Information constrained control analysis of eye gaze distribution under workload

Ron M. Hecht, Aharon Bar Hillel, Ariel Telpaz, Omer Tsimhoni, and Naftali Tishby

***Abstract*—We describe a novel model of human eye gaze behavior under workload, derived from the basic principle of information constrained control. The model assumes two distributions over the visual field: a saliency distribution which is non-goal oriented, and a reward task-related distribution. The eye gaze behavior is determined by the tradeoff between these two distributions, where the goal is to preserve the task-related constraints while remaining as close as possible to the saliency distribution representing a comfort zone. Based on minimum Kullback-Liebler divergence principles, the model gives rise to a family of gaze distributions controlled by a single tradeoff parameter. The model was evaluated experimentally in a driving simulator that consisted of an immersive environment with clear tasks and accurate monitoring capabilities. The findings confirm the theoretical predictions with respect to the low rank manifold and order relations in the data. We show that the model can be used to visualize the unknown reward function associated with a task, and predict human workload based on gaze pattern.**

***Index Terms*—Information constrained control, eye gazing distribution.**

I. Introduction

A major portion of human visual abilities is enabled solely in the receptive field of the fovea, thus making its fixation regions the most important instance of active sensing in humans. The human visual system produces gaze behavior via a complex interaction of bottom-up and top-down processing streams. Although the bottom-up process guides gaze toward salient regions, in the presence of a demanding task, saliency driven processing is partially overridden by a more task-related top-down pattern.

Numerous computational models have been put forward to describe workload-free gaze behavior, mainly using the concept of a non-task related saliency model [1] [2] [3]. These have been validated both as components of computer vision algorithms [2] [3] and models predicting human behavior [1]; see [4] for a review. Task-related attention processes have mainly been addressed using qualitative models in the fields of cognitive psychology and human factors [5] [6] [7] [8] [9] [10]. Experimental studies [6] [5] have shown that when subjects engage in a task that includes high levels of perceptual load, irrelevant visual distractors are not attended to at all, a phenomenon which was termed 'early selection'. There is an ambiguous relationship between cognitive workload, Working Memory (WM) workload and attending to irrelevant distractors. In [11], cognitive workload was shown to cause irrelevant distractors to be more attended to; however, in [12][13] mixed effects were reported. Some WM manipulations were consistent with the effect observed in [11], but others had the opposite effect.

There are many ways to explore the relationship between WM and the visual system under workload. The workload can be imposed by selecting a well-defined controlled environment. Alternatively, a real-world environment such as driving can be used. The behavior of the visual system can be measured at both higher levels of attention and lower levels of eye gaze directions. These differences may account for some of the discrepancies in the literature.

Quantitative analysis has gained momentum over the last few years: databases have been collected, tasks defined, and prediction algorithms proposed and evaluated. The databases cover a broad set of domains. For example, [14] [15] is composed of everyday activities such as talking on the phone and eating. The database in [16] recorded eye gaze distributions collected during the classification task of identifying breeds of dogs. The Dr(eye)ve database focuses on driving scenarios, and was collected on the road. It has roughly half a million frames. Some frames include GPS and speed data as well [17]. Eye gaze related databases can be used to obtain insight into human behavior, predict workload, and better understand the basic features of attention. The recent trend toward deep learning in the field of machine learning has led to the introduction of a set of algorithms to model both saliency distributions and eye gaze behavior under task loads. [18] surveys the use of a CNN (Convolutional Neural Network) for salience estimation. An example of task related gaze modeling was presented by Murabito et al. [16]. They trained a CNN to predict gaze distribution while performing a task and showed that this distribution was very different from a distribution collected while no task was present. In addition, they explored the gaze distributions where the resolution of the original image was blurred.

In this article we suggest a simple quantitative model that accounts for the tradeoff between saliency-based and task-related gaze behavior, and show that this model accurately captures human gaze behavior in a set of controlled experiments. We draw on a reinforcement learning variant introduced by [19] [20], known as the Information Constrained Control (ICC). The ICC can be seen as a Partially Observed Markov Decision

R. M. Hecht is with the Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel and Advanced Technical Center Israel, General Motors, Israel e-mail: hadas-ron@gmail.com.

A. Bar Hillel is with Ben-Gurion University of the Negev, Israel.

A. Telpaz is with Advanced Technical Center Israel, General Motors, Israel.

O. Tsimhoni is Warren Technical Center, General Motors, USA.

N. Tishby is with the Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel.

Manuscript received September 20, 2018; revised February 18, 2019, April 30, 2019.

Process (POMDP) that models the interaction between the perception, states, and actions of an organism. It focuses on the information flow between the organism and its inner state and the environment. The observations are the transmissions of information from the environment to the organism, and the actions and their control are the transmissions in the other directions. Note that the ICC differs from the POMDP in its ability to represent two-sub goals pursued by an organism by maximizing the reward while minimizing the control complexity to perform the action.

Based on the ICC, our model formally describes the tradeoff between saliency-based and task-related gaze behavior as a constrained minimum divergence problem. Saliency-based gaze behavior is formalized as a distribution over the possible gaze locations, which are considered to be the set of possible actions. This is a comfort-zone distribution: the human agent tries to stay as close to it as possible. The task-related signal is introduced using a reward function defined over the action set, such that each possible gaze location is associated with a certain task-related reward. In order to meet task demands, the agent needs to achieve a certain average reward level. The agent's policy is determined by finding the distribution nearest to the comfort zone, in the Kullback-Liebler Divergence sense, which still satisfies the task-related linear constraint. Different levels of workload, introduced by more stringent task conditions or performance of a secondary task in parallel, correspond in this framework to changes in a single problem parameter: the required reward level.

The suggested model is simple, but because of this simplicity it has non-trivial consequences leading to non-trivial empirical predictions. First, it intuitively predicts that average saliency levels of human agents under task pressure will drop monotonically with the level of workload introduced. Second, given two or more policies measured using different workload levels, the implicit reward function can be extracted up to multiplicative and additive constants, thus enabling reward visualization and model validation. The third important implication of the model is that the agent policies, viewed as log-probability vectors, constitute a curve in a three-dimensional linear subspace (spanned by the saliency distribution, the reward function and the unity vector). Importantly, this low manifold structure allows for easy prediction of workload levels from gaze patterns.

We tested this approach empirically in a driving task in a simulator environment which is realistic and immersive, but still enables the introduction of controlled workloads and exact measurements of eye gaze behavior. Workload in our experiments was manipulated by using driving related parameters: road curvature level, location along the route, driving speed, and by performance of a parallel secondary task. Our analysis shows the remarkable fit of the model to the empirically recorded gaze behavior of 18 subjects. Three types of higher workload caused the drivers to deviate further from the saliency model. In certain cases, the reward function could be visualized, and corresponded to a task-relatedness intuition. The effective rank of the empirical policy is shown to be close to three. We then tested the model's prediction abilities using a binary classification task: given two gaze patterns, determine which corresponds to driving with a higher cognitive load. The model-based classifier was able to solve the problem with close to a 90% success rate, and was found to outperform a baseline classifier that relied solely on standard deviations of eye position considerations.

