format:
Peter Møller Nielsen. 2020. Visual Feature Dependent Subjective Difficulty and its Ability to Improve Efficiency. In *Proceedings of 10. Semester project (Nielsen, Peter M.).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Knowledge of a linear relationship between subjective difficulty and features in visual search, makes it possible to make gradual changes to an interface, to control the subjective difficulty affiliated with user interaction. Further understanding of subjective difficulty, will make it possible to approach it as a cognitive system instead of a cognitive construct. A better understanding of the improving effect subjective difficulty might have on efficiency, can be helpful for a variety of user centered product design scenarios. Examples of such scenarios could be the design of products where improved efficiency for the user is a part of the user interaction, or it could be used as an evaluation tool for design alterations to a product in regards to efficiency. It can also help with understanding the cognitive and perceptual processes related to visual search, and these processes relation to subjective difficulty.

Previous research into subjective difficulty as a continuous construct dependent on perception, accessible through self-report, is such an infrequent occurrence in the literature, that an article by Nielsen [29] is the only study on the issue the author is able to find, besides the current attempt. There is however ample literature on subjective difficulty as a quantitative construct dependent on cognition accessible through self-report [1, 7, 22], as well as literature on subjective difficulty dependent on cognition [4, 12, 15, 33] and subjective difficulty being accessible with self-report [36].

The author is unable to find literature of subjective difficulty as a quantitative part of perception, used as a predictor for improvements in task efficiency. However task difficulty have been investigated in relation to the improvement in visual perceptual tasks [2]. In the article by Ahissar and Hochstein [2], improvement is termed perceptual learning, as the effects are consistent when tested in a longitudinal study.

Visual search is a well investigated area of perception, with over 40 years of research [11, 18, 19, 21, 29, 37, 39, 40, 42]. Visual search defines response time as an objective measure of efficiency [18, 39], and thus improvement as a reduction in response time. Visual search is the search after an object (target) in a set of objects (distractors) [39]. Search efficiency for a visual search is based on the linear relationship between response times and the size of the set of objects (i.e setsize) [19, 39], with a dependency on visual features [18], making it a suitable paradigm to investigate visual feature dependent subjective difficulty.

One of the first models constructed for visual attention deployment using the visual search paradigm is Feature Integration Theory (FIT) [21, 37]. FIT segments visual perception into an attentive and a preattentive stage. The attentive stage is used for when a conjunction of features is necessary for visual search, this is the case when a target share all its features in different combinations with distractors [37]. The preattentive stage is when a single feature is enough to conduct that search, as is the case when the target has a feature that differs from the features of all distractors [37]. E.g a red circle between blue circles. The use of the preattentive stage for deployment of attention, would result in an efficient search. Comparably visual search using the attentive stage would result in an inefficient search [37]. The strict separation into two stages for deployment of attention, is an inadequate assumption, as different feature differences leads to different search efficiencies [18]. Because of this issue and others (see Humphreys [21]), FIT in the form proposed by Treisman and Gelade [37] is considered insufficient for explaining attentional deployment. FIT is an inspirational source for Guided Search [39]. Instead of two stages, Guided Search assumes that attention is deployed based on guidance from top-down and bottom-up sources in the preattentive stage, and a selection of objects in the attentive stage [18, 39]. The processing in the preattentive stage is done in parallel, where as the selection in the attentive stage is serial [39]. In latter additions of the model, the selection of objects is done with a diffusion filter, that takes objects in a serial manner, but processes them in parallel, to determine if it is a target or not [18]. The guidance of attention is also separated from the preattentive stage in the latter models[18], making it possible to differentiate between features able to guide attention and features unable to guide attention. In latter years Wolfe et al. [42] and others [10, 11, 23, 40, 41, 43], have expanded the investigation of visual search with new paradigms—hybrid search and hybrid foraging search. Hybrid search is visual search with multiple different search targets in a set [11, 41, 43], where as hybrid foraging search is visual search with multiple instances of multiple different search targets in a set [23, 40, 42]. Hybrid foraging search changes the linear relation between response time and setsize, to a linear relation between response time and *effective setsize* [40, 42]. The *effective setsize* is calculated based on how many targets there are in a set, relative to the total number of objects, making the assumption that the search efficiency is linear with the proportions of targets in a set. It could be argued that this, also should be the case for hybrid search, but it does not seem to be the case in the literature [41, 43]. In case of a hybrid foraging search it opens up for different types of search, dependent on what targets came before the presently searched target [40, 42]. The first target in a search, is the only searched target without a previous target, and as such the search is the only search conducted based on a novel set [40, 42]. After the first target is found, a similar type of target can be found, making it a run after this type of targets, or the target type can be switched, making a change in type of target searched for [40, 42]. The switch is often slower compared to the run [29, 40, 42]. Another type of search is a temporary search where another target type is selected in the middle of a run, because the target is proximate or popping up due to high salience [40, 42]. The last type of search, is when a previously searched target type is returned to [40, 42]. The search types are fittingly termed *first*, *run*, *switch*, *temp* and *return* [42]. Both paradigms add a dimension of memory to the visual search task shown through—memorization as a part of the experimental design [41–43], the data analysis [10, 41, 43] and in direct studies of the relation to working memory [10].

Working memory is the process/processes responsible for storage and functions related to an active task [9, 24]. There is however some disagreement between theories describing working memory [8, 9, 24]. Two of these theories are embedded processes [8] and the multi-component model [24]. Embedded processes assumes working memory to be an activated part of long-term memory, where attention is deployed inside using the Focus of Attention Cowan [8]. Focus of Attention have a capacity if three to four objects at a time [6, 8], consistent with the findings of Luck and Vogel [26]. The upper limit is assumed to be because of interference and loss over time and not a 'hard' upper limit [8]. The Focus of Attention is a crossmodal system, used for all deployment of attention Cowan [8]. In contrast the multi-component model assumes modal specific systems for storage and some functions [5, 9]. The most investigated of these specific systems are the phonological loop and the visuospatial sketchpad. The phonological loop is the storage and rehearsal system for audio stimuli, where as the visuospatial sketchpad is used for visual storage and spatial rehearsal [5]. In the version of the multi-component model presented by Logie [24] the visuospatial sketchpad is divided into the inner scribe where rehearsal of spatial stimuli are done, and the visual catch that stores the visual stimuli. The division of working memory into domain specific components, enables different domain specific storage sizes, explaining the ability of storing seven audio presented digits, but only being able to store five blocks that light up in a sequence [5]. The multi-component model also have some more crossmodal components, such as the central executive and the episodic buffer. The episodic buffer is a crossmodal storage of around four units[5] (some times one [14]), consistent with the results from Luck and Vogel [26]. The central executive started as a placeholder for executive functions [5]. Further analysis into the executive functions have shown at least three different executive functions—inhibition, shifting and updating [28], where inhibition is further divided into three kinds of inhibition—of irrelevant stimuli information, of response and of memory interference [16], and a lot more executive functions are argued for [5, 25].

