# Crossed Eyes: Domain Adaptation for Gaze-Based Mind Wandering Models

Robert E. Bixler
University of Notre Dame

Sidney K. D’Mello
University of Colorado Boulder

## ABSTRACT

The effectiveness of user interfaces are limited by the tendency for the human mind to wander. Intelligent user interfaces can combat this by detecting when mind wandering occurs and attempting to regain user attention through a variety of intervention strategies. However, collecting data to build mind wandering detection models can be expensive, especially considering the variety of media available and potential differences in mind wandering across them. We explored the possibility of using eye gaze to build cross-domain models of mind wandering where models trained on data from users in one domain are used for different users in another domain. We built supervised classification models using a dataset of 132 users whose mind wandering reports were collected in response to thought-probes while they completed tasks from seven different domains for six minutes each (five domains are investigated here: Illustrated Text, Narrative Film, Video Lecture, Naturalistic Scene, and Reading Text). We used global eye gaze features to build within- and cross- domain models using 5-fold user-independent cross validation. The best performing within-domain models yielded AUROCs ranging from .57 to .72, which were comparable for the cross-domain models (AUROCs of .56 to .68). Models built from coarse-grained locality features capturing the spatial distribution of gaze resulted in slightly better transfer on average (transfer ratios of .61 vs .54 for global models) due to improved performance in certain domains. Instance-based and feature-level domain adaptation did not result in any improvements in transfer. We found that seven gaze features likely contributed to transfer as they were among the top ten features for at least four domains. Our results indicate that gaze features are suitable for domain adaptation from similar domains, but more research is needed to improve domain adaptation between more dissimilar domains.

## CCS CONCEPTS

• **Computing methodologies** → Machine learning; • **Human-centered computing** → Human computer interaction (HCI); HCI design and evaluation methods; User models /+.

## KEYWORDS

Mind Wandering, Eye Movements, Domain Adaptation, User Modeling

*ETRA '21 Full Papers, May 25–27, 2021, Virtual Event, Germany*
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8344-8/21/05…$15.00
https://doi.org/10.1145/3448017.3457386

**ACM Reference Format:**
Robert E. Bixler and Sidney K. D’Mello. 2021. Crossed Eyes: Domain Adaptation for Gaze-Based Mind Wandering Models. In *2021 Symposium on Eye Tracking Research and Applications (ETRA '21 Full Papers), May 25–27, 2021, Virtual Event, Germany.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3448017.3457386

## 1 INTRODUCTION

The ideal user interface is sufficiently engaging that the user can focus until their task is completed. Despite our best attempts to prevent it, humans are subject to the phenomenon of *mind wandering* (MW), where their attention shifts from the current task to unrelated thoughts. Although the basis and proper conceptualization of MW are currently being debated [Seli et al., 2018b; Seli et al., 2018c; Christoff et al., 2018; Christoff et al., 2016; Fox and Beaty, 2019], it is widely acknowledged that humans can only attend to a limited amount of information at one time, so when users mind wanders on the task at hand, their performance on that task drops [Randall et al., 2014]. Indeed, it has been found that MW negatively correlated with performance during reading [Feng et al., 2013; Smallwood et al., 2007b; Smallwood et al., 2007a], signal detection [Robertson et al., 1997; Smallwood et al., 2004], and memory tasks [Seibert and Ellis, 1991; Smallwood and Schooler, 2006] to name a few. Further, because the rate of MW can be quite high for certain tasks (it ranges between 20-50% depending on the task and environment [Kane et al., 2007; Killingsworth and Gilbert, 2010; Schooler et al., 2004; Smilek et al., 2010; Seli et al., 2018a]), intelligent user interfaces would be well served to detect when MW occurs and deliver interventions in order to reengage users' attention [Mills et al., 2020; D'Mello et al., 2017; Mills et al., 2019].

One promising method of detecting MW involves analyzing users' *eye gaze,* which is an increasingly popular modality for research and applications in intelligent user interfaces [Nakano et al., 2013]. Eye gaze is typically processed and converted into eye movements, primarily fixations (stationary moments), saccades (rapid movements between fixations), and smooth pursuits (eyes following stimulus). It is useful for revealing user cognitive processing because the direction in which the eye gazes implies a user's locus of visual attention [Roda and Thomas, 2006]. Research has led to a multitude of eye gaze-assisted interfaces that use the information gleaned from eye gaze to improve user performance and experience, spanning domains such as driving [Lemercier et al., 2014; Fletcher and Zelinsky, 2009; Tawari et al., 2014], game playing [Kocur et al., 2020; Lankes et al., 2016; Sundstedt, 2012; Antunes and Santana, 2018], and learning [Hutt et al., 2021; Mills et al., 2020]. One class of these attention-aware user interfaces [Roda and Thomas, 2006] aim to use eye gaze to detect and combat MW in real time [D'Mello et al., 2016]. These approaches typically use supervised machine learning to develop gaze-based detectors of MW, which are subsequently embedded in the user interfaces [D'Mello et al., 2017; Hutt et al., 2021; Rosy et al., In Press]. This entails collecting adequate training data from users (MW labels along with eye gaze features) in a domain to develop models that can be applied to *new* users in the *same* domain.

Because gathering labeled data on a domain-by-domain basis can be expensive, one potential approach is to take advantage of *similar* data collected in a *different* domain. Here we consider a domain as referring to the type of stimulus and a task goal as an activity performed using that stimulus. For example, a model trained using data from users engaged in a scene viewing task (static images) could be used to detect MW while users are watching an instructional video (sequences of images). This approach is a subset of transfer learning called *domain adaptation* [Kouw and Loog, 2018], where a model is built with data from a *source domain* in order to classify data from a *target domain*. Importantly, this type of transfer learning requires that the task goal and features are the same. For example, learning about scientific topics could be accomplished through either reading a textbook or watching an instructional video. In both cases, the task goal is the same but features of the environment (the modality used to learn) will cause differences in eye movements since the stimuli differ. Domain adaptation seeks to discover commonalities between the data for each domain in order to build models that are compatible for both domains despite their differences.

**The goal of this work is to investigate the feasibility of building cross-domain models of mind wandering using eye gaze.** Our aims were to: (1) assess the extent to which models constructed from global, stimulus independent eye gaze features could transfer from one domain to another; (2) determine the effectiveness of locality features and domain adaptation techniques to improve cross-domain model performance, and (3) identify which features were most successful across domains. To investigate this, we collected eye gaze data and MW self-reports from users as they engaged in tasks involving processing static scenes, dynamic scenes, and reading text. We processed the data using standard eye movement filtration techniques [Voßkühler et al., 2008; Komogortsev and Karpov, 2013] and computed features for supervised machine learning models that aimed to distinguish cases of MW from normal attentive processing (not MW). The stimuli are representative of those that might be included in an intelligent user interface focused on facilitating learning and comprehension [D'Mello, 2016].

## 2 RELATED WORK

We focus on efforts to automatically measure mind wandering (MW). Researchers typically use thought sampling techniques where a user reports MW during a task voluntarily or in response to thought probes [Weinstein, 2018]. MW can also be measured retrospectively after a task using questionnaires [Weinstein, 2018]. Using these paradigms, MW has been investigated in a variety of domains including reading [Feng et al., 2013; Reichle et al., 2010; Smallwood, 2011; Unsworth and McMillan, 2013], film viewing [Kopp et al., 2016], interacting with learning technologies [Lindquist and McLean, 2011; Mills et al., 2015], and scene viewing [Krasich et al., 2018; Krasich et al., 2020]. Researchers have also developed MW detectors using a variety of modalities including physiology [Smallwood et al., 2004; Pham and Wang, 2015; Blanchard et al., 2014], neurological signals [Christoff, 2012; Christoff et al., 2009; Jin et al., 2019], acoustic signals [Drummond and Litman, 2010], behavioral measures [Franklin et al., 2011; Mills and D'Mello, 2015; Faber et al., 2018; Dias da Silva and Postma, 2020], eye behaviors [Krasich et al., 2018; Krasich et al., 2020; Bixler and D'Mello, 2016; Hutt et al., 2019; Hutt et al., 2017], and facial features [Bosch and D'mello, 2019; Stewart et al., 2017a].

Our approach to MW detection is based on the eye-mind link [Rayner, 2009; Rayner, 1998; Clifton et al., 2016] that suggests a close coupling between internal processing and eye movements. MW is characterized by a decoupling between internal thoughts and processing of external stimuli [Schooler et al., 2011; Smallwood et al., 2008], meaning the eye-mind link breaks down during MW. Thus, eye movement patterns can be used as indicators of MW. For example, researchers have found that fixations are longer and less frequent [Smilek et al., 2010; Reichle et al., 2010; Bixler and D'Mello, 2016; Frank et al., 2015] and that saccades have greater duration and amplitude [Krasich et al., 2018; Faber et al., 2020] during MW. This research serves as the basis for eye gaze-based models of MW, although there are conflicting accounts and patterns can differ based on domain and task goal [Faber et al., 2020].