This article contributes to the literature by introducing a simple quantitative model unifying bottom-up saliency-based gaze models and top-down task-related behavior, in addition to confirming the model's fit to empirical human gaze data. Beyond its empirical success, our model provides an elegant way of deriving gaze behavior from a very basic rationale: the Information Constrained Control (ICC) [19] [20] and the Minimum Discrimination Information (MDI) principles [21], a natural extension of the principle of insufficient reason. Thus, human gaze behavior is 'simple' in the sense that it deviates the least from the prior saliency-based gaze distribution.

II. The Information Constrained Control Model

This section presents the formal foundations to the approach used in this article which is based on [20] [19] [22] [23] [24] [25]. In a nutshell, the ICC is a satisfier model. It finds a balance between contradictory goals. The goal of the ICC is not only to maximize the reward an organism receives; its ultimate goal takes the motivation to minimize the complexity and information needed to execute a selected behavior into account as well. In the rest of this section, we introduce a single-state derivative to the ICC from scratch and build on top of it. This derivative is formalized using a set of possible actions $A = a_1, ..., a_n$ (where n is the number of actions), and two functions defined over this set: a prior distribution $Q(a)$ and a reward function $R(a)$. Intuitively, from a behavioral perspective, Q can be viewed as a comfort zone of behavior. In our case, it is the saliency distribution. The distribution $Q(a)$ is the unintentional behavior; i.e., $Q(a)$ is the probability of performing action a when no task pressure is present. Each action is associated with a specific value of task-related reward $R(a)$. The actual behavior $P(a)$ of the organism according to our model is defined as the solution to the following inference problem:

$$\begin{aligned} \hat{P}(a) = \arg\min_P \quad & \textstyle\sum_{a\in A} P(a) \log \frac{P(a)}{Q(a)} \\ s.t. \quad & \textstyle\sum_{a\in A} P(a) = 1 \\ & \textstyle\sum_{a\in A} P(a)R(a) \geq \theta \end{aligned} \tag{1}$$

The term $\sum_{a\in A} P(a) \log \frac{P(a)}{Q(a)}$ is known as Kullback Leibler Divergence (D_{KL}) [26]. The D_{KL} measures the information costs associated with selecting distribution P over the set a. This measure is relative to the distribution Q; i.e., Q is a baseline distribution. The information cost associated with selecting P to be identical to Q ($P = Q$) is zero.
Intuitively, the organism is constrained to achieve a certain level of task-related reward θ ($\sum_{a\in A} P(a)R(a) = \theta$), but under this constraint, it minimizes the D_{KL} of its behavior from prior behavior $Q(a)$. Eq. 1 describes a constrained minimization problem. A common way to solve such problem is to transform it into a different unconstrained minimization problem (known as a Lagrangian, and denoted by $L\{P(a)\}$) that has the same solution, and solve it. This methodology is

known as Lagrange multipliers [27]. The Lagrangian contains both the constraining terms and the minimized term as a single term that is later minimized. It is shown in the following equation:

$$L\{P(a)\} = \sum_{a\in A} P(a)\log\frac{P(a)}{Q(a)} \quad (2)$$
$$-\gamma\left(\sum_{a\in A}P(a)-1\right)-\beta\left(\sum_{a\in A}P(a)R(a)-\theta\right)$$

The minimum of the Lagrangian is found by deriving it with respect to β, γ and the probability of each action $P(a)$. The solution with respect to $P(a)$ has the following form:

$$P(a) = \frac{Q(a)e^{\beta R(a)}}{Z(\beta)} \quad (3)$$

where Z is a normalization factor assuring that $\sum_{a\in A}P(a) = 1$.

Insights can be gained by looking at β as a tradeoff parameter selected implicitly by the organism. It provides a tradeoff between the maximization of the reward and the proximity to the prior distribution Q. In this case, Eq. 1 has the following form:

$$\hat{P}(a) = \quad (4)$$
$$\arg\min_P \quad \sum_{a\in A}P(a)\log\frac{P(a)}{Q(a)} - \beta\left(\sum_{a\in A}P(a)R(a)\right)$$
$$s.t. \quad \sum_{a\in A}P(a) = 1$$

A. Divergence from saliency Q

Under this formulation, the organism's policy has one degree of freedom: the change in θ. The distribution of actions should be similar to $Q(a)$ when little reward is needed and should drift away when higher reward demands are set. We associate the reward constraint parameter θ with workload. A high task-related workload means that the task is more demanding, which is formulated as a demand for higher θ values in our model. The indicator for the workload is thus the likelihood of the organism's behavior (actions) according to the distribution $Q(a)$. At a lower workload, the likelihood of $Q(a)$ should remain high because the organism behaves according to the non-task related behavior, whereas at a higher workload, the likelihood of $Q(a)$ should be lower. For ease of notation, we refer to the likelihood score of samples according to $Q(a)$ as the Q-score.

B. Estimating the reward function

In cases where the reward function is unknown and several distributions obtained under varying workload conditions are observed, the reward function can be determined up to multiplicative and additive constants, thus enabling its visualization. By taking the log of the result of Eq. 1 and rearranging terms, we can see that

$$\log\frac{P(a)}{Q(a)} = \beta R(a) + \log Z(\beta) \quad (5)$$

In other words, there is a linear relationship between the reward and the log likelihood ratio. This relationship can be extended to the likelihood ratio between behavioral distributions produced under different workload conditions: $P_h(a)$, $P_m(a)$, and $P_l(a)$ (where h,m,l are the high, medium and low workloads respectively). These three behavior patterns can be expressed as:

$$\log\frac{P_{h,m,l}(a)}{Q(a)} = \beta_{h,m,l}R(a) + \log Z_{h,m,l} \quad (6)$$

where $\beta_{h,m,l}$ and $Z_{h,m,l}$ define three different linear equations, one for each choice of h,m,l. Subtracting the equation for low workload from the equation for high workload yields a $Q(a)$ free solution:

$$\tilde{R}(a) := \log\frac{P_h(a)}{Q(a)} - \log\frac{P_l(a)}{Q(a)} = \log\frac{P_h(a)}{P_l(a)} = \quad (7)$$
$$= (\beta_h - \beta_l)R(a) + \log(Z_h - Z_l)$$

In Eq. 7 we see that the log likelihood ratio between policies that differ in reward level estimates the reward up to multiplicative and additive constants. We denote this estimate by $\tilde{R}(a)$. Intuitively, $\tilde{R}(a)$ maintains a linear relation with respect to $R(a)$. The $\tilde{R}(a)$ values themselves are relative and only useful when compared to other $\tilde{R}(a)$ values. This might seem like a limitation; however, recall that in the larger set of RL tasks the reward is defined as relative in the first place. More formally, the average property holds for $\tilde{R}(a)$:

$$\frac{\tilde{R}(a_i)+\tilde{R}(a_j)}{2} =$$
$$(\beta_h - \beta_l)\frac{R(a_i)+R(a_j)}{2} + \log(Z_h - Z_l) \quad (8)$$

For any i,j

Low rank structure

The results of Eq. 1 can be rewritten as:

$$\log P(a) = \log Q(a) + \beta R(a) - \log Z(\beta) \ \ \forall a \quad (9)$$

Denote by $\log\vec{P} = [\log P(a_1), \cdots, \log P(a_n)]$ the vector of log likelihood of actions. Similar vectors can be defined for the prior distribution $\log\vec{Q}$, and the reward $\vec{R}$. We have:

$$\log\vec{P} = \log\vec{Q} + \beta\vec{R} - \log Z(\beta)\vec{1} \quad (10)$$

$\log\vec{P}$ located in the subspace spanned by $\log\vec{Q}$, $\vec{R}$ and $\vec{1}$. The latter equation shows that regardless of the number of actions, all the policies generated by an ICC model are located on a 3-dimensional sub-space within the policy space. This type of expected behavior can be verified on real data.