As pointed out in the above sections, both models accommodate the object restriction of four shown by Luck and Vogel [26] and latter argued by Zhang and Luck [44]. Studies in expertise [35] and memory loading of different items, [3] do raise questions regarding an object based nature of the memory storage. The results shows a more adaptable, complex and load dependent storage system. The upper limit of four objects is not contested [3, 35], but the nature of the memory storage is up for debate. In regards to the relation between hybrid search and working memory the interpretation of their relation seems unclear. In the relation proposed by Wolfe [41], there is a distinction between activated long term memory and working memory, even though the embedded process model, that activated long term memory stems from, is a model for working memory [8] and not a model supplementing working memory. A study by Doherty et al. [9] shows a larger support for the multi-component model, as the interference from dual tasks are domain dependent, however the predictions of the multi-component model are not perfect [9]. Anoter way of interpreting the interference effect of dual tasks is through the proximity of activated brain regions, where a closer proximity leads to greater interference [30]. These insecurities and others regarding the presented cognitive models, shows that there still is a lot to be learned and understood regarding the systems and their structure.

In this paper it is attempted to investigate the relation between the cognitive systems used in hybrid foraging search and the subjective difficulty related to the execution of the search. The paper primarily attempts to answer two questions in this regard. *Can the visual features of a stimuli predict the subjective difficulty in a hybrid foraging search task, and does the subjective difficulty matter for the improvement of search efficiency over multiple repetitions/time spent conducting hybrid foraging tasks.*

The questions are investigated through a web-based hybrid foraging search experiment where the subjective difficulty is measured using a Visual Analog Scale (VAS). To test if it is a hybrid foraging search experiment, a manipulation check is conducted by making a linear mixed model (LMM) for response time predicted by visual search relevant variables, such as setsize and features. The linear mixed model from the manipulation check is used with relevant variables in a linked linear mixed model (LLM) to predict subjective difficulty. Subjective difficulty measures interacting with trial is used to predict response time through a linear mixed model, to show the effect subjective difficulty have on response time efficiency over trials. Lastly the results are discussed in regards to assumptions,implications, previous findings and theory.

2 METHOD

The experiment is a hybrid foraging search task for sets consisting of circles randomly differing in feature dimensions of size and color. The subjective difficulty is measured using a VAS, based on the assumption that subjective difficulty is a continuous and relative cognitive construct accessible for the participant. Each set consisted of 150 to 200 circles contained in an area there is 700 pixels wide and 500 pixels high (700x500), where 20 to 30 percentage of the circles are targets. There is three random generated unique target types in each set. The experiment is web-based and all recruitment is done through the social media platforms, Facebook and Reddit. Participants are sampled between the 27 of April 2020 and the 27 of May 2020. It is previously shown that web-based experiments can yield similar results to a lab-based experiment [17, 32] As it is a web-based experiment there is added variation between participants, due to differing experimental surroundings and equipment. To avoid the added variation between participants having an effect, the experiment is strictly held as a within subject design. Keeping it as a within subject design also eliminates potential confounding motivational effects, that can become a problem in between subject design, if one condition is boring compared to the other [32]. Web-based experiments often have a high dropout rate [32], this can however be used as an advantage, as unmotivated participants are sorted out. To make sure unmotivated participants are sorted out by heightening the threshold, different methods can be used, such as sampling personal information, that holds them liable, or a wall of text in the beginning of the experiment [32]. In this experiment a wall of text is used, in the form of a consent form.

2.1 Stimuli and apparatus

The experimental software is programmed using **JavaScript**[13], **CSS3** and **HTML5**. The server side is programmed using **node.js** and the server is set up as a **apache2** webserver, with **node.js** compatibility. The server is setup by it-Service on Aalborg University[38] with 2 CPUs and 2 GB RAM. All data is kept on a **SQlite**-database. The experiment can be found on *https://20gr1084a.es.aau.dk/node0/* until the end of June.

Condition:	0	1	2
Size(radius):	unequal	above 1 pixel difference	above 2 pixels difference
Color(hue):	unequal	above 3.6 degree difference	above 7.2 degrees difference
Condition:	3	4	5
Size(radius):	above 3 pixels difference	above 4 pixels difference	above 5 pixels difference
Color(hue):	above 10.8 degrees difference	above 14.4 degrees difference	above 180 degrees difference

Table 1: The table shows the six conditions used in the experiment and the differences between targets and distractors in the size and color dimensions, defined by radius for size and hue for color.

The stimuli used for targets and distractors in the experiment is circles randomized in regards to size and color. Color is defined through Hue-Saturation-lightness (HSL) values. The color randomization is only though hue values ranging from 0 ° to 360 °, with lightness kept at 50% and saturation at 100%. The size is defined by the radius of the circles ranging from 5 to 20 pixels. A manipulation of the feature differences is made, to ensure full use of the subjective difficulty rating on the VAS. The features are manipulated so the minimum differences between the targets and distractors are made with a margin as shown in table 1. So if a set is constructed in condition 2, it means there is a minimum of 2 pixels difference in radius and a 7.2° hue difference between all target types and the other objects, and the same is the case of the other target types. The conditions are randomized between sets.