Most MW detectors have been built and evaluated in a single domain, but we are interested in the ability of eye gaze-based models of MW to transfer between domains. Transfer learning entails building models from datasets where the training and testing sets differ [Kouw and Loog, 2018]. Domain adaptation is a subfield of transfer learning that focuses on problems where the training and testing set differ in their domains, but both share a task goal and set of features. For example, reading a paragraph on physics concepts and viewing a video on cell division would be different domains but could share a task goal (learning) and set of features (eye gaze).

A previous investigation of facial feature-based MW detectors by Stewart et. al. suggested some evidence of transfer between domains [Stewart et al., 2017b]. Data for this study came from two studies in which they recorded videos of users who self-reported MW freely while they either watched the narrative film *The Red Balloon* (1956) [Zacks, 2010] or read an excerpt from the book *Soap-Bubbles and the Forces which Mould Them* [Boys, 1890]. The researchers focused on developing MW models using facial expressions (e.g., facial action units such as eye blinks and dimplers). They found that within domain models and composite models built from both domains performed better than chance. The narrative film model achieved comparable performance on the scientific text data, but the scientific text model performed around chance level when classifying the narrative film data. However, adjusting the classification threshold from the default 0.5 to 0.3, resulted in a scientific text model with similar performance to the narrative film model when classifying the narrative film data. A feature analysis suggested that accuracy was influenced by lip tightener (AU23) and jaw drop (AU26) features.

To our knowledge, the Stewart et al. study is the only one that has investigated MW detection among domains. We considerably expand on this previous study by focusing on eye gaze data collected from the *same users* across five domains. We build on previous work in MW detection, most notably that of [Bixler and D'Mello, 2016;

Hutt et al., 2019], but our work is unique in that it is the first investigation of eye gaze-based MW models that are built from data in one domain and evaluated in another.

## 3 METHODOLOGY

We utilize data from an earlier study on eye gaze and MW [Faber et al., 2020]. However, this previous study did not build MW models as we do here.

### 3.1 Dataset

The participants were 132 students from a mid-sized university in the Midwest of the USA. Participants were seated in front of one of two setups consisting of a keyboard and mouse, computer screen on which stimuli were presented, and an eye tracker. One setup (*n* = 90) used a laptop to display stimuli and a Tobii EyeX eye tracker affixed below the laptop screen to track eye gaze. The other setup (*n* = 42) included a monitor spaced 60 cm. away from an EyeLink 1000 eye tracker with a tower mount for head stability. We collected data from these different setups to increase the generalizability of the models.

After entering the testing room and receiving verbal instructions, participants completed a calibration procedure and calibration test. Each of these consisted of nine fixed points appearing on the screen in a random order that participants were instructed to look at as they appeared. Eye gaze recorded during the calibration test was not used to calibrate the eye tracker, only to evaluate the calibration for post-hoc analysis.

Participants then received typed instructions, which consisted of seven tasks in which a different stimulus was presented in pseudo-random order for six minutes each. Each participant engaged with the same seven stimuli in random order; two were control stimuli and not analysed here. The five analysed stimuli were chosen due to their topicality to various user interfaces for learning and have been used in previous literature [Kopp et al., 2016; Krasich et al., 2018; D'Mello and Graesser, 2014; Hutt et al., 2017; Mills et al., 2016]. In all cases, the learning goal was to comprehend the content being displayed.

The stimuli are listed in Table 1 along with the corresponding domain and total number of MW instances after filtering out those unsuitable for computing features (details below). Stills from each stimulus are shown in Figure 1. The Diagram Set (Illustrated Text) stimulus consisted of two diagrams displayed for three minutes each. Each diagram contained an everyday device (e.g., doorbell) along with an expository text that participants were asked to comprehend, striking a blend between the reading domain and static scene processing. To this point, the Scenes domain was a series of six scenes of cities. It shared characteristics with the Diagram stimulus but did not include descriptive text. Finally, the Red Headed League (Text) stimulus was a series of pages from a Sherlock Holmes novel, presented at a varying rate between 5 and 15 seconds depending on content in order to be consistent with the other domains. The Red Balloon (Film) stimulus was the first six minutes of a French narrative film about a boy following a balloon. It is in the narrative domain since the film tells a story. This is different from the video lecture on population growth since this is in the expository (informative) domain. Like the Red Balloon stimulus, it was also a dynamic scene, but in one continuous shot. A further distinction between the film and video domains is that narrative films employ techniques such as camera cuts in order to guide attention, referred to as the *tyranny of film* [Loschky et al., 2015]. Thought probes similar to those standard in the field [Weinstein, 2018] were used to measure mind wandering at pseudo-random intervals of 90-120s from the onset of each stimulus. These thought probes occurred via an overlay that prompted participants to press a key to indicate if they were MW or not. Participants were instructed to report MW if they found themselves thinking about something other than the task when they received a thought probe. We obtained three thought probes per stimulus resulted in a total of 396 thought probes for each stimulus across all the participants for a total of 1980 data points across the five stimuli used. Upon completion of the main experiment participants completed a posttest and filled out questionnaires collecting demographic information and their perception of each task.

### 3.2 Data Processing

We built models with eye gaze data in short time windows (10, 20, 30, 40, or 50 seconds) prior to the thought probes. Unsuitable instances were deemed those with fewer than 80% valid eye gaze points or fewer than 4 fixations since this was considered the minimum number to compute several gaze features. Overall, only 13% of the data was discarded. Next, fixations and saccades were computed using standard eye movement detection algorithms including fixation dispersion threshold and saccade velocity threshold algorithms [Komogortsev and Karpov, 2013; Voßkühler et al., 2008]. We computed fixation qualitative, fixation quantitative, and saccade quantitative metrics [Komogortsev et al., 2010] from the data collected during the calibration test. We used these metrics to evaluate different parameter settings for each eye movement detection algorithm. Each algorithm performed similarly with minor trade-offs so we decided to use the dispersion based filter with a dispersion threshold of 57 pixels based on the algorithm used in the Open Gaze and Mouse Analyzer software [Voßkühler et al., 2008]

We computed features from the fixation filtered data within each window. We focused on global features based on previous work [Bixler and D'Mello, 2016; Hutt et al., 2019] that are not dependent on the stimuli, such as a prespecified area of interest or unique eye movements such as smooth pursuits that are not present in static stimuli. Computed features include nine statistical functionals (number, mean, median, minimum, maximum, standard deviation, range, skew, and kurtosis) computed from six distributions of fixation duration (ms), saccade duration (ms), saccade amplitude (pixels), relative saccade angle between subsequent saccades (degrees), absolute saccade angle in reference to the horizontal line (degrees), and saccade velocity (pixels/ms). We also computed fixation/saccade ratio, fixation dispersion, proportion of horizontal saccades, and blink count, resulting in 58 features. Fixation/saccade ratio was the total duration of fixations divided by the total duration of saccades. Fixation dispersion was calculated as the root mean square of the distance of each fixation to the average fixation position in the window. Proportion of horizontal saccades was the proportion of saccades that had an angle no more than 30 degrees above or below

**Table 1: Dataset characteristics including the total number of instances, the final instances for each dataset after data processing, the resultant % of missing instances, the number of instances for each class, and the corresponding MW rate.**

| Dataset (Stimulus) | Domain | Total Instances | Final Instances | Missing | MW Yes | MW No | MW Rate |
|---|---|---|---|---|---|---|---|
| Diagram Set (Illustrated Text) | Static Scene/Reading | 396 per domain | 350 | 12% | 187 | 163 | .47 |
| Red Balloon (Film) | Dynamic Narrative Scene | | 356 | 10% | 260 | 96 | .27 |
| Video Lecture | Dynamic Expository Scene | | 352 | 11% | 190 | 156 | .45 |
| City Scene (Image) | Static Scene | | 346 | 13% | 145 | 209 | .59 |
| Red Headed League (Text) | Reading | | 354 | 11% | 212 | 140 | .40 |

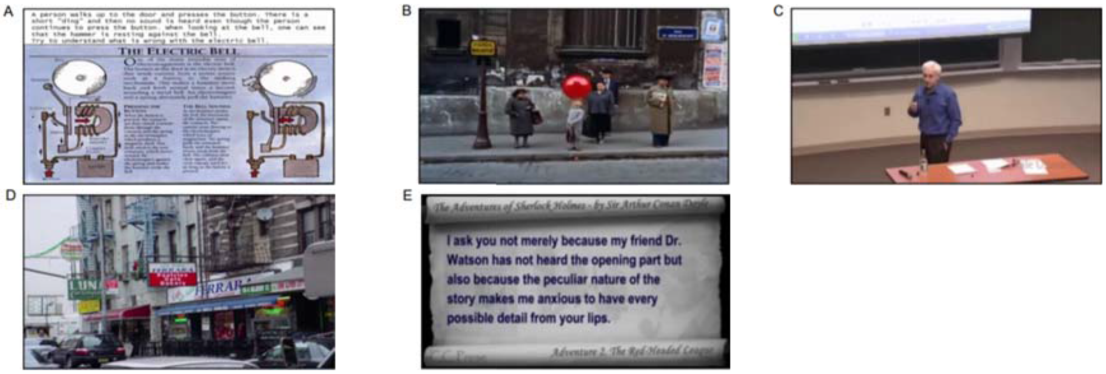

A
A person walks up to the door and presses the button. There is a short "ding" and then no sound is heard even though the person continues to press the button. When looking at the bell, one can see that the hammer is resting against the bell.
Try to understand what is wrong with the electric bell.
THE ELECTRIC BELL
B
C
D
E
The Adventures of Sherlock Holmes - by Sir Arthur Conan Doyle
I ask you not merely because my friend Dr. Watson has not heard the opening part but also because the peculiar nature of the story makes me anxious to have every possible detail from your lips.
Adventure 2. The Red-Headed League

**Figure 1: Example stills from each domain. A) Diagram Set (Illustrated Text), B) Red Balloon (Film), C) Video Lecture, D) Scene, E) Red Headed League (Text).**

the x axis. Pupillometry features were not explored as the EyeX lacked pupil diameter data.