C. Prediction of workload from gaze pattern - $w_h w_l$ score

The reward can be eliminated by using the solution to Eq. 1 twice (once for a high workload condition and once for a medium workload condition); by rewriting the relationship between $P_h(a)$ and $P_m(a)$:

$$\frac{\log\frac{P_h(a)}{Q(a)} + \log Z(\beta_h)}{\beta_h} = R(a) = \frac{\log\frac{P_m(a)}{Q(a)} + \log Z(\beta_m)}{\beta_m} \quad (11)$$

where β_m, β_h are the β values associated with the medium and high workload conditions. By transforming to a vector representation, the equation above becomes:

$$\log \vec{P}_m = \frac{\beta_m}{\beta_h} \log \vec{P}_h + \left(1 - \frac{\beta_m}{\beta_h}\right) \log \vec{Q} + \left(\frac{\beta_m}{\beta_h} \log Z(\beta_h) - \log Z(\beta_m)\right) \vec{1} \quad (12)$$

In a manner similar to Eq. 12, we can define the following equation:

$$\log \vec{P}_m = \frac{\beta_m}{\beta_l} \log \vec{P}_l + \left(1 - \frac{\beta_m}{\beta_l}\right) \log \vec{Q} + \left(\frac{\beta_m}{\beta_l} \log Z(\beta_l) - \log Z(\beta_m)\right) \vec{1} \quad (13)$$

$\log \vec{Q}$ can be eliminated by combining Eq. 12 and Eq. 13.

$$\log \vec{P}_m = (1-w) \log \vec{P}_l + w \log \vec{P}_h + c_{\beta_h \beta_m \beta_l} \vec{1} \quad (14)$$

where the coefficients of $\log \vec{P}_h$ and $\log \vec{P}_l$ are denoted as w and $(1-w)$ respectively,

$$(1-w) = \frac{\beta_m - \beta_l}{\beta_h - \beta_l} \quad (15)$$

and

$$w = \frac{\beta_h - \beta_m}{\beta_h - \beta_l} \quad (16)$$

and the constant

$$c_{\beta_h \beta_m \beta_l} = \frac{(\beta_h - \beta_l)(\beta_m - \beta_l)}{\beta_h \beta_m (\beta_h - \beta_m)} (\beta_h \log z(\beta_l) - \beta_l \log z(\beta_h)) \quad (17)$$

The result shows that all possible policies are located in the subspace spanned by the vectors $\log \vec{P}_h$, $\log \vec{P}_l$, and $\vec{1}$. As observed in the previous section, this is a 3-dimensional subspace. Furthermore, given that $\beta_h > \beta_m > \beta_l$, the two coefficients of w and $(1-w)$ are between zero and one ($0 \leq w \leq 1$) and sum to one. In other words, all the solutions are a convex combination of the highest and lowest workload solutions. Later, we refer to w as w_h, the high workload component and to $1-w$ as w_l as the low workload component.

Intuitively this finding makes more sense if we explore the 2-dimensional subspace spanned by the coefficients w of $\log \vec{P}_h$ and $\log \vec{P}_l$. In this space, the subspace of possible policies forms a line. Specifically, Eq. 14 suggests that one mechanism to estimate workload for a given log-likelihood vector $\log \vec{P}$ would be to calculate a linear regression between it and the vectors $\log \vec{P}_h$, $\log \vec{P}_l$, $\vec{1}$. The workload for $\log \vec{P}$ is w, the coefficient in the regression of $\log \vec{P}_h$.

III. Method

The experiments were carried out in a driving simulator, an environment that mimics a high workload everyday task. This environment enables accurate gaze pattern estimation while introducing variable visual stimuli and variable levels of task-related workloads.

Fig. 1: Simulator setup - National Advanced Driving Simulator (NADS) MiniSim with three screens.

A. Participants

Twenty healthy volunteers (9 females), ranging in age from 25 to 59 (Mean = 33), were recruited to participate in the study. The participants were required to have a valid driving license for at least three years, and confirm that they drove on a daily basis. All were naïve to the purpose of the study. Participants stated they had normal (or corrected to normal) vision. Prior to the start of the experiment, the participants gave their informed consent in compliance with the guidelines of the General Motors Institutional Review Board. At the end of the experiment, each participant was paid a fixed fee of 200 NIS (about 50 USD).

B. Apparatus

1) Driving Simulator: The experiment was conducted in a low fidelity driving simulator, the NADS Mini-Sim, which is a driving simulator developed by the National Advanced Driving Simulator (NADS) and the University of Iowa [28]. The simulator consists of three 42" monitors which are used to display the frontal and side road views, yielding a 130° field of view (see Fig. 1). An additional screen mounted in the simulator cockpit dashboard was used to present the gauges on the instrument panel. Simulator-related sounds were delivered through a 2.1 audio system. Just like real world vehicles, the simulator was equipped with a steering wheel, as well as gas and brake pedals, which allowed the participants to operate the simulator vehicle naturally. The simulator software logs information on the participants' driving behavior and location on the route at a $60Hz$ rate. In addition, the images displayed on the three simulator monitors are captured and saved at an average rate of about $10Hz$.

Eye Tracking: The Smart-Eye pro 5 eye tracking system (manufactured and developed by Smart Eye AB, Gothenburg, Sweden; http://smarteye.se/) was used to track the participants' eye movements and point of gaze during the experiment. The system is composed of two IR cameras and IR LEDs. The two cameras were positioned on the upper part of the dashboard facing the participant. This eye-tracking system exploits corneal reflections from the IR LEDs at the center of the pupil to estimate the gaze intersection with the screens. The calibration process of the system takes place in two stages. In the first stage, the system generates a head model of the participant, followed by a gaze calibration. During the gaze

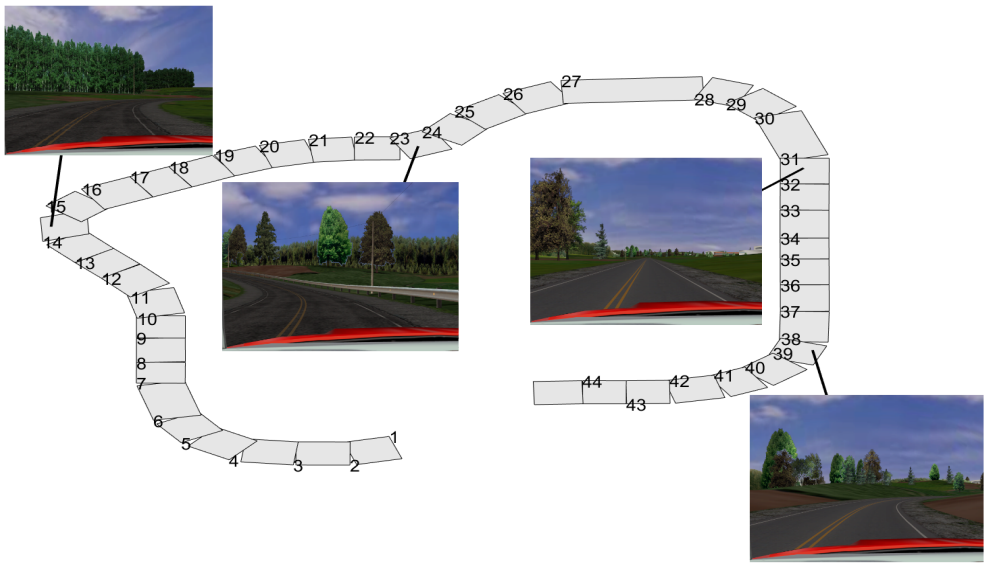


Fig. 2: The route driven in the experiment. The route started at segment 1 and ended at segment 44. The snapshots depict several representative segments along the route.

calibration phase, the participants are asked to look at twelve different locations in the vehicle cockpit. The process ends when the calibration quality is high enough, as determined by the average and standard deviation of the shift between the recorded gaze points and the 12 target locations. Specifically, the average and standard deviation were set to below 2°. The eye tracking data were collected at a rate of 60Hz.