2.2 Procedure

Before the experiment the participant is presented with a consent form, experimental instructions and some demographic questions regarding age and sex. When the experiment starts, the participant is shown the page in upper left corner of figure 1. The page contains the four targets presented in the middle of the 700x500 pixels area. Under the area a trial counter showing "Trial **X** of 20" and a button with the text "Start" is visible. Above the area the text "Look carefully at the targets and try to remember them" is present, to remind the participant of their instructed task. To the right of the area a short list of six points is present, to remind the participant of the overall procedure of the experiment, in case the participant forgets the experimental procedure. It was chosen to make this short list to the right of the area, to avoid misunderstandings of the experiment or the participant forgetting the experimental procedure, as a participant in a web-based experiment, lacks the support of a researcher [32]. The participant is instructed to press the button, when they are able to remember the three presented targets. A focus cross will appear on the screen for 2000 milliseconds, with the center of the focus cross 120 pixels from the border of the 700x500 area, to ensure the focus of attention for the participant is placed somewhere surrounded by circles. The 120 pixels distance from the border, is made because if the focus of attention is at the border of the area when the experiment starts, it might prime the participant to use a 'scanning/reading' search strategy, they would not have used, had their focus of attention started surrounded by circles. The page with the focus cross is shown in the middle of figure 1. When the focus cross disappears the set of 150 to 200 circles appears

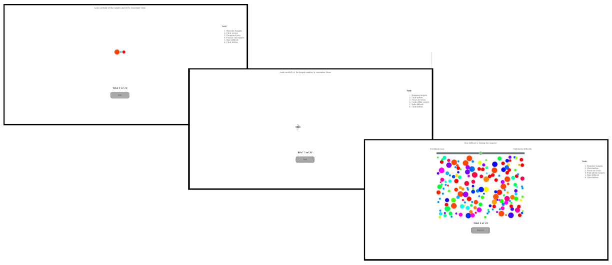

Figure 1: The figure shows the experimental progression for a trial, starting in the upper left corner going towards the lower right corner.

inside the area, and the text above the area on the previous pages are replaced by a VAS with the text "How difficult is finding the targets?" above. The VAS has two end points marked by text, on the left side it says "Extremely easy" and on the right side it says "Extremely difficult". The text on the button is changed to "Next trial". The page can be seen in the lower right side of the figure 1. The changes to the page should help in giving the participant feedback regarding the current state of the experiment. When the participant have found all instances of the three target types, it is able to, the participant rates the subjective difficulty on the VAS and clicks the button. This is repeated over 20 trials, followed by a page indicating they are done with the experiment.

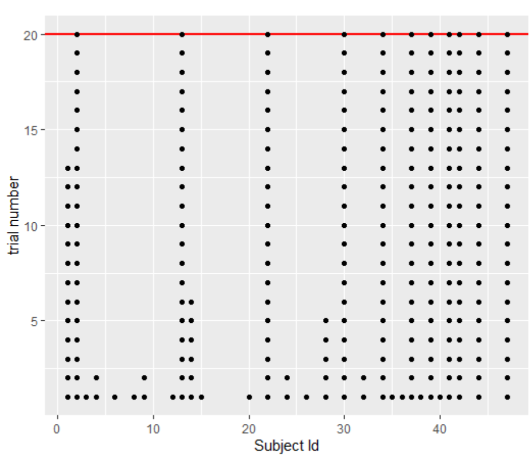

Figure 2: The figure shows the drop out for specific participants. The horizontal axis shows the participant ids,*Subject Id*, and the vertical axis shows the trial number,*trial number*. Each trial is indicated by a dot, and the red line in the top of the screen indicates the twentieth trial and a completed experiment.

2.3 Participants

47 participants began the experiment (22 woman, median age 24, age range 19-38). Out of the 47 participants, around 23.4% completed the experiment. Figure 2 shows at what trial the participants dropped out of the experiment, where each trial is shown with a dot and the red horizontal line at the top indicates a completed experiment. The holes between the dots at the bottom of figure in horizontal direction, indicates a participant that quit the experiment before completing first trial. There is a half and half in participants that quit before first trial and participants that quit before completion but after conducting first trial. This gives a dropout rate on 38.3% for both kinds of drop out.

11 completed the experiment (6 women, median age 26, age range 21-31).

3 RESULTS

For the data analysis the software **R** [31] version 4.0.0 with **Rstudio** [34] version 1.1.447 is used with packages **rjson**, **jsonlite**, **formatR**, **pbkrtest**,**lme4**, **car**, **psych** and **tidyverse**. The linear mixed models are developed and reported using the best practise guideline from Meteyard and Davies [27]. This means all random effects are made first in the parameter selection with log likelihood, then the fixed effects are developed using Kenward-Rogers degrees of freedom, as they make a correction to the sample variance, instead of assuming it equal to the population variance, making the degrees of freedom more robust towards errors than normal degrees of freedom. The only exception from using Kenward-Rogers degrees of freedom, is the manipulation check, where the fixed effects are developed using log likelihood, to minimise the computational needs for the data analysis. All models are compared using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). For further information regarding the parameter selection and the model comparisons, see the supplementary materials.

The measured response time is scaled by the use of log, to minimise the effect of extreme values. The VAS used to measure subjective difficulty gives values between 0 and 100, where 0 is "Extremely easy" and 100 is "Extremely hard". To get single feature values related to each target, the distance between the target and the other objects in a set are calculated and averaged with an arithmetic mean for each feature dimension. For spatial distance the euclidean distance is calculated, and averaged for each target in a set for a given participant. The difference between a selected targets radius and the other objects radius are calculated and averaged for the given set. And the hue difference is calculated and averaged for a set. The hue difference can be both negative and positive, as it got an effect on what mean color difference the set have relative to the given target. This gives 3756 data points, one for each selected target, distributed over 11 participants where each conducted 20 trials. The data shows that some of the participants only clicked three targets for each set, which could indicate that the participants have misunderstood the task as a hybrid search task instead of a hybrid foraging search task. To include possible effects this might have on the results of the experiment the variable *search paradigm* is added. *search paradigm* divides the trials into two groups; one group with the trials where more than three targets are clicked in a set, *hybrid foraging search*, and one group where three or less targets are clicked in a set, *hybrid search*. Through the data analysis trial is treated as a factor, as it is unknown if the distance between units are equidistant, which is a premise for continuous variables. However in cases where trial is used as a fixed effect, as is the case in the model for response time improvement, trial is treated as a continuous variable in regards to the confidence intervals and the estimates for readability and interpretability of the model. A consequence of making this choice is that the estimates of the model becomes relative in nature, as they do not showcase a model where the theoretical assumptions are meet. This means the estimates, but not their precise magnitude, still can be compared to each other and whether or not they have a decreasing or increasing effect on the dependent variable.