## 3.3 Supervised Classification

*3.3.1 Parameter Exploration for Baseline Models.* Different parameter values may result in better performance for certain domains but using a consistent set of parameters allows direct comparisons between models. Hence, before building cross-domain models we performed a parameter exploration to select appropriate baseline models. The parameters we explored, their different possible values, and the value that we selected for the baseline models are shown in Table 2. Window size was varied in order to measure the effect different amounts of data had on model performance. Larger window sizes risk inclusion of data unrelated to MW, while smaller window sizes risk missing pertinent events. Balancing the class distribution can sometimes lead to improved accuracy so we built a set of models with balanced classes using the Synthetic Minority Oversampling Technique (SMOTE) [Chawla et al., 2002] to compare to those without any balancing. Importantly, SMOTING was only done on training data; testing distributions were not modified. Outliers can also skew models, especially with smaller amounts of data, so we built a set of models that used a process called winsorization where outliers for each feature were replaced with a value three standard deviations away from the mean value for that feature. We built sets of models where a certain proportion of the worst performing features were removed in order see if the models could be improved with a smaller set of predictive features. We used a variety of classifiers as there is no prior indication of which would work best, including support vector machines with a linear kernel, k-nearest neighbors, logistic regressions, Bayesian networks, and random forests. Each classifier used the default hyperparameters used in WEKA [Hall et al., 2009] except for the k-nearest neighbors, which had k set to 5. We chose the parameter settings in Table 2 for the final models because it was among those best models and used SMOTE to address class imbalance.

*3.3.2 Feature Selection and Cross-Validation.* We selected parameters using participant-level 5-fold cross-validation (see Figure 2). For a given fold and iteration the training set consisted of data from the source domain and from participants that were not within the fold. The testing set consisted of data from the held-out participants from the source domain (for baseline models) and from each target domain (to evaluate transfer). Data processing steps including the removal of multicollinear features, within-domain normalization, and within-domain outlier removal were applied to the training set within each fold. We determined multicollinear features by computing the variance inflation factor (VIF) for each feature and successively removing the feature with the highest VIF above a VIF of 5. The data were z-score standardized within each domain to address gaze-related domain differences independent of MW.

After these data processing steps were completed, correlation-based feature selection [Hall, 1999] was used to rank each feature. Feature selection was done five times with a random 66% of the data in the training set. Features were ranked and their rankings

**Table 2: Selected and possible values for window size, sampling method, outlier removal, feature selection cutoff, and classifier model building parameters.**

| Parameter | Selected | Other Options |
|---|---|---|
| Window Size | 40 seconds | 10, 20, 30, 50 |
| Sampling Method | SMOTE | None |
| Outlier Removal | Winsorization | None |
| Feature Selection Cutoff | 100% of Features | 30%, 70% of Features |
| Classifier | Support Vector Machine | KNN, Logistic, BayesNet, Random Forest |

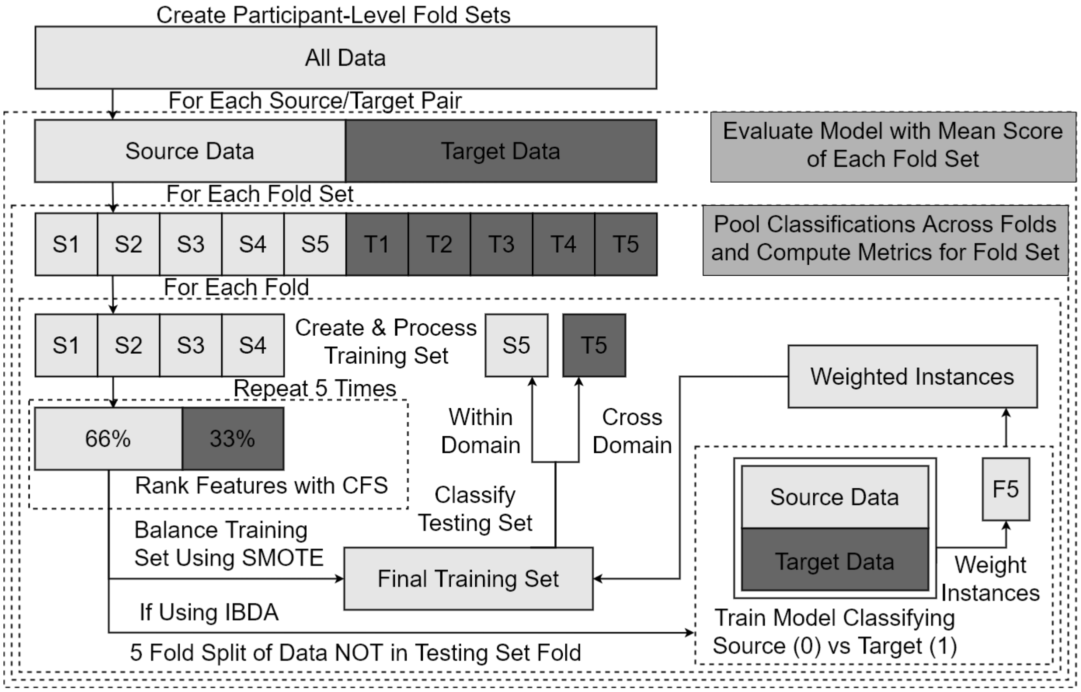

Create Participant-Level Fold Sets
All Data
For Each Source/Target Pair
Source Data
Target Data
Evaluate Model with Mean Score of Each Fold Set
For Each Fold Set
S1
S2
S3
S4
S5
T1
T2
T3
T4
T5
Pool Classifications Across Folds and Compute Metrics for Fold Set
For Each Fold
S1
S2
S3
S4
Create & Process Training Set
S5
T5
Weighted Instances
Repeat 5 Times
66%
33%
Within Domain
Cross Domain
Rank Features with CFS
Classify Testing Set
Source Data
Target Data
F5
Balance Training Set Using SMOTE
Final Training Set
Weight Instances
If Using IBDA
Train Model Classifying Source (0) vs Target (1)
5 Fold Split of Data NOT in Testing Set Fold

**Figure 2: Illustration of model building process with examples of how the data were processed in each step. The training set was processed prior to feature selection by removing collinear features, normalizing the data within each domain, and removing outliers within each domain, in that order. Domain adaptation techniques were applied after the classes in the training set were balanced using SMOTE, with Instance-Based Domain Adaptation (IBDA) shown here.**

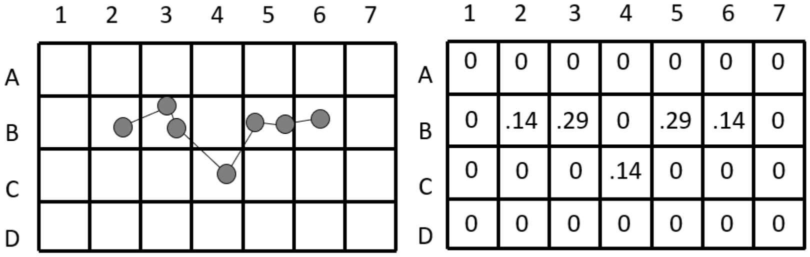
were summed across each of the five feature selection iterations. We then selected the best performing percent of features based on this ranking. The final step before classification was to balance the classes in the training set using SMOTE. The training and testing sets were then used to train and evaluate the classifier. This process was repeated for each fold, upon which each instance from the target domain had been classified once. We computed the accuracy metrics by pooling classifications across testing folds. upon which we averaged accuracies across the 50 iterations.