C. Design and Procedure

Prior to the start of the experiment, the participants were requested to fill in a Simulator Sickness Questionnaire (SSQ) [29] to help avoid simulator sickness while driving in the simulator. Based on the results, two participants were eliminated. This left us with 18 participants. After the eye tracking calibration was completed, the participants engaged in a seven minute training scenario which enabled them to experience and learn how to operate the simulator vehicle. After this training session, the participants were given instructions on the experimental scenarios. The study was a within-subjects design consisting of five experimental scenarios divided into two blocks of two-three scenarios each. In all five scenarios, the participants drove an identical route. Fig 2 shows the driving path used in the experimental scenarios. The route path length was about five miles. The entire drive was in cruise control mode so that driving was at constant speed regardless of how the driver behaved. The driving took place in a simulated rural environment. The path involved different curvature levels including straight road segments. The scenery was relatively dull and monotonic without radical visual changes. In addition, for simplification purposes, no other vehicle except for the participant's vehicle were involved in the scenario. The participants were informed that they would not encounter any other vehicles during the scenario.

D. Independent Measures

To evaluate our model predictions for gaze pattern under different levels of workload, three independent factors were manipulated: N-Back task, vehicle speed, and curvature level:

TABLE I: Driving scenarios in the experiment. Within each block, the order of scenarios was counterbalanced across participants.

Block	Scenario	Speed	N-Back
1	1	35	without
	2	45	without
	3	55	without
2	1	35	with
	2	55	with

1) N-Back (2-Back) task: During the two scenarios in the second block, and for each participant, the experimenter read aloud random numbers between 1 and 9 at intervals of about 20 seconds. After each number that was read to the participants, they were asked to state whether the number was larger or smaller than the number presented to them 2 steps earlier. This task imposes additional working memory storage demands [10] and is known to interfere (i.e. increase drivers' mental workload) on a driving task in dual task settings [8]. In the first block, there was no secondary task. Since driving improves over time, administering the secondary task during the second block would be indicative that N-Back had an even more significant effect.

2) Vehicle speed: As shown previously, higher speed levels are associated with higher cognitive workload [30] [31] [32]. For instance, [31] found that the addition of a secondary demanding task caused drivers to reduce their vehicle speed, suggesting that it enabled them to allocate more cognitive resources to the secondary task. Therefore, within each block, in one scenario the speed was set to 35 MPH, in another scenario it was set to 55 MPH, and in the first blocks a 45 MPH scenario was added (Table I). The order of the scenarios within each block was counter-balanced across participants. The speed values were selected in such way that it would be possible to drive the entire route without having to press the brake pedal.

3) Curvature level: [33] reported a relationship between visual demands and curvature and showed in particular that there is an increase in task demand even before the curvature starts. For this reason, the route was made of several curved and straight segments. As shown in Fig. 2, the path was divided into 44 driving segments and curvature was estimated in each of them.

E. Dependent Measures

1) Likelihood of Q (Q-score): The likelihood of Q was measured by evaluating the similarity between the observed gaze pattern and the expected gaze pattern according to the Graph Based Visual Saliency (GBVS) model [1]. Similarity was estimated on a frame-by-frame basis and then averaged for each of the driving route segments. Specifically, for every frame, a GBVS activation map was generated and the activation values of the pixel in the participant's (x, y) gaze location coordinates (Q score) were averaged for each driving route segment (see example in Fig. 3). In some instances for individual participants, there were several gaze location samples within the same frame. In these cases, the Q-score represents the average of the log GBVS probability value for these samples. This analysis was only conducted for the central screen in the simulator because the eye tracking accuracy was lower for the side screens. This did not create bias in the analysis outputs since the proportion of gazes to the side screens was very low. The Q-scores were normalized by subtraction of the participants' mean Q-score.

2) Statistical Analysis: To examine the statistical significance of the Q score predictions under different levels of workload, we conducted a repeated measure analysis of variance (ANOVA), with N-Back condition (with vs. without), Speed level (35 vs. 55), and driving route segment as within subjects factors. To ensure the quality of the data analysis presented in the results section, any region that did not include data from all participants was eliminated from subsequent analysis. Therefore, due to the technical limitation of the eye tracking system, only 21 regions were taken into account for the analysis presented in the results section. For each analysis, we report the effect size using the partial eta squared (η_p^2) where values from 0.13 to 0.40 typically represent medium to large effects, respectively.

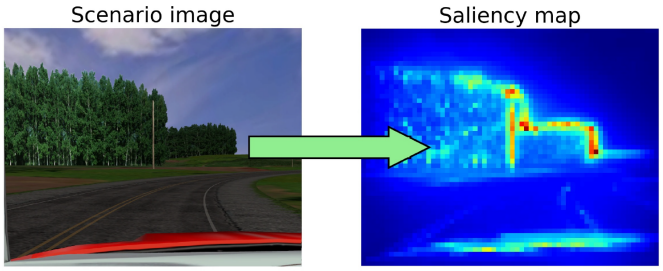


Fig. 3: Illustration of the generation process of the likelihood of Q scores. The figure on the left is a scenario image. The figure on the right shows the corresponding GBVS probability values. The Q score is simply the average of the log value of the pixels in the gaze location coordinates.

3) Visualization of the reward: The estimation of the reward $\tilde{R}(a)$ was measured using Eq. 7. This was measured by selecting two gaze log likelihood distributions.

4) Dimension of low rank structure: As Eq. 10 suggests, all the eye gaze distributions were expected to be located in a 3-dimensional subspace. Here, the dimension of the subspace was estimated using Principal Component Analysis (PCA) [34]. More specifically, the drivers' actions, which are the equivalent of sight locations, were measured at high resolution (dimension 1024X1280). Later, we down-sampled the raw data to concentrate the probabilistic mass (dimension 21X26). Then, rarely observed actions were removed, yielding the policy vector $\log \vec{P}$ in a low dimensional action space (dimension 21). A $\log \vec{P}$ vector was generated for each driver, driving condition, and road segment, yielding roughly one hundred vectors per segment. For each segment, based on its roughly hundred vectors, a covariance matrix was estimated. The PCA algorithm receives a covariance matrix as input, and identifies a set of orthogonal vectors with high variance. These vectors are known as eigenvectors. Each eigenvector is associated with a scalar known as an eigenvalue. The eigenvalue indicates the importance and variance of that direction. The dependent measures are the set of eigenvalues for each segments.