3.1 Manipulation Check

The first part of the analysis consists of checking if the assumption of a visual search task is kept. If this is not the case, then the theoretical relation between perception of the features and the subjective difficulty, might not hold true. The model is developed through a top-down parameter selection, which is an approach often used for confirmatory experiments [27]. For a comparison to a model developed using a bottom-up parameter selection, which often is used for exploratory experiments [27], see the supplementary analysis. The model contains two random effects, in the form of intercepts for the participant and an intercepts for the interaction between the participant and trial. Trial and participant is used as these give variance to the data, the model needs to account for, but not include in the analysis of the fixed effects. It is especially important to have participant as a random effect in a web-based within subject experimental design, as there is a added variation between participants, due to the differing experimental equipment and scenarios. The fixed effects for the model is *Size difference*, *color difference*, *search paradigm*, *setsize* and the interaction between *color* and *setsize*. The estimates for the model can be seen in table 2. As can be seen even though some of the parameters, such as color and Setsize, explained a significant amount of variance for the model, it is not significant when it comes to prediction of the response time. The model estimates and p-values shows that the prediction of response time is somewhat consistent with the theory, as both features and setsize have a significant effect. It is however not consistent with

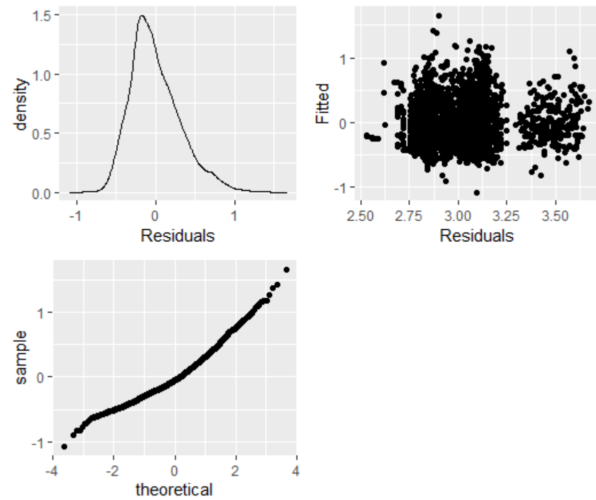

Figure 3: The figure shows three graphical representations of the residuals for the manipulation check model made to test linear model assumptions. In the upper left of the figure is a density plot for the residuals. In the upper right is the residuals held against the fitted values. The plot in the bottom is a qqplot comparing the sample residuals with a theoretical normal distribution.

the theory, that a larger setsize leads to a decreased response time, and a larger color difference leads to an increased response time.

To ensure the validity of the model fit, the assumptions for the residuals are graphically represented with a qqplot, a density plot of the residuals and a comparison of residuals and fitted values. The plots can be seen on figure 3. In the upper left of figure 3 a density plot of the residuals is shown, with a qqplot below it. These two plots shows a somewhat multivariate normal distribution of the predictor variables, as the residuals are normal distributed, indicating a valid fit of the model. The plot in the upper right of figure 3 shows the residuals relative to the fitted values. The residuals are a bit skewed with higher frequency in the positive range, indicating a badly fit model. The skewness of the residuals is not very large, so the validity of the model is accepted, and further analysis is conducted on the assumption of a it being a visual search task.

Parameters:	Df	Estimates	CI 2.5%	CI 97.5%	P-values
Size	1	-1.93e-02	-2.75e-02	-1.11e-02	3.59e-06*
Color	1	7.40e-03	1.80e-03	1.29e-02	0.86
Search paradigm = hybrid search	1	2.14e-01	9.16e-02	3.40e-01	<0.01*
Setsize	1	-1.68e-03	-3.04e-03	-3.24e-04	0.35
Color:Setsize	1	-4.26e-05	-7.43e-05	-1.07e-05	0.01*

Table 2: The table shows the Kenward-Rogers degrees of freedom (DF), Estimates, Confidence Interval (lower at 2.5% and upper at 97.5% and the P-values are made with Kenward-Rogers DF for the manipulation check model. The '*' on the table indicates significant values based on alpha value of 0.05.

Response time	AIC	BIC
Without	30952	31133
Measured	30913	31100
Prediction	30509	30696
Prediction+Residuals	30505	30698

Table 3: The table shows the Akaike Information Criterion and Bayesian Information Criterion for the models predicting subjective difficulty constructed using a bottom-up approach. The first column shows what type of response time there is added to the model seen in 1, the second column shows the AIC values for the models and the last column shows the BIC values for the models.

3.2 Subjective Difficulty predicted by features

To investigate if the Subjective difficulty can be predicted in a visual search paradigm, a liner mixed model is made, in accordance with Meteyard and Davies [27], by finding the random effects first, and then find the fixed effects using Kenward-Rogers degrees of freedom. The parameter selection is done with a bottom-up approach, as there is a limited (see Nielsen [29]) theoretical basis for the association. For a comparison to a top-down approach for parameter selection, see the supplementary material. As in Nielsen [29], it is assumed that the prediction of subjective difficulty follows the perception of the features.

$$\begin{aligned} subjectiveDiff \sim (1|participant) + (1|Trial) + \\ Condition + SetSize + Size + Condition : Color + \\ Condition : Setsize + Condition : Size + \\ Size : Color + Setsize : Color \end{aligned} \tag{1}$$

This sequential relation is shown through a linked linear model. A linked linear model assumes a sequential dependency between two distributions, shown through the use of predictions from one dependent variable in one model, response time, in the prediction of a second dependent variable in a second model, subjective difficulty [20]. For further explanation of the linked linear models see Hohenstein et al. [20], and for validation of the method results, see the supplementary materials for Hohenstein et al. [20]. Before implementing the predictions from the response time mode, made as the manipulation check, all other independent parameters is selected, ending out in the model seen in equation 1. The effects indicated with the form $(1|x)$, is random intercept effects. In the current model it is trial and participant.