## 3.4 Domain Adaptation Approaches

*3.4.1 Investigating Domain Shift.* One of the primary hurdles in domain adaptation is domain shift. Two forms of this are covariate shift where the feature distributions are different from one domain to the next and prior probability (baserate) shift where the distribution of labels changes. As seen in Table 1, we have moderate prior probability shift due to the differences in MW rates between the domains. The least amount of MW occurs in the Film data with a rate of .27 compared to the .59 in the Scenes data. This is an indication that it will be more difficult to build a cross-domain model between these two domains. Covariate shift can be measured using the Jensen-Shannon distance, which yields a value of 0 for matching distributions. We computed the Jensen-Shannon distance for each distribution of feature values between each pair of domains. We then computed the average across all features that were not multicollinear in any domain, resulting in the values in Table 3. We found that the Illustrated Text and Scene domains were the most similar and the Film and Lecture domains were the least similar.

**Table 3: Jensen Shannon Distance between domains (min and max values bolded).**

| Source | Jensen Shannon Distance | | | |
|---|---|---|---|---|
| | Film | Lecture | Scene | Text |
| Illustrated Text | .26 | .27 | **.22** | .23 |
| Film | | **.29** | .26 | .25 |
| Lecture | | | .27 | .26 |
| Scene | | | | .23 |



| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| A | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| B | 0 | .14 | .29 | 0 | .29 | .14 | 0 |
| C | 0 | 0 | 0 | .14 | 0 | 0 | 0 |
| D | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

**Figure 3: An example of locality features using a 4x7 grid size. Each cell is a different feature (28 features in this example).**

These results indicate that covariate shift could cause some difficulties building cross-domain models. Accordingly, we investigated two approaches to address domain shift as discussed next.

*3.4.2 Locality Features.* The first approach to address domain shift was to investigate locality features that captured the proportional distribution of eye movements with respect to the stimulus as seen in Figure 3. These features are distinct from global features because they measure the spatial distribution of eye gaze in a stimulus-independent fashion. We selected grid sizes of 4x4 and 10x10, resulting in 16 and 100 features respectively. We built models for the locality features using the same type of SVM, data processing, and cross-validation procedure as the global models, with a few key differences. Specifically, we did not remove multicollinear features, normalize features, or remove outliers for the locality features in order to retain the raw values. The features were already normalized and comparable between domains and any removal of features would reduce screen coverage.

*3.4.3 Domain Adaptation.* The second approach was to apply either feature-level domain adaptation (FLDA) [Kouw et al., 2016] or instance-based domain adaptation (IBDA) [Shimodaira, 2000]. We chose techniques that operate under the assumption that labels were available for the source domain but not for the target domain. We used each of these methods with the same global features and hyperparameters as the baseline models. The training set was developed as in Figure 2 and participants in the training set folds were used for either FLDA or IBDA.

FLDA entails reweighting features based on how their distribution differs between the source and target domain. We performed FLDA using the python package libTLDA. This was done by analyzing the distribution of features in each domain, agnostic to class labels, and similar features received higher weights prior to training the models.

IBDA weights individual instances in the source domain based on their similarity to the target domain, which can then be used to train a weighted classifier. For IBDA, each instance in the training set was given a weight using a meta-classifier trained to distinguish instances of the source domain from those of the target domain. The first step was to split the participants in the training set into 5 random folds. The instances from the source domain for each of the participants in each of these folds constituted a nested testing set and the data from the source AND target domains for the participants from the other four folds were used as a nested training set. Each instance in the nested training set was labeled a 0 if they were from the source domain and a 1 if they were from the target domain. A separate SVM was trained using the nested training set and used to classify the instances in the nested testing set, resulting in a value between 0 and 1 for each of the instances in the nested testing set. The weight of each instance was set to this value, and once all folds were classified by the separate SVM, the weighted instances constituted a weighted training set. This weighted training set was then used to train the main SVM model to classify the testing set of the outer cross-validation fold.

## 4 RESULTS

We used two metrics to evaluate our models. We used the area under the receiver operating characteristic (AUROC) to select the best within-domain model (but also provide several comparison metrics). The AUROC is a commonly used metric where .50 represents chance and 1 represents perfect classification. Cross domain models were evaluated using the transfer ratio (TR) [Glorot et al., 2011]. The TR gives a measure of how well a cross-domain model performs in comparison to a within-domain model. It is computed using $TR = \frac{e(S,T)}{e_b(T,T)}$, where $e(S,T)$ is the performance of a model trained on a source domain $S$ and tested on a target domain $T$, and $e_b(T,T)$ is the performance of the baseline model trained and tested on the target domain $T$. Here we computed model performance as the AUROC

**Table 4: Metrics and Characteristics for Baseline Within-domain Models. Metrics include overall F1 score, F1 score for the MW class, precision, recall, kappa, and AUROC. Characteristics are the number of instances and corresponding MW rate.**

| Dataset | F1 | F1 MW | Precision | Recall | Kappa | **AUROC** | Instances | MW Rate |
|---|---|---|---|---|---|---|---|---|
| I. Text | 0.65 | 0.63 | 0.65 | 0.65 | 0.30 | **0.72** | 350 | 0.47 |
| Film | 0.61 | 0.40 | 0.65 | 0.58 | 0.10 | **0.58** | 356 | 0.27 |
| Lecture | 0.57 | 0.53 | 0.57 | 0.57 | 0.13 | **0.59** | 352 | 0.45 |
| Scene | 0.60 | 0.63 | 0.61 | 0.59 | 0.18 | **0.63** | 346 | 0.59 |
| Text | 0.57 | 0.48 | 0.58 | 0.57 | 0.12 | **0.57** | 354 | 0.40 |

**Table 5: AUROC and Transfer Ratio (TR) for each domain pairing. Values in each cell are the value when using the domain in the row to classify the domain in the column. The within-domain models are highlighted in light grey. Bolded cells indicate models that transferred well, with a TR greater than 0.50.**

| Source | Target AUROC | | | | | Target Transfer Ratio (TR) | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | I. Text | Film | Lecture | Scene | Text | I. Text | Film | Lecture | Scene | Text |
| I. Text | 0.72 | 0.56 | 0.60 | 0.64 | 0.51 | | **0.67** | **1.11** | **1.08** | 0.14 |
| Film | 0.56 | 0.58 | 0.53 | 0.54 | 0.56 | 0.27 | | 0.33 | 0.31 | **0.86** |
| Lecture | 0.68 | 0.54 | 0.59 | 0.63 | 0.50 | **0.82** | 0.44 | | **1.00** | 0.00 |
| Scene | 0.68 | 0.56 | 0.60 | 0.63 | 0.55 | **0.82** | **0.67** | **1.11** | | **0.71** |
| Text | 0.47 | 0.58 | 0.48 | 0.50 | 0.57 | -0.14 | **0.89** | -0.22 | 0.00 | |

of the model minus the AUROC baseline of .50 as we are concerned with how much better or worse the models performed relative to chance. With this modification, a value of 1 indicates that the cross-domain model was equivalent to the within-domain model, a value greater than 1 indicates improved performance, whereas values less than 1 indicate reduced performance (e.g., 0.5 would indicate a cross-domain model that is about half as much above chance as the baseline within-domain model). All metrics were computed using the average AUROC across the 50 participant-independent 5-fold cross validation iterations after pooling classifications across folds.

## 4.1 Within-Domain Models

We first report our baseline within-domain models using the parameters from Table 2 which serve as the point of comparison for all other models. As seen in Table 4, our models outperformed chance for all domains. The best performing model was built from the Illustrated Text data, with an AUROC of .72, while the worst performing model was in the Text domain with an AUROC of .57. There did not appear to be a clear relationship between performance and the number of instances or prior MW rate. These results are comparable with those from previous studies using similar data [Bixler and D'Mello, 2016; Hutt et al., 2017; Mills et al., 2016], with a best kappa value of .30 compared to .31 in [Bixler and D'Mello, 2016], an F1 MW of .57 (chance .45) for the Lecture domain compared to .41 (chance .30) in [Hutt et al., 2017], and an F1 of .61 for the Film domain compared to .53 in [Mills et al., 2016].

## 4.2 Cross Domain Models

We next report AUROC and TR (as discussed above) for cross domain models of MW (Table 5) built on global eye gaze data from one domain and tested on another. We consider models with a TR of .50 or above to show some evidence of transfer to the target domain. Each cell reflects performance of a given model trained using data from the domain in the rows and evaluated using data from the domain in the columns (diagonals reflect within-domain models). There are three main takeaways from our cross domain results: (1) The Illustrated Text, Scene, and Lecture domains transferred well among one another; (2) the Film and Text domains transferred well to one another; and (3) the Scene domain transferred to all other domains. The best TRs (1.11) were from the Illustrated Text and Scene source domains to the Lecture domain. Models trained on the Lecture domain, conversely, transferred to these other two domains to a lesser extent.