IV. Results

For one participant there were abnormal Q scores exceeding more than 3 absolute standard deviations from the sample mean. Therefore, this participant was excluded from further analyses, leaving us with 17 participants.

A. N-Back, speed and driving route segment main effects

As seen in Figure 4, the Q score was lower (M = 0.16, SD = 0.006) in the N-Back condition compared to the without N-Back condition (M = 0.17, SD = 0.008). This difference between the two conditions was significant ($p < 0.05, \eta_p^2 = 0.28$). As stated in [13], the effect of WM manipulation can be similar or opposite to the one observed at [11]. Here the effect was opposite to [11], and was similar to perceptual workload. Similarly, speed had a significant effect on the Q score, with higher Q score values when the speed was set to 35 MPH (M = 0.17, SD = 0.007) compared to 55 MPH (M = 0.16, SD = 0.007); $p < 0.05, \eta_p^2 = 0.28$. Not surprisingly, region was also found to be a significant factor ($p < 0.0001, \eta_p^2 = 0.81$), indicating that the Q scores varied among the 21 driving route segments that were included in the analysis. Since the main differentiator between the driving segments was the curvature level we conducted further analyses to examine the association between curvature and Q score.

All the participants performed the N-Back task better than random, indicating that the task was not ignored. The worst performer achieved 73% accuracy. There was no significant difference between the performance of the task at 35MPH and 55MPH. There was no monetary award for high performers on the N-Back task.

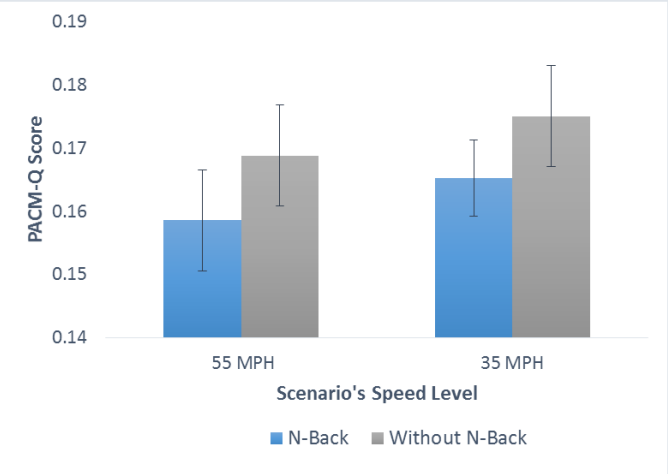


Fig. 4: Q score as a function of speed level and N-Back condition. Error bars denote standard error of the means.

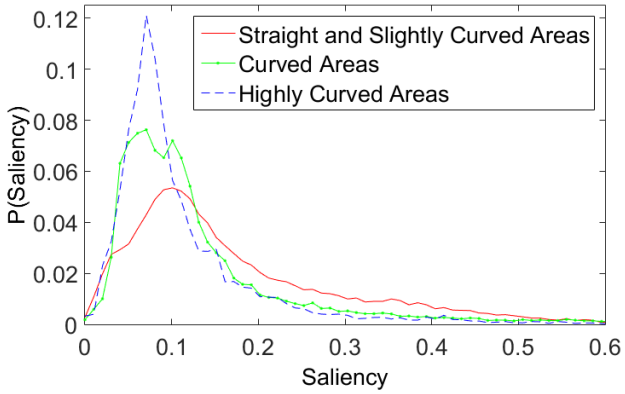


Fig. 5: Distributions of Q scores for three curvature levels.

B. Q-Score and Curvature Level

For each frame, we matched its Q score to the road curvature level. The analysis involved four scenarios $[35, 55]X[with, wo]$ and all participants, totaling more than 100,000 frames. The aim of this analysis was to examine the shape of the Q-score distribution under different levels of curvature. We divided the curvature spectrum into three parts: straight and slightly curved road, curved road, and highly curved road. As can be seen in Fig. 5, the higher the curvature (i.e. the more demanding the task), the lower the mean Q scores (Mean of straight and slightly curved: 0.176 Mean of curved road: 0.136 Mean of highly curved road: 0.115). Interestingly, the Q-score distribution became narrower under higher levels of curvature (SD of straight and slightly curved: 0.131 SD of curved road: 0.119 SD of highly curved road: 0.095).

C. Visualization of the reward

The mechanism to estimate reward was presented in the Method section; it boils down to estimations of the log likelihood ratio between two gaze distributions. Figure 6 illustrates the visualization capabilities of the ICC. The figure presents the ratio between the gaze log likelihood distributions. The distributions were smoothed based on likelihood and in places where no gazes were observed, only the background is seen. The remainder of the areas could be divided into cold and warm colored blobs. The warm colored blobs represent less important areas. The cold blobs are more important from a reward perspective. These areas were located on the road itself as expected, whereas the low reward areas were located on the sides of the road. The high reward area had a non-trivial shape.

D. Low rank structure

As explained in the Method section, the vectors used to estimate the covariance matrix were generated using a three-stage process. A visualization of the transformation during this process is presented in Fig. 7. The original high resolution distribution is presented in Fig 7.a (dimension 1024X1280). The down sampled distribution is presented in Fig. 7.b (dimension 21X26). The distribution after the removal of rare actions is presented in Fig. 7.c (dimension 21).

After the $\log \vec{P}$ vectors were extracted, a covariance matrix was estimated. The eigenvalues of the covariance matrix were used to gain insight into the dimensions of the subspace. Fig. 8 presents the sum of the first N normalized eigenvalues. The sum of the eigenvalues was averaged across the segments and the standard deviation is presented as well. The figure shows that the first three eigenvectors included more than 0.80 of the overall signal variance. It depicts the result when vectors from different participants were used; hence, the unexplained variance tail (the energy in eigenvectors > 3) can be attributed to differences between individuals.

E. Load prediction from gaze patterns

As described in the Model section, the coefficients of $\log \vec{P_l}$ and $\log \vec{P_h}$ (denoted w_l and w_h) sum to one. If they are chosen as the two extreme conditions, any other intermediate condition $\log \vec{P_m}$ is a convex combination of the two extremes. In our case the two extreme vectors corresponded to 35MPH without N-Back and 55MPH with N-Back.

A leave-one-out approach was selected to define the vectors $\log \vec{P_l}$ and $\log \vec{P_h}$. for each tested driver and route segment, where the reference vectors $\log \vec{P_l}$ and $\log \vec{P_h}$ were the average vectors in scenarios 35MPH without N-Back and 55MPH with N-Back of all other drivers in that segment except the tested driver. The tested driver vector was expressed as a linear combination of the two reference vectors and the constant $\vec{1}$. The results are presented in Fig. 9. It is clear that the actual findings are located along the line that was predicted by the model. The color represents the workload condition in different segments. Even a cursory glance shows the different colors at different locations across the curve.