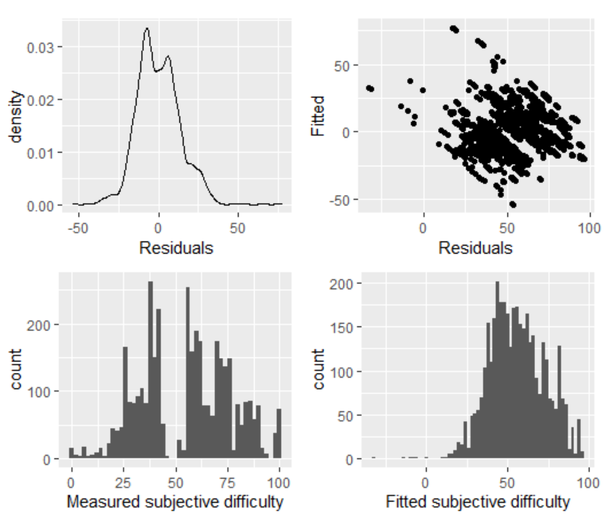

Figure 4: The figure shows four graphical representations of the subjective difficulty model, made to test linear model assumptions. In the upper left of the figure is a density plot for the residuals. In the upper right is the residuals held against the fitted values. The plot in the bottom left shows a histogram over the measured difficulty and the plot in the bottom right shows a histogram over the fitted subjective difficulty.

Variable:	Df	Estimates	CI 2.5%	CI 97.5%	P-values
Condition	5	-1.13e+01	-1.59e+01	-6.77e+00	7.17e-16*
Setsize	1	1.98e-02	-7.15e-02	1.11e-01	<2.2e-16*
Size	1	7.74e-02	-7.83e-01	9.29e-01	9.12e-14*
Predicted Response Time	1	1.25e+02	1.12e+02	1.38e+02	<2.2e-16*
Residual Response Time	1	1.38	-1.39e-01	2.90e+00	0.01*
Condition:Color	6	2.23e-03	-8.21e-03	1.27e-02	<2.2e-16*
Condition:SetSize	5	5.85e-02	3.24e-02	8.46e-02	<2.2e-16*
Condition:Size	5	4.17e-01	2.02e-01	6.33e-01	<2.2e-16*
Size:Color	1	-6.04e-03	-1.86e-02	6.58e-03	0.02*
SetSize:Color	1	3.46e-04	-9.05e-05	7.79e-04	<0.01*

Table 4: The table shows the Kenward-Rogers degrees of freedom (DF), Estimates, Confidence Interval (lower at 2.5% and upper at 97.5% and the P-values are made with Kenward-Rogers DF for the subjective difficulty model. The Confidence Intervals and Estimates are made with Condition as a continuous variable and should only be interpreted relative. The '*' on the table indicates significant values based on alpha value of 0.05.

Three models are made— the first adding the measured response time as a fixed effect, the second adding the predictions from the manipulation check as a fixed effect, and the last adding the predictions and the residuals from the manipulation check as fixed effects. The three models and a model without response time, as seen in equation 1, are compared using AIC and BIC. The results can be seen in table 3.

The model with predictions and residuals, from the response time model, as fixed effects are having the better fit. Model assumptions regarding the residuals, are graphically validated using the plots in figure 4. Based on the density plot in the upper left of figure 4, the residuals does not seem normal distributed, indicating a transformation of data might be needed, the model is a somewhat inadequate fit for the data or there is not enough data. Furthermore the spread of the residuals when held against the fitted values shows some heteroscedasticity as the variance changes, indicating there is more affecting the subjective difficulty, than what is accounted for in the model, seen on the plot in the upper right of figure 4. The two histograms over measured and fitted values in the bottom of figure 4, clearly shows that the model does not approximate values from the lower half of the scale very often, even though there is answered in the lower end of the scale. The answers are however centered around the middle values of the VAS, so it might be due to the VAS being an inadequate tool.

The estimates for the model are shown in table 4. The estimates and confidence intervals are made with Condition as a continuous variable, to make the table more readable. It is arguable, that Condition is a continuous variable, as each condition is equidistant to the nearest conditions, meaning there is a difference of 1 pixel and a difference of 3.6° in hue every time condition increases by one. This is also shown in table 1. However in the current study the conditions are assumed distinct as a grouping variable showing what rule the distractors are made with relative to the target. This means the estimates and confidence intervals are relative, and the exact magnitude are unusable for the analysis. Table 4 shows that the largest increase in subjective difficulty is from an increase in the predicted response times from the manipulation check. For most effects an increase leads to an increase in the subjective difficulty. The only two exceptions for this is the condition manipulation, that leads to a decrease in the subjective difficulty, and the interaction between the averaged radius difference and the averaged color difference.

3.3 Response time improvement

In the previous analysis trial is treated as a random effect, meaning the variance between trials is accounted for, but not taking in as a part of the analysis, as potential learning effects, would disrupt the analysis and interpretation of the models. As this part of the analysis is concerned with improvement over time, trial is used as a fixed effect, so the progression over time/difference between trials is a part of the model interpretation. The model is constructed bottom-up, as there is no previous theory for the relations and dependencies between variables. The procedure is the same as with previous models, where the random effects are developed first. In this model the only random effect is an intercept for participant. Much like with the model for subjective difficulty where the linked fixed effects are added last, the interaction between subjective difficulty and trial is the last addition to this model. In table 5 the estimates, p-values, Kenward-Rogers degrees of freedom and confidence intervals for the model is shown. The model shows an increase in the subjective difficulty with an increase in trial, leads to an increased response time. The effect of the interaction is lower than the decreasing effect of an increase in trial. The model shows decreasing effects from the averaged size difference, and the interaction between averaged color difference and trail, where as a shift from the hybrid foraging search paradigm, to a hybrid search paradigm leads to an increase in response time.