## 4.3 Composite Models

We also built composite models with two different source domains and compared them to the baseline cross-domain models in three ways (Table 6). The Source Average TR was computed as the average across all models that were trained using a given domain. For example, the Source Average TR for the Text domain was computed as the average of the four TR values in the bottom row of Table 5. This gives an indication of how well that domain transfers to the others, on average. The Target Average TR was computed as the average across all models that were evaluated on a given domain. For example, the Target Average TR for the Text domain was computed as the average of the four TR values in the last column of Table 5. Finally, we computed an average TR for all models, similar to how TR has previously been used [Glorot et al., 2011], which was the average of the 20 TR values in Table 5. The means for the composite models were computed similarly, except there were TR values for 30 models instead of 20. These results show that the composite models performed better on average, with a TR of .63 versus .54 for the single domain models. This was driven by improved Source TRs for models trained using data from the Film or Text domain.

**Table 6: Average Transfer Ratio (TR) for single domain models versus composite models. The average TR is computed as the mean across all models of the type corresponding to the row and with the same source (left side of table) or target (right side of table). The right most column is an average across all models for the given model type.**

| Model Type | Source Average TR | | | | | Target Average TR | | | | | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | I. Text | Film | Lecture | Scene | Text | I. Text | Film | Lecture | Scene | Text | TR |
| Baseline | 0.75 | 0.44 | 0.57 | 0.83 | 0.13 | 0.44 | 0.67 | 0.58 | 0.60 | 0.43 | 0.54 |
| Composite | 0.73 | 0.59 | 0.53 | 0.73 | 0.58 | 0.59 | 0.61 | 0.89 | 0.76 | 0.31 | 0.63 |

**Table 7: Transfer Ratio (TR) for the locality (L) and domain adaptation (FLDA and IBDA) models. The average Source and Target TR is computed as the mean across all models using the method corresponding to the row and with the same source (left side of table) or target (right side of table). Bolding indicates model performance above global models. The right most column is an average across all models in a row for the given method.**

| Method | Source Average TR | | | | | Target Average TR | | | | | Avg |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | I. Text | Film | Lecture | Scene | Text | I. Text | Film | Lecture | Scene | Text | TR |
| Global | 0.75 | 0.44 | 0.57 | 0.83 | 0.13 | 0.44 | 0.67 | 0.58 | 0.60 | 0.43 | 0.54 |
| L 4x4 | **0.90** | 0.10 | **0.67** | 0.70 | **0.70** | 0.38 | 0.31 | **0.88** | **0.61** | **0.89** | **0.61** |
| L 10x10 | 0.53 | 0.11 | 0.32 | 0.56 | 0.11 | 0.32 | 0.38 | 0.34 | 0.18 | 0.43 | 0.33 |
| G (FLDA) | 0.12 | 0.08 | 0.28 | 0.28 | **0.15** | 0.13 | 0.06 | 0.28 | 0.32 | 0.11 | 0.18 |
| G (IBDA) | 0.75 | 0.43 | 0.56 | 0.83 | 0.13 | 0.44 | 0.67 | 0.58 | **0.61** | 0.41 | 0.53 |

Conversely, composite models were worse at classifying these domains (i.e., lower Target TRs). This analysis indicates that there are likely small improvements that can be gained by combining multiple domains in a single model.

## 4.4 Locality and Domain Adaptation

Results for locality models and domain adaptation techniques are shown in Table 7. The key takeaways from this table are as follows. First, the locality models with a grid size of 4x4 performed better than the global eye gaze models on average. This is due to improvements in Source Average TR for the models trained on the Illustrated Text, Lecture, and Text data and comes with a tradeoff of reduced performance in Source Average TR for the models trained on Film and Scene data. Second, locality models with a grid size of 10x10 performed worse than the global models suggesting that larger grid sizes might spread the information out between too many features. Third, feature level domain adaptation (FLDA) performed worse than the global eye gaze model on average, and fourth instance-based domain adaptation (IBDA) performed similarly to the global eye gaze model on average. Thus, the domain adaptation techniques we used did not result in improved performance over the global models.

## 4.5 Feature Analysis

We analyzed our features by ranking the top 10 features within each domain. Features were ranked during feature selection as shown in Figure 2 and features with a higher rank received a higher score. The highest-ranking feature received points equal to the number of features included in the feature selection for that round, with each successively lower ranked feature receiving one less point. For example, if 20 features were ranked then the highest ranked feature would receive a score of 20 and the second highest feature would receive a score of 19. We then converted this score into a proportion of the total points available in a round. A round with 20 features would have a total of 210 points allocated among features, so the highest ranked feature would receive a score of .095. We then summed these scores across all folds and cross-validation within each domain. The scores for the top ten highest scoring features in each domain are shown in Table 8 ordered by their mean score across domains. The scores for features that were not within the top 10 highest scoring features in a domain are not included in the table.

We found that the saccade velocity median, saccade velocity kurtosis, and saccade amplitude kurtosis were the only three features that scored in the top ten in all domains. The horizontal saccade proportion and fixation saccade ratio were the highest scoring feature in two domains but absent in the Text domain, which may explain why the other domains had difficulty transferring to it. Seven features scored in the top ten for four of the domains, indicating that these features might have contributed to the transfer in our data.

# 5 DISCUSSION

Mind wandering (MW) detection within a single domain is a challenging task. In this work we investigated going beyond this to detect MW using eye gaze data from a different domain. We demonstrated the extent to which models constructed from global, stimulus independent eye gaze features could transfer from one domain to another, determined the effectiveness of three additional techniques to improve cross-domain model performance, and identified which features were most effective for building models. We review the main findings, discuss limitations and future work, and identify potential applications.

**Table 8: Top ten scoring features for each domain. The score for each feature in each domain is listed below the domain label, provided it was among the top ten scoring features for that domain. Features are ordered by mean score across all domains. The highest score in each domain is bolded, as are the features that scored in the top ten across all five domains.**

| Feature | I. Text | Film | Lecture | Scene | Text |
|---|---|---|---|---|---|
| Horizontal Sac Prop | **16.13** | 13.57 | 16.58 | **16.38** | |
| **Sac Vel Med** | 13.74 | 14.13 | 13.26 | 13.44 | 14.35 |
| Fix Sac Ratio | 14.33 | **17.05** | **17.29** | 15.11 | |
| Sac Angle Rel Max | 15.96 | 14.99 | 13.28 | 15.04 | |
| Sac Dur Range | 16.01 | | | 11.95 | **18.63** |
| **Sac Vel Kur** | 10.96 | 12.09 | 11.95 | 11.85 | 12.01 |
| Blink Count | 12.06 | | 15.71 | 10.85 | 10.10 |
| **Sac Amp Kur** | 12.46 | 11.46 | 11.89 | 11.61 | 9.91 |
| Sac Angle Rel SD | | 15.40 | 15.25 | 15.21 | |
| Fix Dur Med | | 12.06 | 12.42 | | 12.64 |
| Sac Angle Abs Med | | 10.84 | | | |
| Sac Angle Rel Mean | | | 14.43 | 14.31 | |
| Sac Angle Rel Med | | | | | 17.89 |
| Sac Dur Med | 11.55 | | | | |
| Sac Vel SD | | | | | 15.01 |
| Sac Dur Kur | | | | | 14.51 |
| Sac Angle Rel Skew | 12.46 | 11.85 | | | |

## 5.1 Main Findings

Our main finding was that MW models using global eye gaze features can transfer across domains with an average TR of .54, suggesting accuracies a bit over half than of within-domain models. We found that the Illustrated Text, Lecture, and Scenes domains transfer well to another, as did Film and Text domains. The Scenes domain was the only one that resulted in transfer across all domains, though it performed the worst on the Film and Text domains. Building composite models trained on two source domains resulted in a slight improvement with an average TR of .63, suggesting benefits to using data from multiple domains.

We also found that models built using locality features with a grid size of 4x4 resulted in improved performance overall compared to global features (on single domains), with an average TR of .61. This was due to an increase in performance for models built on the Illustrated Text, Lecture, and Text domains despite lower TRs for the Film and Scene domains. Adding feature-level domain adaptation to our global feature model building process resulted in an overall decrease in performance. Using instance-based domain adaptation resulted in models that were very similar to our baseline global feature models.

Finally, we investigated the top ten scoring features and found three that scored highly among all domains and three that scored highly in all domains except the Text domain. The domains with the most highly scored features in common were the Lecture and Scene domains with nine and the Illustrated Text and Scene domains with eight, which may have contributed to the high transfer ratio between these domains.

## 5.2 Limitations

This work has several limitations. First, the volume of data was low. This study had a total number of instances that is comparable to previous work in this area [Bixler and D'Mello, 2016; Hutt et al., 2019; Krasich et al., 2018; Mills et al., 2016; Bosch and D'mello, 2019] but experimental constraints on the amount of time per user effectively reduced the number of instances per domain, reducing performance in general. The lack of data also made it difficult to explore deep learning methods. The second limitation was the lack of an effective smooth pursuit classifier. The calibration procedure involved a 9-point calibration test, but the points did not move on the screen so smooth pursuit classification metrics could not be used to classify smooth pursuits. These features are not applicable to static domains such as the Illustrated Text, Scene, and Text but the alternative is lumping in smooth pursuits with fixations in the dynamic (Film and Lecture) domains. A third limitation is that this is a lab study with a somewhat homogeneous sample. It is therefore unclear how well these models would generalize when used outside the lab or to a different population. A final limitation is the use of domain-specific normalization, which improves the performance of within-domain models but can limit the ability of a system to generalize when employed in an unfamiliar new domain.