As the model suggests, w_h is a workload measure. When closer to 0, low workload is assumed, and high w_h values are associated with high workload. Another measure for workload is the Standard Deviation (SD) along the x axis [8]. We used the SD to set a reference baseline for recognition performance of workload (see Table II). A comparison within pairs of conditions for that hierarchy of workload was conducted. About

(a) (b) (c)

Fig. 6: Subfigures (a),(b),(c) present $\tilde{R}(a)$ values - the reward visualization capability of the ICC. As suggested by Eq. 8, the reward is estimated up to multiplicative and additive constants. i.e. the units themselves are arbitrary and only the relations among them counts. In areas where almost no gazes were observed, only the original image is seen. Areas colors with hot colors represent less important areas (low $\tilde{R}(a)$ values), and areas with cold colors represent the more important areas (high $\tilde{R}(a)$ values). The cold colored areas (more important) are mostly located on the road itself, while the hot colored areas (less important) are mostly located off road. Since those values are relative, there is no need to present the exact value. Distributions were estimated based on data from all participants. Three distinct locations along the route where selected to demonstrate the route diversity. In addition, each image depicts reward visualization that was generated from contrast between different driving conditions. (a) was generated from the contrast between 55 MPH N-Back and 35 MPH N-Back conditions. (b) was generated from the contrast between 55 MPH N-Back and 35 MPH no N-Back conditions. (c) was generated from the contrast between 55 MPH N-Back and 45 MPH no N-Back conditions.

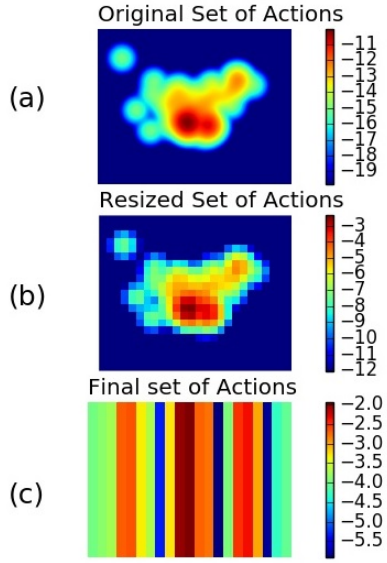


Fig. 7: Action definition. (a) Raw distribution of eye gazes. (b) Distribution after down-sampling. (c) Selection of dominant actions. All subfigures present log likelihood values.

800 comparisons were made within each pair of conditions. On average our approach outperformed the predictions using Standard Deviation measurements.

As mentioned in the introduction, WM workload is associated with mixed effects. In our experiment, it led to a reduction in the gaze SD. This suggests some disparity between our findings and some of the attentional findings in the psychology literature. Note, however, that the experiments were not identical and differed in their manipulations, measurements, and environments.

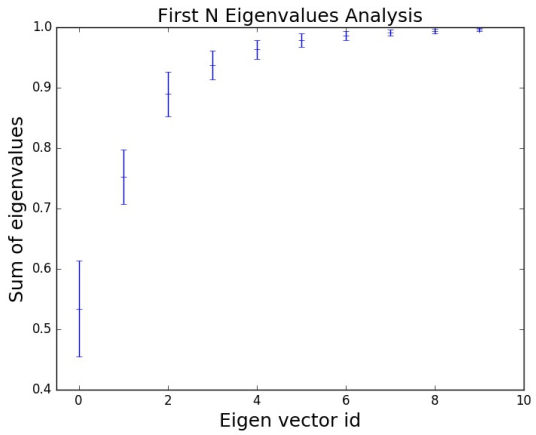


Fig. 8: Sum of the first N normalized eigenvalues of the distributions averaged over route segments.

V. Discussion

This study examined whether an information constrained control approach could account for gaze behavior under workload. Specifically, we modeled eye gaze while participants drove in a simulator under varying workload conditions. We assumed that the baseline (workload free) distribution Q would be estimated by a visual saliency mechanism (more specifically, by GBVS). We tested the efficiency of the model in terms of the relationship between workload and deviation from the baseline gaze distribution, estimation and visualization of the reward, existence of a low rank structure, and workload detection capabilities.

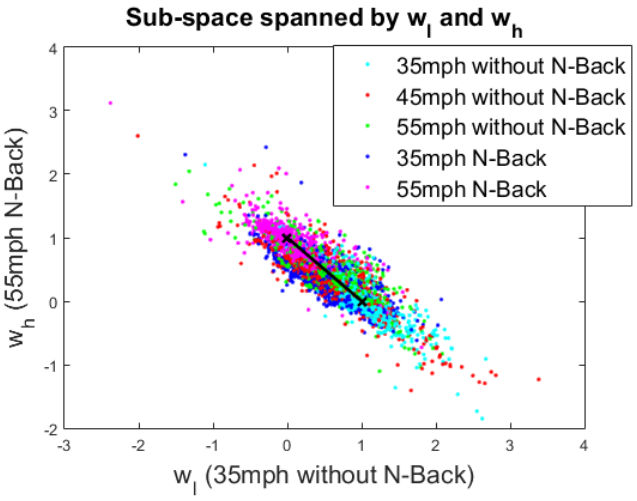


Fig. 9: Expected and observed tradeoff between high and low workload in the w_l and w_h sub-space. We refer to w as w_h and to $w-1$ as w_l. Recall that according to equations 15 and 16, the sum of w_h and w_l equals one. In addition, $0 \leq w_h, w_l \leq 1$.This suggests that theoretically all the possible combinations of w_h, w_l form a line between the point $(1,0)$ and the point $(0,1)$. This line is represented by the solid black line in the figure. The dots represent empirical estimations of w_h, w_l for different segments and driving conditions. The empirical data forms a scatter that is located close to the theoretical black solid line. The colors of the dots represent the driving conditions. The cyan represents the simplest ride condition (35 MPH no N-Back) and the magenta the hardest one (55 MPH N-Back). They are located on the opposite ends of the points cloud. The other driving conditions/colors are located more towards the middle of the cloud.

TABLE II: Workload recognition performance for different pairs of scenarios. SD is the standard deviation score, and $w_h w_l$ score as presented in Eq. 15

Scenarios Comparison	SD	$w_h w_l$ score
35 vs 45	0.64	0.81
35 vs 55	0.82	0.90
45 vs 55	0.77	0.63
35 vs 35 N-Back	0.81	0.85
35 N-Back vs 55 N-Back	0.81	0.96
35 vs 55 N-Back	0.93	0.99
45 vs 55 N-Back	0.88	0.95
55 vs 55 N-Back	0.69	0.91
Overall correct	0.79	0.88

A within-subject repeated measure factorial analysis of variance was used to test the deviation from the baseline distribution Q. The results showed that curvature, speed, and N-Back, which are known to increase driver workload, caused deviations from the baseline gaze distribution. We increased the workload by two manipulations of the main task and one manipulation by introducing a secondary task. All the manipulations caused the gaze pattern to shift from a more bottom-up pattern to a more top-down pattern, in a way which was clearly and significantly measurable using saliency Q scores.

These findings thus show that the model generates an accurate mechanism for reward visualization. This visualization was possible in cases where we had two gaze distributions that were estimated under two workload conditions (high and low). This visualization is of potential interest to researchers working on task-related perception in that it can identify the areas in the field of view that are important to the participants. Although the reward is only visualized up to a linear transformation, it is nevertheless a powerful technique, since our estimation maintains the reward-relations among the locations in the field of view. Whenever the reward associated with location A is higher than location B, the visualization value of A is higher than the value associated with point B. More specifically, the locations with the highest and the lowest rewards are the ones with the highest and lowest computed and visualized values. Hence the visualization image generated by the model identifies the most and the least important areas in a scene. Our examples show that in our experiments we were able to identify the important gaze locations for a driver entering a curve.