To the left of figure 5 the two plot shows the residuals being somewhat normally distributed, making the model fit more likely to be trustworthy. The plot in the upper right corner, shows that the residuals for the model as somewhat skewed towards positive values, indicating the model fit can be improved, as the residuals are not equally distributed on each side of zero for the fitted values. The density plot showing the distribution of the fitted values in the

Variable:	Df	Estimates	CI 2.5%	CI 97.5%	P-values
size	1	-1.88e-02	-2.62e-02	-1.15e-02	4.92e-05*
Search paradigm = Hybrid Search	1	2.25e-01	1.04e-01	3.43e-01	<0.01*
Trial	19	-1.13e-02	-1.43e-02	-8.27e-03	4.26e-05*
Trail:Color	20	-2.68e-05	-5.78e-05	4.22e-06	<0.01*
Trial:Subjective Difficulty	20	1.25e-04	8.16e-05	1.69e-04	5.68e-08*

Table 5: The table shows the Kenward-Rogers degrees of freedom (DF), Estimates, Confidence Interval (lower at 2.5% and upper at 97.5% and the P-values are made with Kenward-Rogers DF for the improvement model. The confidence intervals and estimates are made with trial as a continuous variable and should only be interpreted relativ. The p-values and Kenward-Rogers degrees of freedom is calculated using trial as a distinct variable. The '*' on the table indicates significant values based on alpha value of 0.05.

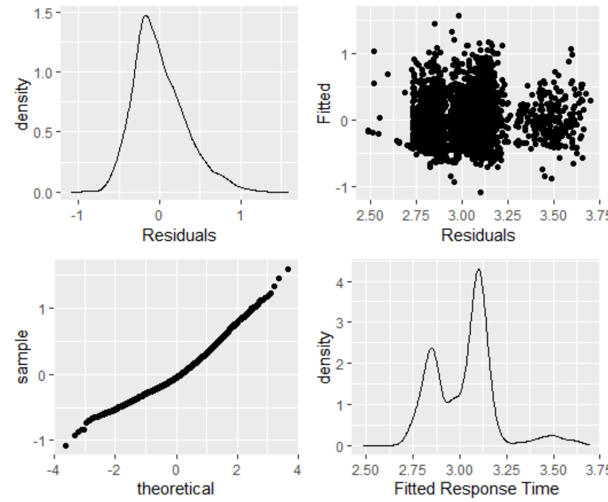

Figure 5: The figure shows four graphical representations of the improvement model, made to test linear model assumptions. In the upper left of the figure is a density plot for the residuals. In the upper right is the residuals held against the fitted values. The plot in the bottom left is a qqplot comparing the sample residuals with a theoretical normal distribution and the plot in the bottom right shows a histogram over the fitted subjective difficulty.

bottom right of figure 5, could indicate there is a missing variable, that could group the response times into a fast paced search and a slower paced search.

4 DISCUSSION

The results from the experiment will be discussed in regards to visual search and working memory. Implications of the experimental method and the restrictions to the interpretation is discussed with a perspective to the usability of the results in a user centered product development context.

Manipulation check: In the manipulation check the visual features, the conditions, setsize and search paradigm all had a significant effect on the search efficiency, as can be seen from table 4. These results are in accordance with the literature for visual search. Setsize and Color is not significant unless it is the interaction between the two variables. This deviates from the theory, as especially Setsize is used when constructing response time curves for visual search experiments. Another deviation is that the model indicates a larger difference in the feature space leads to slower response times, and an increase in setsize leads to decreased reponse time. However as noted in the introduction 1 previous experiments have used the effective setsize instead, under the assumption that every target uses the same number of objects in a set. This assumption might not hold true, as could be argued, if using a hypothetical 100 objects set with two targets. Under the assumption of effective setsize, these two targets would have an equal number of objects that should be search to find them (i.e 50 objects), however if one of the targets is near the border of the search area, and the other target is around the center, than the target near the border is not in the set equal to the target near the center. Previous studies have however shown that the effective setsize is usable when creating response curves for a hybrid foraging search [29, 40, 42], making it a subject for further investigation, also in context of subjective difficulty studies such as this. The color difference calculation, might be the reason for the results being contradictory to the literature. As the differences contains negative and positive hue differences, some of these might have contradicted each other. The argument of the a negative and positive hue difference being logically different still holds true, so a scaling of the variable might be necessary, so the negative values becomes zero, and the differences is given from 0° to 360°. The search paradigm variable constructed in this experiment, got some limitations due to the web-based experimental design used in the current study. The variable is limited as it is unknown if the participants in the study, actually conducted a hybrid search task, in cases where less than or equal to three targets where found in a set. It could also be possible, that the participants was unable to find more targets, but understood the task correct and as such was conducting a hybrid foraging search.

Subjective Difficulty Model: In the model predicting subjective difficulty, estimate can be seen in table 4, the largest effect is the fixed effect of the predictions from the manipulation check model, indicating perception having a large impact on the subjective difficulty. As the model is a linked linear model, the model assumes a sequential relation between perception and subjective difficulty, where subjective difficulty is temporally followed by perception. If this is interpreted in regards to the theory, than the cognitive construct used for accessing the difficulty of a task is affected by the search efficiency of the visual search.

Based on the model assumption, visual search being a task and working memory being the cognitive system used in relation to active tasks, it is possible that the performance of the visual search task is perceived, and processed in a part of working memory. Based on the evaluation of the perceptual performance in the visual search task, the subjective difficulty is created. This view is consistent with the findings by Drew et al. [10], where working memory is affected by visual search tasks, but not the other way around. As pointed out in the introduction 1, the exact relation between working memory and visual search is still unclear, and the model proposed by Drew et al. [10], does not see visual search as a task under working memory. However as the efficiency of the visual search task is accessible, by a cognitive construct able to validate the difficulty of the task, and working memory is the cognitive system used for current tasks, it could be argued that there is some relation between the visual search and working memory, making the subjective difficulty measure possible. This is of course under the assumption that difficulty is not experienced by the perceptual system related to visual search. It is necessary with further study of visual search and working memory, before the exact nature of the relation between the two can be revealed, but the current study indicates visual search as an accessible for evaluation by working memory.