## 5.3 Applications and Future Work

One avenue for achieving better models is to expand data collection. Two to three times as much data would likely lead to better models and adapting the calibration test procedure to provide smooth pursuit metrics could lead to more accurate insights into building models for dynamic domains. Another important endeavor is to experiment with a wider array of domain adaptation techniques. These techniques would also provide insight into domain-general gaze-based models for other applications, such as those that seek to detect other mental states such as emotion or cognitive workload.

Future work should also focus on integrating these models in attention-aware technologies. One potential use case is a learning system that seeks to improve learning by attending to users' attention [D'Mello, 2019; Roda and Thomas, 2006]. Effective learning can involve multiple types of stimuli, but training data is not always easy to obtain. This work provides some evidence that cross-domain models such as the ones explored in this work can provide sufficient predictions for stimuli from some domains without the need for domain-specific training data. Exploring the most effective interventions to use in conjunction with the MW models is another important avenue for future work. A variety of interventions are possible, but it is important that they are robust to false positives due to the modest model accuracy. For example, switching the modality from text reading to viewing a video at a natural breaking point in the system in response to an elevated MW rate would be covert and have a lower chance of interfering with learning. Another potential application is a post hoc analyses that identifies periods of the interaction for content with high MW rates for future refinement.

## 5.4 Conclusion

One of the primary challenges to completing a task is our own wandering minds suggesting a potential for detecting and addressing the occurrence of MW in intelligent user interfaces. Because training MW models on a per-domain basis is a challenging endeavor, using previously collected data from one domain to detect MW in another could facilitate quicker and more widespread adoption of MW detection and interventions. This work demonstrated a modest ability to transfer eye gaze-based models of MW across domains, serving as a basis for future efforts to develop more accurate and generalizable models.

## ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (NSF) (DRL 1235958, IIS 1523091l, and DRL 1920510). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.

## REFERENCES

João Antunes and Pedro Santana. 2018. A Study on the Use of Eye Tracking to Adapt Gameplay and Procedural Content Generation in First-Person Shooter Games. Multimodal Technol. Interact. 2, 2 (May 2018), 23. DOI:https://doi.org/10.3390/mti2020023

Robert Bixler and Sidney D'Mello. 2016. Automatic gaze-based user-independent detection of mind wandering during computerized reading. User Model. User-Adapt. Interact. 26, 1 (March 2016), 33–68. DOI:https://doi.org/10.1007/s11257-015-9167-1

Nathaniel Blanchard, Robert Bixler, Tera Joyce, and Sidney K. D'Mello. 2014. Automated Physiological-Based Detection of Mind Wandering During Learning. In Intelligent Tutoring Systems, Springer, 55–60. DOI:https://doi.org/10.1007/978-3-319-07221-0_7

Nigel Bosch and Sidney K. D'mello. 2019. Automatic Detection of Mind Wandering from Video in the Lab and in the Classroom. IEEE Trans. Affect. Comput. (April 2019). DOI:https://doi.org/10.1109/TAFFC.2019.2908837

Charles V. Boys. 1890. Soap-bubbles, and the forces which mould them. Cornell University Library. Retrieved August 6, 2016 from http://projecteuclid.org/euclid.chmm/1424377189

Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-Sampling Technique. J. Artif. Intell. Res. 16, 1 (2002), 321–357.

Kalina Christoff. 2012. Undirected thought: Neural determinants and correlates. Brain Res. 1428, (January 2012), 51–59. DOI:https://doi.org/10.1016/j.brainres.2011.09.060

Kalina Christoff, Alan M. Gordon, Jonathan Smallwood, Rachelle Smith, and Jonathan W. Schooler. 2009. Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. Proc. Natl. Acad. Sci. 106, 21 (May 2009), 8719–8724. DOI:https://doi.org/10.1073/pnas.0900234106

Kalina Christoff, Zachary C. Irving, Kieran C. R. Fox, R. Nathan Spreng, and Jessica R. Andrews-Hanna. 2016. Mind-wandering as spontaneous thought: a dynamic framework. Nat. Rev. Neurosci. 17, 11 (November 2016), 718–731. DOI:https://doi.org/10.1038/nrn.2016.113

Kalina Christoff, Caitlin Mills, Jessica R Andrews-Hanna, Zachary C Irving, Evan Thompson, Kieran CR Fox, and Julia WY Kam. 2018. Mind-Wandering as a Scientific Concept: Cutting through the Definitional Haze. Trends Cogn. Sci. 22, 11 (November 2018), 957–959. DOI:https://doi.org/10.1016/j.tics.2018.07.004

Charles Clifton, Fernanda Ferreira, John M. Henderson, Albrecht W. Inhoff, Simon P. Liversedge, Erik D. Reichle, and Elizabeth R. Schotter. 2016. Eye movements in reading and information processing: Keith Rayner's 40 year legacy. J. Mem. Lang. 86, (January 2016), 1–19. DOI:https://doi.org/10.1016/j.jml.2015.07.004

M.R. Dias da Silva and M. Postma. 2020. Wandering minds, wandering mice: Computer mouse tracking as a method to detect mind wandering. Comput. Hum. Behav. 112, (November 2020), 106453. DOI:https://doi.org/10.1016/j.chb.2020.106453

Sidney K. D'Mello. 2019. Gaze-Based Attention-Aware Cyberlearning Technologies. In Mind, Brain and Technology: How People Learn in the Age of Emerging Technologies. Springer International Publishing, 87–106. DOI:https://doi.org/10.1007/978-3-030-02631-8_6

Sidney K. D'Mello and Art Graesser. 2014. Confusion and its dynamics during device comprehension with breakdown scenarios. Acta Psychol. (Amst.) 151, (September 2014), 106–116. DOI:https://doi.org/10.1016/j.actpsy.2014.06.005

Sidney K. D'Mello. 2016. Giving Eyesight to the Blind: Towards Attention-Aware AIED. Int. J. Artif. Intell. Educ. 26, 2 (June 2016), 645–659. DOI:https://doi.org/10.1007/s40593-016-0104-1

Sidney K. D'Mello, Caitlin Mills, Robert Bixler, and Nigel Bosch. 2017. Zone out No More: Mitigating Mind Wandering during Computerized Reading. In Proceedings of the 10th International Conference on Educational Data Mining, International Educational Data Mining Society, 8–15.

Sidney K. D'Mello, Kristopher Kopp, Robert Earl Bixler, and Nigel Bosch. 2016. Attending to Attention: Detecting and Combating Mind Wandering during Computerized Reading. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16), Association for Computing Machinery, New York, NY, USA, 1661–1669. DOI:https://doi.org/10.1145/2851581.2892329

Joanna Drummond and Diane Litman. 2010. In the Zone: Towards Detecting Student Zoning Out Using Supervised Machine Learning. In Intelligent Tutoring Systems, Springer, Berlin, Heidelberg, 306–308. DOI:https://doi.org/10.1007/978-3-642-13437-1_53

Myrthe Faber, Robert Bixler, and Sidney K. D'Mello. 2018. An automated behavioral measure of mind wandering during computerized reading. Behav. Res. Methods 50, 1 (February 2018), 134–150. DOI:https://doi.org/10.3758/s13428-017-0857-y

Myrthe Faber, Kristina Krasich, Robert E. Bixler, James R. Brockmole, and Sidney K. D'Mello. 2020. The eye–mind wandering link: Identifying gaze indices of mind wandering across tasks. J. Exp. Psychol. Hum. Percept. Perform. 46, 10 (October 2020), 1201–1221. DOI:https://doi.org/10.1037/xhp0000743

Shi Feng, Sidney D'Mello, and Arthur C. Graesser. 2013. Mind Wandering While Reading Easy and Difficult Texts. Psychon. Bull. Rev. 20, 3 (June 2013), 586–592. DOI:https://doi.org/10.3758/s13423-012-0367-y

L. Fletcher and A. Zelinsky. 2009. Driver Inattention Detection based on Eye Gaze–Road Event Correlation. Int. J. Robot. Res. 28, 6 (June 2009), 774–801. DOI:https://doi.org/10.1177/0278364908099459

Kieran CR Fox and Roger E Beaty. 2019. Mind-wandering as creative thinking: neural, psychological, and theoretical considerations. Curr. Opin. Behav. Sci. 27, (June 2019), 123–130. DOI:https://doi.org/10.1016/j.cobeha.2018.10.009

David J. Frank, Brent Nara, Michela Zavagnin, Dayna R. Touron, and Michael J. Kane. 2015. Validating older adults' reports of less mind-wandering: An examination of eye movements and dispositional influences. Psychol. Aging 30, 2 (2015), 266–278. DOI:https://doi.org/10.1037/pag0000031

Michael S. Franklin, Jonathan Smallwood, and Jonathan W. Schooler. 2011. Catching The Mind in Flight: Using Behavioral Indices to Detect Mindless Reading in Real Time. Psychon. Bull. Rev. 18, 5 (October 2011), 992–997. DOI:https://doi.org/10.3758/s13423-011-0109-6

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach. In Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML'11), Omnipress, Madison, WI, USA, 513–520. DOI:https://doi.org/10.5555/3104482.3104547

Mark A. Hall. 1999. Correlation-Based Feature Selection for Machine Learning. PhD Thesis. The University of Waikato.

Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA Data Mining Software: an Update. ACM SIGKDD Explor. Newsl. 11, 1 (November 2009), 10–18. DOI:https://doi.org/10.1145/1656274.1656278

Stephen Hutt, Jessica Hardey, Robert Bixler, Angela Stewart, Evan Risko, and Sidney K.

D'Mello. 2017. Gaze-based Detection of Mind Wandering during Lecture Viewing. In Proceedings of the 10th International Conference on Educational Data Mining, 226–231.

Stephen Hutt, Kristina Krasich, Sidney K. D'mello, and James R. Brockmole. 2021. Breaking out of the Lab: Mitigating Mind Wandering with Gaze-Based Attention-Aware Technology in Classrooms. In Proceedings of the ACM CHI Conference on Human Factors in Computing Systems, ACM, New York.

Stephen Hutt, Kristina Krasich, Caitlin Mills, Nigel Bosch, Shelby White, James R. Brockmole, and Sidney K. D'Mello. 2019. Automated gaze-based mind wandering detection during computerized learning in classrooms. User Model. User-Adapt. Interact. 29, 4 (September 2019), 821–867. DOI:https://doi.org/10.1007/s11257-019-09228-5

Christina Yi Jin, Jelmer P. Borst, and Marieke K. van Vugt. 2019. Predicting task-general mind-wandering with EEG. Cogn. Affect. Behav. Neurosci. 19, 4 (August 2019), 1059–1073. DOI:https://doi.org/10.3758/s13415-019-00707-1

Michael J Kane, Leslie H Brown, Jennifer C McVay, Paul J Silvia, Inez Myin-Germeys, and Thomas R Kwapil. 2007. For Whom the Mind Wanders, and When An Experience-Sampling Study of Working Memory and Executive Control in Daily Life. Psychol. Sci. 18, 7 (July 2007), 614–621. DOI:https://doi.org/10.1111/j.1467-9280.2007.01948.x.

M. A. Killingsworth and D. T. Gilbert. 2010. A Wandering Mind is an Unhappy Mind. Science 330, 6006 (November 2010), 932–932. DOI:https://doi.org/10.1126/science.1192439

Martin Kocur, Martin Johannes Dechant, Michael Lankes, Christian Wolff, and Regan Mandryk. 2020. Eye Caramba: Gaze-Based Assistance for Virtual Reality Aiming and Throwing Tasks in Games. In ACM Symposium on Eye Tracking Research and Applications (ETRA '20 Short Papers), Association for Computing Machinery, New York, NY, USA. DOI:https://doi.org/10.1145/3379156.3391841

Oleg V. Komogortsev, Sampath Jayarathna, Do Hyong Koh, and Sandeep Munikrishne Gowda. 2010. Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10), Association for Computing Machinery, New York, NY, USA, 65–68. DOI:https://doi.org/10.1145/1743666.1743682

Oleg V. Komogortsev and Alex Karpov. 2013. Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. Behav. Res. Methods 45, 1 (March 2013), 203–215. DOI:https://doi.org/10.3758/s13428-012-0234-9

Kristopher Kopp, Caitlin Mills, and Sidney K. D'Mello. 2016. Mind wandering during film comprehension: The role of prior knowledge and situational interest. Psychon. Bull. Rev. 23, 3 (June 2016), 842–848. DOI:https://doi.org/10.3758/s13423-015-0936-y

Wouter M Kouw, Jesse H Krijthe, and Marco Loog. 2016. Feature-Level Domain Adaptation. J. Mach. Learn. Res. 17, 171 (January 2016), 1–32. DOI:https://doi.org/10.5555/2946645.3053453

Wouter M. Kouw and Marco Loog. 2018. An introduction to domain adaptation and transfer learning. ArXiv181211806 Cs Stat (December 2018). Retrieved from http://arxiv.org/abs/1812.11806

Kristina Krasich, Greg Huffman, Myrthe Faber, and James R. Brockmole. 2020. Where the eyes wander: The relationship between mind wandering and fixation allocation to visually salient and semantically informative static scene content. J. Vis. 20, 9 (September 2020), 10. DOI:https://doi.org/10.1167/jov.20.9.10

Kristina Krasich, Robert McManus, Stephen Hutt, Myrthe Faber, Sidney K. D'Mello, and James R. Brockmole. 2018. Gaze-based signatures of mind wandering during real-world scene processing. J. Exp. Psychol. Gen. 147, 8 (August 2018), 1111–1124. DOI:https://doi.org/10.1037/xge0000411

Michael Lankes, Bernhard Maurer, and Barbara Stiglbauer. 2016. An Eye for an Eye: Gaze Input in Competitive Online Games and Its Effects on Social Presence. In Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology (ACE '16), Association for Computing Machinery, New York, NY, USA. DOI:https://doi.org/10.1145/3001773.3001774

Céline Lemercier, Christelle Pêcher, Gaëlle Berthié, Benoit Valéry, Vanessa Vidal, Pierre-Vincent Paubel, Maurice Cour, Alexandra Fort, Cédric Galéra, Catherine Gabaude, Emmanuel Lagarde, and Bertrand Maury. 2014. Inattention behind the wheel: How factual internal thoughts impact attentional control while driving. Saf. Sci. 62, (February 2014), 279–285. DOI:https://doi.org/10.1016/j.ssci.2013.08.011

Sophie I. Lindquist and John P. McLean. 2011. Daydreaming and its correlates in an educational environment. Learn. Individ. Differ. 21, 2 (April 2011), 158–167. DOI:https://doi.org/10.1016/j.lindif.2010.12.006

Lester C. Loschky, Adam M. Larson, Joseph P. Magliano, and Tim J. Smith. 2015. What Would Jaws Do? The Tyranny of Film and the Relationship between Gaze and Higher-Level Narrative Film Comprehension. PLOS ONE 10, 11 (November 2015), 1–23. DOI:https://doi.org/10.1371/journal.pone.0142474

Caitlin Mills, Robert Bixler, Xinyi Wang, and Sidney K. D'Mello. 2016. Automatic Gaze-Based Detection of Mind Wandering during Film Viewing. In Proceedings of the 9th International Conference on Educational Data Mining, International Educational Data Mining Society, 30–37.

Caitlin Mills, Nigel Bosch, Kristina Krasich, and Sidney K. D'Mello. 2019. Reducing Mind-Wandering During Vicarious Learning from an Intelligent Tutoring System. In Artificial Intelligence in Education, Seiji Isotani, Eva Millán, Amy Ogan, Peter Hastings, Bruce McLaren and Rose Luckin (eds.). Springer International Publishing, Cham, 296–307. DOI:https://doi.org/10.1007/978-3-030-23204-7_25

Caitlin Mills, Sidney K. D'Mello, Nigel Bosch, and Andrew M. Olney. 2015. Mind Wandering During Learning with an Intelligent Tutoring System. In Artificial Intelligence in Education, Springer, Cham, 267–276. DOI:https://doi.org/10.1007/978-3-319-19773-9_27

Caitlin Mills and Sidney K. D'Mello. 2015. Toward a Real-time (Day) Dreamcatcher: Detecting Mind Wandering Episodes During Online Reading. In Proceedings of the 8th International Conference on Educational Data Mining, International Educational Data Mining Society (IEDMS), 69–76. Retrieved from http://www.educationaldatamining.org/EDM2015/proceedings/full69-76.pdf

Caitlin Mills, Julie Gregg, Robert Bixler, and Sidney K D'Mello. 2020. Eye-Mind reader: an intelligent reading interface that promotes long-term comprehension by detecting and responding to mind wandering. Human–Computer Interact. (2020), 1–27. DOI:https://doi.org/10.1080/07370024.2020.1716762

Yukiko I Nakano, Cristina Conati, and Thomas Bader. 2013. Eye gaze in intelligent user interfaces: gaze-based analyses, models and applications (1st ed.). Springer-Verlag London.