Our entire analysis is focused on the model space (distribution space). In this space, each eye gaze distribution is represented by a single point. For discrete distributions, the dimension of this space is the number of elements in the distribution minus one. For example, all the Bernoulli distributions $(p, 1-p)$ are points in a one-dimensional space. The larger the number of elements, the higher the dimension becomes. Workload manipulation causes a shift from one point in the distribution space (from one distribution) to another point in that space (another distribution). When a set of workload manipulations is applied, each manipulation yields a different point in the distribution space (different distribution). The ICC predicts that all these points are embedded in a low dimensional subspace in the distribution space. Fig 8 shows that the majority of the signal is located in the first dimension. A single dimension has more energy than all the rest of the dimensions. The first three dimensions have about ninety percent of the energy. This suggests an ICC-like behavior where all the distributions are highly bounded and do not tend to exit a small subspace.

Not only are all the points embedded in the small dimensional subspace, the ICC suggests that these points form a curved line in that subspace. This line can be projected to a straight line in a certain weighted space. In this geometric description, the task free distribution is the initial point of the curve, the goal-only distribution is its terminal point, and β is the index of the location along the curve.

The distance between any two points along the curve can be measured. Specifically, one can measure the distance from any point to both ends of the curved line (one end is associated with the lowest workload and the other with the highest workload). Intuitively, normalization of these distances generates the $w_h w_l$ score. This work suggests that the w_h,w_l are a measure of workload. This was compared to the SD of a gaze position measure. We tested these predictions capabilities and found that the w_h,w_l predictions

outperformed the commonly used SD of gaze position measure, and achieved a correct identification rate of 0.88. Fig. 9 visualizes the w_h,w_l distances used in this comparison. Several workload manipulations were compared, one of which was N-Back manipulation. This manipulation imposes a Working Memory (WM) workload and is very different from the cognitive workload presented at [35] which involved memorizing a set of digits before the task began and testing whether the digits were recalled after the task ended, and has no perceptual features, unlike N-Back. Overall, while some cognitive workload [11] [35] [36][37] [38] causes participants to pay more attention to goal-irrelevant stimuli, other produce the opposite effect [39][40]. In addition, it was shown that N-Back is associated with narrower eye gaze distributions [41][42][43]. it is worth bearing in mind, that there are differences across the experiments in terms of the manipulations used, the visual system measures, and experimental environment. Hence, the relationship among these factors is not straightforward, and further work is needed to address and understand these differences.

VI. Limitations and Conclusions

Overall, an information constrained control model with a saliency baseline emerges as an efficient tool to model driver gaze behavior under varying workload conditions. At a highly practical level, this model provides a set of useful tools for researchers.

One of the drawbacks of our experiments is that the commonly used mechanism of visual distractors and response time measurements was not used, which can preclude generalization. The effect we measured was achieved by varying other conditions, such as curvature. Future experiments should introduce visual distractors and test response time. However, the tendency of distractors to become very salient needs to be overcome, since this can alter the scene‘s baseline distribution. In addition, future studies should focus on modeling tasks with more complex reward distributions. This type of approach would be better able to distinguish between the SD-based model and our suggested model. A delta-like or Gaussian distribution of reward is the least useful for discriminating between the two. Future experiments should attempt to generate a bi-modal, or otherwise complex reward function, thus providing more interesting test cases for the theory.

More broadly, given the results presented here on the visual system and the findings presented in [22] concerning the role of information constrained control in word selection for dialog, it is worth exploring whether the information constrained control principle is effective in modeling other human behavioral contexts.

References

[1] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” in *Advances in neural information processing systems*, 2006, pp. 545–552.

[2] L. Guanbin and Y. Yizhou, “Visual saliency based on multiscale deep features,” in *CVPR*, 2015.

[3] M. N. Heess and A. Graves, “Recurrent models of visual attention,” in *NIPS*, 2014.

[4] A. Borji and L. Itti, “State-of-the-art in visual attention modeling,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 185–207, 2013.

[5] N. Lavie and S. Cox, “On the efficiency of visual selective attention: Efficient visual search leads to inefficient distractor rejection,” *Psychological Science*, vol. 8, no. 5, pp. 395–396, 1997.

[6] N. Lavie, “Attention, distraction, and cognitive control under load,” *Current Directions in Psychological Science*, vol. 19, no. 3, pp. 143–148, 2010.

[7] C. D. Wickens, “Multiple resources and mental workload,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, no. 3, pp. 449–455, 2008.

[8] T. W. Victor, J. L. Harbluk, and J. A. Engström, “Sensitivity of eye-movement measures to in-vehicle task difficulty,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 2, pp. 167–190, 2005.

[9] U. Cartwright-Finch and N. Lavie, “The role of perceptual load in inattentional blindness,” *Cognition*, vol. 102, no. 3, pp. 321–340, 2007.

[10] N. Cowan, “Processing limits of selective attention and working memory: Potential implications for interpreting,” *Interpreting*, vol. 5, no. 2, pp. 117–146, 2000.

[11] N. Lavie, “Distracted and confused?: Selective attention under load,” *Trends in cognitive sciences*, vol. 9, no. 2, pp. 75–82, 2005.

[12] N. Konstantinou, E. Beal, J.-R. King, and N. Lavie, “Working memory load and distraction: dissociable effects of visual maintenance and cognitive control,” *Attention, Perception, & Psychophysics*, vol. 76, no. 7, pp. 1985–1997, 2014.

[13] N. Konstantinou and N. Lavie, “Dissociable roles of different types of working memory load in visual detection.” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 39, no. 4, p. 919, 2013.

[14] S. Mathe and C. Sminchisescu, “Actions in the eye: Dynamic gaze datasets and learnt saliency models for visual recognition,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 7, pp. 1408–1424, 2015.

[15] M. Marszalek, I. Laptev, and C. Schmid, “Actions in context,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 2929–2936.

[16] F. Murabito, C. Spampinato, S. Palazzo, D. Giordano, K. Pogorelov, and M. Riegler, “Top-down saliency detection driven by visual classification,” *Computer Vision and Image Understanding*, 2018.

[17] S. Alletto, A. Palazzi, F. Solera, S. Calderara, and R. Cucchiara, “Dr (eye) ve: a dataset for attention-based tasks with applications to autonomous and assisted driving,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 54–60.

[18] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, “Salient object detection: A benchmark,” *IEEE transactions on image processing*, vol. 24, no. 12, pp. 5706–5722, 2015.

[19] J. Rubin, O. Shamir, and N. Tishby, “Trading value and information in mdps,” in *Decision Making with Imperfect Decision Makers*. Springer, 2012, pp. 57–74.

[20] N. Tishby and D. Polani, “Information theory of decisions and actions,” in *Perception-action cycle*. Springer, 2011, pp. 601–636.

[21] C. R. Smith, G. J. Erickson, and P. O. Neudorfer, *Maximum Entropy and Bayesian Methods*, 1991.

[22] R. M. Hecht, A. Bar-Hillel, S. Tiomkin, H. Levi, O. Tsimhoni, and N. Tishby, “Cognitive workload and vocabulary sparseness: Theory and practice,” in *Sixteenth Annual Conference of the International Speech Communication Association*, 2015.

[23] R. M. Hecht, A. Bar-Hillel, A. Telpaz, and N. Tishby, “Adaptation of eye gazing while driving using a perception - action cycle model,” in poster without proceedings at the ICRI-CI retreat, 2016.