Improvement Model: The results in table 5 indicates a significant effect from trial, both as a fixed effect in itself and in the form of interaction effects. Because trial is an ordinal measure for progression, there is some reservations regarding the interpretation of the results. Because trail can be seen as a measure of time or a measure of repetition. This means it is unsure if the effect on response time happens based on number of times an individual is exposed to the task, or the time spent on the task. The implications of this on product design, is manipulations of visual features to control the subjective difficulty, should happen either based on the time spent on a given task or number of times a given task is done. An example where these implications mattered for a product design, could be a learning tool for pilots to improve search efficiencies, in regards to detecting abnormalities in a cluttered display. In this case, it is important to optimise correctly, to ensure the biggest benefits from the learning tool, either by having a lot of exposure to the search task or by spending a lot of time on searches. The interpretation of trial also have implications for the understanding of the processes happening when improving on a visual search task. In the case where trial is a measure of time, it could mean an individual spending a lot of time on a single search would lead to the similar response time improvement, as an individual that spends a similar amount of time on visual search tasks, but conducts a lot more searches. To answer this it is necessary to further study the precise nature of trial. The interpretation of trial also affects how the interaction effects are understood. As shown in table 5, the interaction between subjective difficulty and trial is significant and an increase in the interaction leads to an increase in response time. If trial is a time measure, it means this effect is continuous, and manipulations of visual features to control subjective difficulty, can be made gradually. If trial is a measure of repeated exposure, than the feature manipulations to control subjective difficulty, should be made in a more distinct fashion, so the manipulations should be made between times of exposure, and not gradually over time. If the second interpretation of trial and subjective difficulty is true, it indicates a need for a break between the tasks before improvement is possible. This also have implications for how a user centered product design should be made, as a system accommodating the need for breaks, should stop the user in conducting the task or somehow segment the tasks to highten improvement.

The lower right corner of figure 5, shows a density plot for the improvement models fitted values. The model have two peaks, that could indicate a missing grouping factor in the analysis. Such a factor could be the search types used in other hybrid foraging search experiments [29, 40, 42], and shortly explained in the introduction 1. As one of the peaks is at lower values indicating faster search times, it would be consistent with the run and switch search types most common for hybrid foraging searches (as can be seen in the results of Nielsen [29] and Wolfe et al. [42]).

The model shown in table 5 indicates that improved efficiency can be controlled by subjective difficulty, and the model shown in table 4 shows subjective difficulty, as dependent on feature manipulated visual perception as well as the visual features. If the results are combined it shows that improvement can be affected through feature manipulations. For this to be useful for learning tools, difficulty adapted progression in product design etc. the improvement needs to last (i.e. learning). It is still possible to use the results for design evaluation of difficulty, or designing for improving short term efficiency however a longitudinal study is necessary for ensuring long term effects, and enabling the results to have a larger significance for user centered product design.

The interaction between subjective difficulty and trial shows that if subjective difficulty increases as trials increases, than the search efficiency decreases. A possible explanation for this could be that the individuals get demotivated, if the individual do not get a feeling of improvement. As the experiment did not contain any feedback on the individuals performance, the increased subjective difficulty as they kept on with the experiment, might have a demotivating effect, leading to longer search times. This does however assume motivation as a variable able to affect search efficiency over time, meaning higher cognition affecting perceptual systems. The assumption might hold true, as cognition affecting perception is the premise for aspects of visual search, such as top-down guidance [18, 39]. It is important to note that the decrease in search efficiency due to an increase in the interaction between subjective difficulty and trial, is less than the increased efficiency from an increase in trials, indicating it is only decreasing improvement, not stopping it.

REFERENCES

[1] Rachel F Adler, Deena Rubin, Abdul Rahman Mohammad, Amna Irfan, Haridu Senadeera, and Timothy Nguyen. 2015. That's Easy! The Effects of Objective and Subjective Task Difficulty when Multitasking. *Procedia Manufacturing* 3 (2015), 5467–5474.

[2] Merav Ahissar and Shaul Hochstein. 1997. Task difficulty and the specificity of perceptual learning. *Nature* 387, 6631 (1997), 401–406.

[3] G.A Alvarez and P Cavanagh. 2004. The Capacity of Visual Short-Term Memory is Set Both by Visual Information Load and by Number of Objects. *Psychological Science* 15, 2 (2004), 106–111.

[4] Frédérique Autin and Jean-Claude Croizet. 2012. Improving Working Memory Efficiency by Reframing Metacognitive Interpretation of Task Difficulty. *Journal of Experimental Psychology: General* 141, 4 (2012), 610–618.

[5] Alan Baddeley. 2012. Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology* 63, 1 (2012), 1–29.

[6] Olivia Beaudry, Ian Neath, Aimée M Surprenant, and Gerald Tehan. 2014. The focus of attention is similar to other memory systems rather than uniquely different. *Frontiers in human neuroscience* 8, 1 (2014), 1–7.

[7] Rémi L. Capa, Michel Audiffren, and Stéphanie Ragot. 2008. The effects of achievement motivation, task difficulty, and goal difficulty on physiological, behavioral, and subjective effort. *Psychophysiology* 45, 5 (2008), 859–868.

[8] Nelson. Cowan. 2005. *Working memory capacity.* Psychology Press, New York. 32–59 pages.

[9] Jason M. Doherty, Clement Belletier, Stephen Rhodes, Agnieszka Jaroslawska, Pierre Barrouillet, Valerie Camos, Nelson Cowan, Moshe Naveh-Benjamin, and Robert H. Logie. 2019. Dual-Task Costs in Working Memory: An Adversarial Collaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45, 9 (2019), 1529–1551.

[10] Trafton Drew, Sage Boettcher, and Jeremy Wolfe. 2016. Searching while loaded: Visual working memory does not interfere with hybrid search efficiency but hybrid search uses working memory capacity. *Psychonomic Bulletin & Review* 23, 1 (2016), 201–212.

[11] Trafton Drew, Sage E. P. Boettcher, and Jeremy M. Wolfe. 2017. One visual search, many memory searches: An eye-tracking investigation of hybrid search. *Journal of Vision* 17, 11 (2017), 5. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5596794/pdf/

[12] Beate Dunst, Mathias Benedek, Emanuel Jauk, Sabine Bergner, Karl Koschutnig, Markus Sommer, Anja Ischebeck, Birgit Spinath, Martin Arendasy, Markus Bühner, Heribert Freudenthaler, and Aljoscha C Neubauer. 2014. Neural efficiency as a function of task demands. *Intelligence* 42, 100 (2014), 22–30. http://search.proquest.com/docview/1826583226/

[13] ECMA Ecma. 2019. 262: Ecmascript language specification. *ECMA (European Association for Standardizing Information and Communication Systems), pub-ECMA: adr,* (2019).