Phuong Pham and Jingtao Wang. 2015. AttentiveLearner: Improving Mobile MOOC Learning via Implicit Heart Rate Tracking. In Artificial Intelligence in Education, Springer, Cham, 367–376. DOI:https://doi.org/10.1007/978-3-319-19773-9_37

Jason G. Randall, Frederick L. Oswald, and Margaret E. Beier. 2014. Mind-wandering, cognition, and performance: A theory-driven meta-analysis of attention regulation. Psychol. Bull. 140, 6 (November 2014), 1411–1431. DOI:https://doi.org/10.1037/a0037428

Keith Rayner. 1998. Eye Movements in Reading and Information Processing: 20 Years of Research. Psychol. Bull. 124, 3 (November 1998), 372–422. DOI:https://doi.org/10.1037/0033-2909.124.3.372

Keith Rayner. 2009. Eye movements and attention in reading, scene perception, and visual search. Q. J. Exp. Psychol. 62, 8 (August 2009), 1457–1506. DOI:https://doi.org/10.1080/17470210902816461

Erik D. Reichle, A. E. Reineberg, and J. W. Schooler. 2010. Eye Movements During Mindless Reading. Psychol. Sci. 21, 9 (August 2010), 1300–1310. DOI:https://doi.org/10.1177/0956797610378686

Ian H Robertson, Tom Manly, Jackie Andrade, Bart T Baddeley, and Jenny Yiend. 1997. "Oops!": Performance Correlates of Everyday Attentional Failures in Traumatic Brain Injured and Normal Subjects. Neuropsychologia 35, 6 (June 1997), 747–758. DOI:https://doi.org/10.1016/s0028-3932(97)00015-8

Claudia Roda and Julie Thomas. 2006. Attention aware systems: Theories, applications, and research agenda. Comput. Hum. Behav. 22, 4 (July 2006), 557–587. DOI:https://doi.org/10.1016/j.chb.2005.12.005

Southwell Rosy, Julie Gregg, Robert Bixler, and Sidney K. D'mello. In Press. What Eye Movements Reveal about Comprehension during Naturalistic Reading of Long, Connected Texts. Cogn. Sci. (In Press).

Jonathan W. Schooler, Erik D. Reichle, and David V. Halpern. 2004. Zoning Out While Reading: Evidence for Dissociations Between Experience and Metaconsciousness. In Thinking and seeing: visual metacognition in adults and children, Daniel T Levin (ed.). MIT Press, Cambridge, Mass., 203–226.

Jonathan W. Schooler, Jonathan Smallwood, Kalina Christoff, Todd C. Handy, Erik D. Reichle, and Michael A. Sayette. 2011. Meta-awareness, perceptual decoupling and the wandering mind. Trends Cogn. Sci. (June 2011). DOI:https://doi.org/10.1016/j.tics.2011.05.006

Pennie S. Seibert and Henry C. Ellis. 1991. Irrelevant Thoughts, Emotional Mood States, and Cognitive Task Performance. Mem. Cognit. 19, 5 (1991), 507–513. DOI:https://doi.org/10.3758/BF03199574

Paul Seli, Roger E Beaty, James Allan Cheyne, Daniel Smilek, Jonathan Oakman, and Daniel L Schacter. 2018a. How pervasive is mind wandering, really? Conscious. Cogn. 66, (2018), 74–78. DOI:https://doi.org/10.1016/j.concog.2018.10.002

Paul Seli, Michael J. Kane, Thomas Metzinger, Jonathan Smallwood, Daniel L. Schacter, David Maillet, Jonathan W. Schooler, and Daniel Smilek. 2018b. The Family-Resemblances Framework for Mind-Wandering Remains Well Clad. Trends Cogn. Sci. 22, 11 (November 2018), 959–961. DOI:https://doi.org/10.1016/j.tics.2018.07.007

Paul Seli, Michael J. Kane, Jonathan Smallwood, Daniel L. Schacter, David Maillet, Jonathan W. Schooler, and Daniel Smilek. 2018c. Mind-Wandering as a Natural Kind: A Family-Resemblances View. Trends Cogn. Sci. 22, 6 (June 2018), 479–490. DOI:https://doi.org/10.1016/j.tics.2018.03.010

Hidetoshi Shimodaira. 2000. Improving predictive inference under covariate shift by weighting the log-likelihood function. J. Stat. Plan. Inference 90, 2 (October 2000), 227–244. DOI:https://doi.org/10.1016/S0378-3758(00)00115-4

Jonathan Smallwood. 2011. Mind-wandering While Reading: Attentional Decoupling, Mindless Reading and the Cascade Model of Inattention. Lang. Linguist. Compass 5, 2 (February 2011), 63–77. DOI:https://doi.org/10.1111/j.1749-818X.2010.00263.x

Jonathan Smallwood, Emily Beach, Jonathan W. Schooler, and Todd C. Handy. 2008. Going AWOL in the brain: Mind wandering reduces cortical analysis of external events. J. Cogn. Neurosci. 20, 3 (2008), 458–469. DOI:https://doi.org/10.1162/jocn.2008.20037

Jonathan Smallwood, John B. Davies, Derek Heim, Frances Finnigan, Megan Sudberry, Rory O'Connor, and Marc Obonsawin. 2004. Subjective Experience and the Attentional Lapse: Task Engagement and Disengagement During Sustained Attention. Conscious. Cogn. 13, 4 (December 2004), 657–690. DOI:https://doi.org/10.1016/j.concog.2004.06.003

Jonathan Smallwood, Daniel J. Fishman, and Jonathan W. Schooler. 2007a. Counting the Cost of an Absent Mind: Mind Wandering as an Underrecognized Influence on Educational Performance. Psychon. Bull. Rev. 14, 2 (2007), 230–236. DOI:https://doi.org/doi.org/10.3758/BF03194057

Jonathan Smallwood, Merrill McSpadden, and Jonathan W. Schooler. 2007b. The Lights are On but No One's Home: Meta-Awareness and the Decoupling of Attention when the Mind Wanders. Psychon. Bull. Rev. 14, 3 (2007), 527–533. DOI:https://doi.org/doi.org/10.3758/BF03194102

Jonathan Smallwood and Jonathan W. Schooler. 2006. The Restless Mind. Psychol. Bull. 132, 6 (2006), 946–958. DOI:https://doi.org/10.1037/0033-2909.132.6.946

D. Smilek, J. S. A. Carriere, and J. A. Cheyne. 2010. Out of Mind, Out of Sight: Eye Blinking as Indicator and Embodiment of Mind Wandering. Psychol. Sci. 21, 6 (June 2010), 786–789. DOI:https://doi.org/10.1177/0956797610368063

Angela Stewart, Nigel Bosch, Huili Chen, Patrick Donnelly, and Sidney K. D'Mello. 2017a. Face forward: Detecting mind wandering from video during narrative film comprehension. In International Conference on Artificial Intelligence in Education, Springer, Cham, 359–370. DOI:https://doi.org/doi.org/10.1007/978-3-319-61425-0_30

Angela Stewart, Nigel Bosch, and Sidney K. D'Mello. 2017b. Generalizability of Face-Based Mind Wandering Detection Across Task Contexts. In International Educational Data Mining Society, International Educational Data Mining Society, 88–95.

Veronica Sundstedt. 2012. Gazing at games: An introduction to eye tracking control. Synth. Lect. Comput. Graph. Animat. 5, 1 (2012), 1–113. DOI:https://doi.org/10.2200/S00395ED1V01Y201111CGR014

Ashish Tawari, Sayanan Sivaraman, Mohan Manubhai Trivedi, Trevor Shannon, and Mario Tippelhofer. 2014. Looking-in and looking-out vision for urban intelligent assistance: Estimation of driver attentive state and dynamic surround for safe merging and braking. In Intelligent Vehicles Symposium Proceedings, IEEE, 115–120. DOI:https://doi.org/10.1109/IVS.2014.6856600

Nash Unsworth and Brittany D. McMillan. 2013. Mind Wandering and Reading Comprehension: Examining the Roles of Working Memory Capacity, Interest, Motivation, and Topic Experience. J. Exp. Psychol. Learn. Mem. Cogn. 39, 3 (2013), 832–842. DOI:https://doi.org/10.1037/a0029669

Adrian Voßkühler, Volkhard Nordmeier, Lars Kuchinke, and Arthur M. Jacobs. 2008. OGAMA (Open Gaze and Mouse Analyzer): Open-Source Software Designed to Analyze Eye and Mouse Movements in Slideshow Study Designs. Behav. Res. Methods 40, 4 (November 2008), 1150–1162. DOI:https://doi.org/10.3758/BRM.40.4.1150

Yana Weinstein. 2018. Mind-wandering, how do I measure thee with probes? Let me count the ways. Behav. Res. Methods 50, 2 (April 2018), 642–661. DOI:https://doi.org/10.3758/s13428-017-0891-9

Jeffrey M. Zacks. 2010. The brain's cutting-room floor: segmentation of narrative cinema. Front. Hum. Neurosci. 4, (2010). DOI:https://doi.org/10.3389/fnhum.2010.00168