[24] R. M. Hecht, A. Telpaz, G. Kamhi, A. Bar-Hillel, and N. Tishby, “Disentanglement of top-down and bottom-up processes using information constrained control,” in poster without proceedings at the fifth Conference on Cognition Research, 2018.

[25] R. M. Hecht, A. Telpaz, G. Kamhi, A. Bar Hillel, and N. Tisbhy, “Information constrained control for visual detection of important areas,” in *International Conference on Acoustics, Speech, and Signal Processing*, 2019.

[26] S. Kullback and R. A. Leibler, “On information and sufficiency,” *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[27] N. Cristianini and J. Shawe-Taylor, “An introduction to support vector machines and other kernel-based learning methods,” 2000.

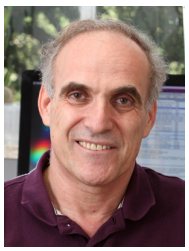
[28] Y. He. (2006) Technical report nads minisim driving simulator. [Online]. Available: http://www.nads-sc.uiowa.edu/publicationStorage/200610111034530.N06-025%20NADS%20MiniSim.pdf
[29] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The international journal of aviation psychology*, vol. 3, no. 3, pp. 203–220, 1993.
[30] C. D. Wickens, "Multiple resources and performance prediction," *Theoretical issues in ergonomics science*, vol. 3, no. 2, pp. 159–177, 2002.
[31] R. Srinivasan and P. P. Jovanis, "Effect of in-vehicle route guidance systems on driver workload and choice of vehicle speed: Findings from a driving simulator experiment," *Ergonomics and safety of intelligent driver interfaces*, pp. 97–114, 1997.
[32] A. H. Jamson and N. Merat, "Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 2, pp. 79–96, 2005.
[33] O. Tsimhoni and P. Green, "Visual demand of driving curves determined by visual occlusion," in *Vision in Vehicles 8 Conference, Boston, MA*, 1999.
[34] I. Jolliffe, "Principal component analysis," in *International encyclopedia of statistical science*. Springer, 2011, pp. 1094–1096.
[35] N. Lavie, A. Hirst, J. W. De Fockert, and E. Viding, "Load theory of selective attention and cognitive control." *Journal of Experimental Psychology: General*, vol. 133, no. 3, p. 339, 2004.
[36] J. W. de Fockert, G. Rees, C. D. Frith, and N. Lavie, "The role of working memory in visual selective attention," *Science*, vol. 291, no. 5509, pp. 1803–1806, 2001.
[37] N. Lavie and J. De Fockert, "The role of working memory in attentional capture," *Psychonomic bulletin & review*, vol. 12, no. 4, pp. 669–674, 2005.
[38] J. Rissman, A. Gazzaley, and M. Desposito, "The effect of non-visual working memory load on top-down modulation of visual processing," *Neuropsychologia*, vol. 47, no. 7, pp. 1637–1646, 2009.
[39] M. Rose, C. Schmid, A. Winzen, T. Sommer, and C. Büchel, "The functional and temporal characteristics of top-down modulation in visual selection," *Cerebral Cortex*, vol. 15, no. 9, pp. 1290–1298, 2004.
[40] K. K. Sreenivasan and A. P. Jha, "Selective attention supports working memory maintenance by modulating perceptual processing of distractors," *Journal of cognitive neuroscience*, vol. 19, no. 1, pp. 32–41, 2007.
[41] M. Niezgoda, A. Tarnowski, M. Kruszewski, and T. Kamiński, "Towards testing auditory–vocal interfaces and detecting distraction while driving: A comparison of eye-movement measures in the assessment of cognitive workload," *Transportation research part F: traffic psychology and behaviour*, vol. 32, pp. 23–34, 2015.
[42] B. Reimer, B. Mehler, Y. Wang, and J. F. Coughlin, "A field study on the impact of variations in short-term memory demands on drivers visual attention and driving performance across three age groups," *Human Factors*, vol. 54, no. 3, pp. 454–468, 2012.
[43] Y. Wang, B. Reimer, J. Dobres, and B. Mehler, "The sensitivity of different methodologies for characterizing drivers gaze concentration under increased cognitive demand," *Transportation research part F: traffic psychology and behaviour*, vol. 26, pp. 227–237, 2014.

Ron M. Hecht received his B.Sc. degree from the Hebrew University, Jerusalem, Israel, in 1996 and his M.Sc. degree in Computer Science from the University of Tel Aviv, Tel Aviv, Israel, in 2008. Currently Ron is a PhD student at the Hebrew University. He also holds an M.B.A. from the Tel Aviv University. He is a senior researcher at the User Experience Technologies Lab, General Motors Advanced Technical Center-Israel, Herzliya.

Aharon Bar Hillel is a senior lecturer at the Department of Industrial Engineering and Management at Ben-Gurion University, Israel. He earned his Ph.D. thesis in neural computation from the Hebrew University of Jerusalem in 2006. Since then he has been a machine learning and computer vision researcher at Intel Research (2006-2008), GM Research (2009-2012), and Microsoft Research (2013-2016). He is interested in interpretation, acceleration and decomposition of deep networks, and applications of computer vision and machine learning to various domains.

Ariel Telpaz is a senior researcher in the user experience technologies group at the General Motors Advanced Technical Center in Israel. His main expertise is in using analytical techniques and applying statistical methods to develop technologies to improve human-vehicle interaction. Ariel earned his PhD and M.Sc (2012 and 2008 respectively) from the faculty of Industrial Engineering and Management at the Technion, Israel Institute of Technology. His dissertation focused on the utility of physiological measures, specifically EEG and eye tracking, in predicting decision-making behavior. During 2012 he was a Leverhulme postdoc scholar conducting research on computational models of adaptive decision making. From 2009 to 2013 Ariel was an adjunct lecturer at the Technion, Israel Institute of Technology, where he taught courses on statistics, programming, research methods and human factors engineering.

Omer Tsimhoni received his M.S.E. and Ph.D. degrees in Industrial and Operations Engineering from the University of Michigan, Ann Arbor, in 1997 and 2004, respectively.
From 2005 to 2008, he was an Assistant Research Scientist with the Human Factors Division, University of Michigan Transportation Research Institute, and an Adjunct Assistant Professor with the Department of Industrial and Operations Engineering, University of Michigan. Since 2008, he has been working for General Motors. Currently, he is a Technical Fellow and a group manager in R&D. His research interests include transportation human factors, speech interfaces, adaptive interfaces, computational cognitive modeling of driving, driving safety, and automotive night-vision systems.



Naftali Tishby is the Flinkman professor of Computer Science and Brain Research at the Hebrew University of Jerusalem. Prof. Tishby is one of the leaders of machine learning research and computational neuroscience in Israel. He was the PI of Intels Collaborative Research Institute for Computational Intelligence (IICRI CI) in Israel. Tishby was the founding chair of the computer-engineering program, and a director of the Leibnitz research center in computer science, at the Hebrew university.
He received his PhD in theoretical physics from the Hebrew university in 1985 and was a research staff member at MIT and Bell Labs from 1985 and 1991. Prof. Tishby was also a visiting professor at Princeton NECI, University of Pennsylvania, and UCSB. His current research is at the interface between computer science, statistical physics, and computational neuroscience. He is best known for applications of Information Theory to Machine Learning and Neuroscience, in particular the Information Bottleneck principle. He is the winner of the Landau prize in Computer Science (2015) and the IBT prize for Mathematical Neuroscience (2019).