[14] Martin Eimer. 2014-10. The neural basis of attentional control in visual search. *Trends in Cognitive Sciences* 18, 10 (2014-10), 526,535.

[15] Iman Feghhi and David A. Rosenbaum. 2019-08. Judging the Subjective Difficulty of Different Kinds of Tasks. *Journal of Experimental Psychology: Human Perception and Performance* 45, 8 (2019-08), 983,994.

[16] Naomi P. Friedman and Akira Miyake. 2004. The Relations Among Inhibition and Interference Control Functions: A Latent-Variable Analysis. *Journal of Experimental Psychology: General* 133, 1 (2004), 101–135.

[17] Laura Germine, Ken Nakayama, Bradley Duchaine, Christopher Chabris, Garga Chatterjee, and Jeremy Wilmer. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review* 19, 5 (2012), 847–857.

[18] Wayne D Gray. 2007. *Integrated models of cognition systems / edited by Wayne D. Gray* (elektronisk udgave ed.). Oxford University Press, New York. 97–119 pages. http://zorac.aub.aau.dk/login?url=http://dx.doi.org/10.1093/acprof:oso/9780195189193.001.0001

[19] James Hoffman. 1979. A two-stage model of visual search. *Perception & Psychophysics* 25, 4 (1979), 319,327.

[20] Sven Hohenstein, Hannes Matuschek, and Reinhold Kliegl. 2017. Linked linear mixed models: A joint analysis of fixation locations and fixation durations in natural reading. *Psychonomic Bulletin & Review* 24, 3 (2017), 637,651.

[21] Glyn W Humphreys. 2016. Feature confirmation in object perception: Feature integration theory 26 years on from the Treisman Bartlett lecture. *The Quarterly Journal of Experimental Psychology* 69, 10 (2016), 1910,1940. http://www.tandfonline.com/doi/abs/10.1080/17470218.2014.988736

[22] Nicole Jochems, Sebastian Vetter, and Christopher Schlick. 2013. A comparative study of information input devices for aging computer users. *Behaviour & Information Technology* 32, 9 (2013), 902–919. http://www.tandfonline.com/doi/abs/10.1080/0144929X.2012.692100

[23] Arni Kristjansson, Omar I. Johannesson, and Ian M. Thornton. 2014. Common Attentional Constraints in Visual Foraging.(Research Article). *PLoS ONE* 9, 6 (2014), e100752.

[24] Robert H Logie. 2011. The Functional Organization and Capacity Limits of Working Memory. *Current Directions in Psychological Science* 20, 4 (2011), 240–245.

[25] Robert H Logie. 2016. Retiring the central executive. *The Quarterly Journal of Experimental Psychology: QJEP Classics Revisited* 69, 10 (2016), 2093–2109. http://www.tandfonline.com/doi/abs/10.1080/17470218.2015.1136657

[26] Steven J. Luck and Edward K. Vogel. 1997. The capacity of visual working memory for features and conjunctions. *Nature* 390, 6657 (1997), 279–281.

[27] Lotte Meteyard and Robert A.I Davies. 2020. Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language* 112 (2020).

[28] Akira Miyake, Naomi P Friedman, Michael J Emerson, Alexander H Witzki, Amy Howerter, and Tor D Wager. 2000. The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology* 41, 1 (2000), 49–100.

[29] Peter Møller Nielsen. 2019. An Investigation of Visual Features Effect on the Subjective Experience of Difficulty. (2019).

[30] Menno Nijboer, Jelmer Borst, Hedderik van Rijn, and Niels Taatgen. 2014. Single-task fMRI overlap predicts concurrent multitasking interference. *NeuroImage* 100 (2014), 60–74.

[31] R Core Team. 2018. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[32] Ulf-Dietrich Reips. 2002. Standards for Internet-Based Experimenting. *Experimental Psychology* 49, 4 (2002), 243–256.

[33] David Rosenbaum and Bill Bui. 2019. Does task sustainability provide a unified measure of subjective task difficulty? *Psychonomic Bulletin & Review* 26, 6 (2019), 1980–1987.

[34] RStudio Team. 2016. *RStudio: Integrated Development Environment for R.* RStudio, Inc., Boston, MA. http://www.rstudio.com/

[35] Thomas Alrik Sørensen and Søren Kyllingsbæk. 2012. Short-term storage capacity for visual objects depends on expertise. *Acta Psychologica* 140, 2 (2012), 158–163.

[36] Manoj Thomas and Claire I. Tsai. 2012-08-01. Psychological Distance and Subjective Experience: How Distancing Reduces the Feeling of Difficulty. *Journal of Consumer Research* 39, 2 (2012-08-01), 324,340.

[37] Anne M Treisman and Garry Gelade. 1980. A feature-integration theory of attention. *Cognitive Psychology* 12, 1 (1980), 97,136.

[38] Aalborg University. 2020. https://www.its.aau.dk/

[39] Jeremy Wolfe. 1994. Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review* 1, 2 (1994), 202,238.

[40] Jeremy Wolfe, Matthew Cain, and Avigael Aizenman. 2019-04-15. Guidance and selection history in hybrid foraging visual search. *Attention, Perception, & Psychophysics* 81, 3 (2019-04-15), 637,653.

[41] Jeremy M Wolfe. 2012. Saved by a Log: How Do Humans Perform Hybrid Visual and Memory Search? *Psychological Science* 23, 7 (2012), 698–703.

[42] Jeremy M Wolfe, Avigael M Aizenman, Sage E.P Boettcher, and Matthew S Cain. 2016. Hybrid foraging search: Searching for multiple instances of multiple types of target. *Vision Research* 119 (2016), 50–59.

[43] Jeremy M. Wolfe, Sage E. P. Boettcher, Emilie L. Josephs, Corbin A. Cunningham, and Trafton Drew. 2015. You Look Familiar, but I Don't Care: Lure Rejection

in Hybrid Visual and Memory Search Is Not Based on Familiarity. *Journal of Experimental Psychology: Human Perception and Performance* 41, 6 (2015), 1576.

[44] Weiwei Zhang and Steven J. Luck. 2008. Discrete fixed-resolution representations in visual working memory. *Nature* 453, 7192 (2008), 233–5.